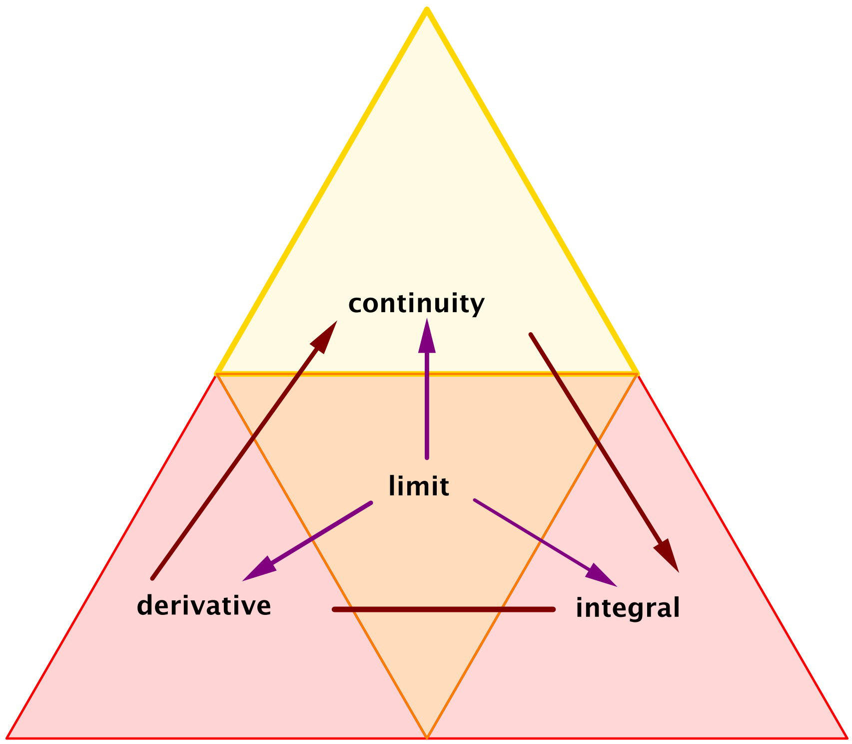
MARIANO ECHEVERRIA

CALCULUS


continuity
limit
derivative
integral

What is Calculus All About?

In a nutshell, Calculus is the study of mathematical operations that can't be realized in a *finite* number of steps and so an *infinite* process is necessary to perform them. The kind of infinity that appears in Calculus is handled through the *limit* concept and it provides the foundations for the key ideas of Calculus: continuity, differentiability and integrability.

Limits turn out to be very different from the previous mathematics you have studied so far, and the conceptual change required is what gives Calculus its reputation for being difficult. The following examples show how the idea of limit appears in calculus.

Approximating $\sqrt{2}$

Suppose we construct a square triangle whose legs are of length 1, and we want to determine the length of the hypothenuse. If we call the length of this side L Pythagoras' Theorem says that

$$L^2 = 1^2 + 1^2 = 2 \tag{1}$$

and the number that represents this length is called $\sqrt{2}$. Now, saying that $L = \sqrt{2}$ is just a way of naming this new length; in practice we are interested of representing L in more familiar terms. For example, to find $1 + \frac{1}{2}$ we can first write $\frac{1}{2} = 0.5$ and then $1 + \frac{1}{2} = 1 + 0.5 = 1.5$. However, if we want to find $1 + \sqrt{2}$ we would be stuck unless we represent $\sqrt{2}$ like $\frac{1}{2}$ was represented.

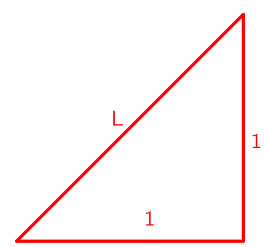


Figure 1: Pythagoras' Theorem

To find a representation for $\sqrt{2}$ we might try the following: since $\sqrt{2}$ must satisfy $\left(\sqrt{2}\right)^2 = 2$ and $1^2 = 1$, $2^2 = 4$ it follows that [1] $\sqrt{2}$ is between 1 and 2. This means that $\sqrt{2}$ should be something like $1.[]$ where $[]$ denotes a "black box" in which we need to add more digits.

[1] We are secretly using the property that if a, b, c are three positive numbers satisfying $a^2 < b^2 < c^2$ then it must be the case that $a < b < c$

To find what is inside our black box, observe that $(1.4)^2 = 1.98$ and $(1.5)^2 = 2.25$ which implies that $\sqrt{2}$ is between 1.4 and 1.5, therefore $\sqrt{2}$ must be of the form $1.4[]$ where $[]$ is a new black box. If we continue in this fashion we can build a table for $\sqrt{2}$:

$$\begin{array}{c} 1 \\ 1.4 \\ 1.41 \\ 1.414 \\ 1.4142 \\ \vdots \\ 1.414213562 \\ \vdots \end{array} \tag{2}$$

Table 1: Approximations to $\sqrt{2}$

After a while one might start getting frustrated because it seems that we can go on and on. Is there a way to know whether this process will end or not? The answer is that this process will never end! This is because, as the pythagoreans showed, $\sqrt{2}$ is irrational, that is, $\sqrt{2}$ can't be written as a quotient of two integers like $\frac{2}{3}$, $\frac{7}{9}$, etc. The fact that $\sqrt{2}$ is irrational implies that the table 2 will never end, and there won't be a pattern to the successive digits that appear (which is different from numbers like $\frac{10}{3}$, which can be written as an "infinite" sequence $3.33333\cdots$ where there is a clear pattern). However, we would like to claim that $\sqrt{2}$ is in some sense "the limit" of the approximations in 2, and although we won't address this particular problem any more, this sort of limit is similar to the ones that will be studied during the course. The most important aspect to note at this point is that the approximations in 2 get closer and closer to $\sqrt{2}$, without ever being equal to $\sqrt{2}$.

What is $2^{\sqrt{2}}$?

In a similar vein, suppose that you need to explain what is $2^{\sqrt{2}}$: at first this may sound like a silly problem but it is actually more complicated than it seems. For example, the meaning of 2^3 is clear: you just need to multiply the number 2 three times with itself, that is,

$$2^3 = 2 \cdot 2 \cdot 2 \tag{3}$$

However, since numbers can only be multiplied a integer number of times, we can't say that $2^{\sqrt{2}}$ means multiply 2 square root of two times. Therefore, some work is needed to give a precise meaning to this notion.

At this point it is useful to recall how other powers of 2 are defined, for example, how 2^0, 2^{-4}, $2^{1/3}$ and $2^{4/3}$ are defined. The key idea is that we need to define the previous numbers in such a way that the familiar properties of the powers of 2 keep working in this new setting. For example, when a and b are positive integers we know that $2^a \cdot 2^b = 2^{a+b}$. Therefore, we define 2^0 in such a way that this property works even when a or b is zero. For example, if we take $a = 3$ and $b = 0$ this means that we want

$$2^3 \cdot 2^0 = 2^{3+0} \tag{4}$$

and since $2^{3+0} = 2^3$ the last equation is equivalent to

$$2^3 \cdot 2^0 = 2^3 \tag{5}$$

which forces 2^0 to be equal to 1 since we can cancel 2^3 on both sides of the equation. So we conclude that we should define[2]

$$2^0 \equiv 1 \tag{6}$$

[2] from now on the symbol $\equiv$ will be used every time a definition is made

To define 2^{-4} we use a similar procedure: now we want $2^a \cdot 2^b = 2^{a+b}$ to be true even if a and/or b are negative. In particular, if we take $a = 4$ and $b = -4$ this means that

$$2^4 \cdot 2^{-4} = 2^{4-4} \tag{7}$$

Since we already defined 2^0 the last equation is equivalent to

$$2^4 \cdot 2^{-4} = 1 \tag{8}$$

so in order for this equation to be true we must define

$$2^{-4} \equiv \frac{1}{2^4} \tag{9}$$

To define $2^{1/3}$ we use the fact that $(2^a)^b = 2^{ab}$ whenever a and b are positive integers and as the previous formula we assume that this property will still be true even if a and/or b are rational numbers. That is, if we take $a = \frac{1}{3}$ and $b = 3$ we want the following equation to be true

$$\left(2^{\frac{1}{3}}\right)^3 = 2^{\frac{1}{3}\cdot 3} \tag{10}$$

since $\frac{1}{3} \cdot 3 = 1$ the previous equation is the same as

$$\left(2^{\frac{1}{3}}\right)^3 = 2 \tag{11}$$

so we should define $2^{\frac{1}{3}}$ as

$$2^{\frac{1}{3}} \equiv \text{the number } x \text{ with the property that } x^3 = 2 \tag{12}$$

Maybe this last definition seems a little suspicious: why should there be a number with such a property? For example, if we had defined $2^{\frac{1}{3}}$ as the number x of brothers of planet Venus then we would be stuck since planets have no brothers so the definition is ill defined. Therefore, some argument has to be given to guarantee that our definition for 12 is legitimate. This argument will be given later in the course with the help of the intermediate value theorem. Notice that we wrote '*the* number' instead of '*a* number'; that is because it can be shown that there is only one number which satisfies the property specified in 12. [3]

[3] This is because we will be working only with real numbers, if we allow complex numbers (for those of you who know what they are) there are in fact two more numbers that satisfy property 12

Finally, to define a number like $2^{4/3}$ we simply stipulate that

$$2^{\frac{4}{3}} \equiv \left(2^{\frac{1}{3}}\right)^4 \tag{13}$$

A similar procedure allows us to calculate $2^{\frac{a}{b}}$ whenever a, b are integers, positive or negative with b different from 0. However, this is still not enough to calculate a number like $2^{\sqrt{2}}$. To do this we need to use the rational approximations we found to $\sqrt{2}$ in 2. The idea would be to approximate $2^{\sqrt{2}}$ as the following table shows: observe that for clarity we rewrote each power as a fraction so that it is clear what to do.

$$\begin{array}{c} 2^1 \\ 2^{\frac{14}{10}} \\ 2^{\frac{141}{100}} \\ 2^{\frac{1414}{1000}} \\ 2^{\frac{14142}{10000}} \\ \vdots \\ 2^{\frac{1414213562}{1000000000}} \\ \vdots \end{array} \tag{14}$$

Table 2: Approximations to $\sqrt{2}$

Again, we would like to say that $2^{\sqrt{2}}$ is the "limit" of the previous approximations. As will be made clear later, there are many other situations in which one needs a similar type of approximation: the two most important for calculus being finding the derivative and the integral of a function.

How to find the biggest rectangle with a given perimeter?

Suppose that we have 40 feet of cable and we want to use it to build a rectangle. Since there are different choices we can make for the length and width of the rectangle (as the table on the right shows) we could build rectangles having very different areas.

Width (ft)	Length (ft)	Area $(\text{ft})^2$
0.1	19.9	1.99
0.5	19.5	9.75
3	17	51
5	15	75
8	12	96
10	10	100

Table 3: Choices for the sides of the rectangle

Suppose we are interested in building the rectangle with the *biggest* amount of area. From the previous table it seems that the rectangle with the biggest area is actually the square. Now, although empirical evidence should be taken seriously, it cannot be considered sufficient if we want to use it to show the validity of a mathematical statement. We need to provide a *proof* that shows that no matter which sides our rectangle have, its area will never be as big as the square. The following argument provides such a proof.

If we start with square then any rectangle of length 40 ft can be obtained from that square by increasing a side of the square by a quantity $\triangle x$ and decreasing the other side of the square by the same amount $\triangle x$. The area of that rectangle is

$$\text{area rectangle } = (10+\triangle x)\,(10-\triangle x) = 100-(\triangle x)^2 \tag{15}$$

and since $(\triangle x)^2$ is always a positive number then $100-(\triangle x)^2$ will be less than 100, which is the area of the square. Now, there was nothing special about the number 40, any other length for the cable would have worked just as well so we have shown that[4]

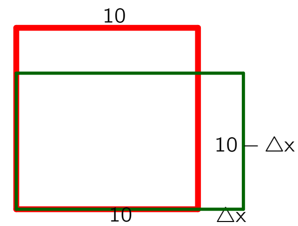


Figure 2: Square is the Rectangle with the Biggest Area

[4] It is interesting to observe that if had wanted to find the rectangle with *smallest* amount of area then our problem would not have had a solution, that is, it is possible to build rectangles of arbitrarily small area. This may not be obvious but can be easily proven once we know some calculus. Moreover, this also shows that there is no reason to expect correlations between quantities like perimeter and area, that is, the area of an object can increase without having to increase its perimeter and vice-versa. This helps to understand why the circulatory system can have around 96 560 km (60 000 miles) of capillaries with a total surface area of some 800 -1000 m^2 (an area greater than three tennis courts) and a volume of just 5 liters.

Among all rectangles with a *given* perimeter, the square is the one with the biggest area

For a more complicated problem it might not be possible to find such an elegant solution so it is useful to restate it in a more suitable way to Calculus. This restatement will also show another advantage of Calculus: its ability to solve simultaneously an *infinite* number of problems. We will state our problem in the following way:

Among all rectangles with a *given* perimeter P, find which rectangle has the biggest area.

Observe that P is not a *variable* but rather a *parameter*; that is, it is a constant whose value we choose not to use explicitly. The advantage of not using explicitly the value of the perimeter is that in this way we can solve simultaneously infinitely many problems, one for each different value of P. For example, when $P=40$ we are solving the previous problem and if $P=20$ then we are solving a different, albeit similar problem. Also, another advantage of not using the value of P explicitly is that we don't have to worry about the *units* of our quantities. For example, since 40 ft is approximately equal to 12.2 meters then we should find that the square has the biggest area when $P=40$ or when $P=12.2$ but since in mathematics we typically don't show units explicitly it looks like we are solving two different problems even tough they represent the same physical situation.

Returning to our problem we want to build rectangles with a perimeter P. If we call x and y the sides of the rectangle and we let A be its area then we have the following two equations

$$\begin{cases} P = 2x+2y \\ A = xy \end{cases} \tag{16}$$

Since P is actually a constant, we can use the first equation to write one of the variables as a function of the other variable, for example

$$y = \frac{P - 2x}{2} \tag{17}$$

If we substitute y in the second equation then this allows us to consider the area A as a function of x, that is,

$$A(x) = x\left(\frac{P - 2x}{2}\right) = \frac{Px - 2x^2}{2} \tag{18}$$

Therefore, we are trying to find the value of x which gives us the biggest value of $A(x)$. In Calculus jargon we are trying to find the value x which *optimizes* the function $A(x)$. If we plot the graph of $A(x)$ it is obvious[5] (as the next figure shows) that the biggest value of A occurs when $x = \frac{P}{4}$. Substituting in 17 we obtain that $y = \frac{P}{4}$ so we see that we must build a square whose sides have length $\frac{P}{4}$. If we take $P = 40$ then we obtain our previous result and we also see that since P could have any value then in fact we solved simultaneously an infinite number of problems as we mentioned earlier.

[5] the methods of Calculus will give us a technique to find the graph of a function like 18, the other option would be to remember that equation 18 represents a parabola.

Now, let's assume that we want to characterize geometrically the value x which gives the biggest value of $A(x)$. As we can see from the figure, the graph of $A(x)$ is actually a curve and as we will learn soon Calculus allows us to find for any (reasonable) curve the tangent line to any given point on such a curve. From the figure it should be clear that the tangent line (which is purple in the figure) to the curve at the value of the maximum is actually horizontal, that is, has slope zero. That is,

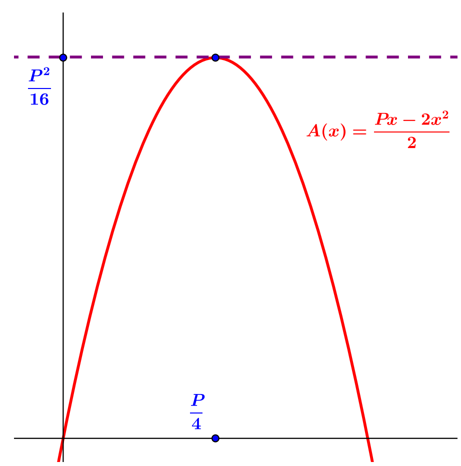


Figure 3: Graph of $A(x) = \frac{Px-2x^2}{2}$ and tangent line

The tangent line to the curve going through the maximum value of the curve has slope zero.

We will see later that the previous statement is true in a general situation, however, this property of having a zero slope will be a *necessary* but not *sufficient* condition for the existence of a maximum. [6]

[6] As an analogy, *being human* is a necessary condition for *being a man*; however, it is not a sufficient condition since you can be a human without being a man.

Describing Motion in Geometric Terms

Suppose that we have three people running a race as the next animation shows. It is clear from the animation that the people run at different speeds during the race and their motion is not too difficult to describe. However, suppose that for some reason we are only able to see the entire trajectory of each runner, not what is happening at each instant. For example, suppose that each runner had some paint and as they run the road is being painted with their particular color. In that case it would not be clear what was happening with each runner during the race since we would only see the same segment (painted in three different ways) and we would not be able to distinguish one from the other.

Therefore, it would be very useful to find a way to represent their motion that would allow us to distinguish the three runners, even if we have access only to their entire trajectory. The way to do this is with the help of a space-time diagram.

The idea of a space-time diagram is as follows. Given that each runner occupies a particular location at each instant of time, we can associate to each runner a function of time $r_1(t)$, $r_2(t)$, $r_3(t)$ such that

$$\begin{cases} r_1(t) = \text{position of runner 1 at time } t \\ r_2(t) = \text{position of runner 2 at time } t \\ r_3(t) = \text{position of runner 3 at time } t \end{cases} \tag{19}$$

Now, a space-time diagram is simply a diagram in which we plot the functions $r_1(t)$, $r_2(t)$, $r_3(t)$; its name comes from the fact that one of the axis of our diagram is used to represent the position (space) of the runners and the other axis is used to represent each particular instant of time of the race. On the space-time diagram[7], each function is represented by a curve which is called the world line of the runner. The following animation shows the space time diagram of our runners. Using the space-time diagram, it is clear that the runners have different motions. In fact, as we will see later, with a space-time diagram we can find any kinematical quantity that we care about like velocity and acceleration by computing geometric properties of the world lines.

[7] It should be noted that we use the vertical axis to represent the space variable and the horizontal axis to represent the time variable while physicists use the reverse convention in Relativity Theory.

Solving Mathematical Models

Many elementary algebra problems are about solving equations like

$$x^2 = 4 - 3x \tag{20}$$

It is easy to see that the solutions of the previous equation are $x = 1$ and $x = -4$. These solutions can be found using an entirely algebraic approach (for example, via the factorization method) but it is interesting to see if there is a geometric interpretation to equation 20. The idea is to consider each side of the equation as a different function. For example, the left hand side would represent the function $y = x^2$ which is the equation of a parabola while the right hand side would represent the function $y = 4 - 3x$ which is the equation of a straight line. Therefore, we interpret equation 20 in the following way: $x^2 = 4 - 3x$ can be solved if and only if the curves $y = x^2$ and $y = 4 - 3x$ intersect.

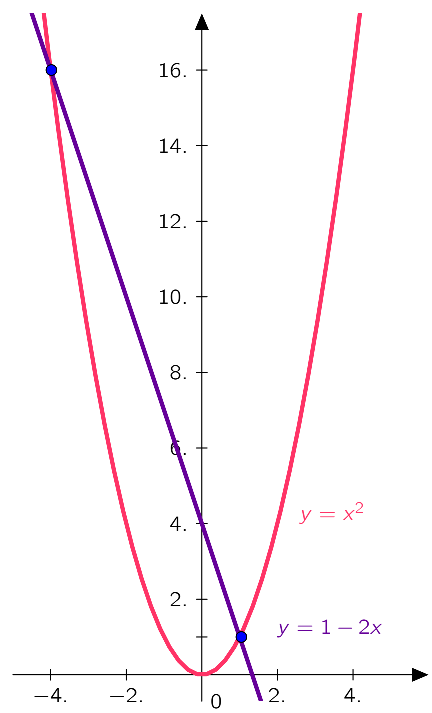


Figure 4: $x^2 = 4 - 3x$ and intersection of curves

The figure on the right shows that the two curves intersect in two points whose x coordinates are the solutions of the equation $x^2 = 4 - 3x$. Moreover, using this interpretation, it is clear why some equations like $x^2 = -1$ have no (real) solution.

In Calculus we will be interested in solving equations which are called differential equations. An example of a differential equation is (don't worry if you don't understand what it says!)

$$\frac{dN(t)}{dt} = kN(t) \tag{21}$$

The idea is to think that equation 21 is a mathematical model for some phenomena in the natural world. For example, if we think of $N(t)$ as representing the size of some population at time t, then 21 gives a model on how we expect the population to change in time. The techniques of this course will tell us how to solve equation 21. Instead of finding a finite number of points (which is what happens when we solve an algebraic equation) when solving a differential equation we will find a curve which "follows" some "arrows" specified by 21 as the next figure shown. Solving the differential equation requires a technique known as integration, which is the second pillar of Calculus.

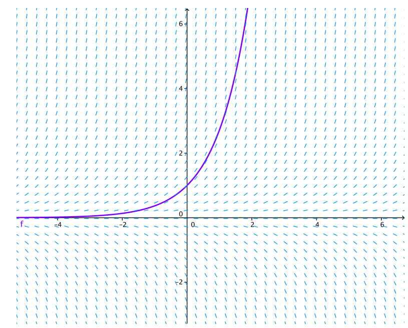


Figure 5: Differential Equation $\frac{dN}{dt} = kN$

In fact, the Fundamental Theorem of Calculus states that in some sense differentiation and integration are inverse processes (as an analogy, differentiation would correspond to running a movie forward while integration would correspond to running a movie backwards). What makes Calculus so useful is that many quantities/variables in the social and natural sciences are related via the process of differentiation and integration, as the following table illustrates.

Quantity A	$\longrightarrow$ differentiation $\longleftarrow$ integration	Quantity B	Variable
Position of a particle		Velocity	time
Velocity of a particle		Acceleration	time
Cost of Living		Inflation Rate	time
Total cost of some goods		Marginal Cost	quantity
Height above sea level on a trail		Steepness	distance
Mass of a rod		linear density	length
Height of a tree		growth rate	time
Work (physics)		Power (physics)	time
Potential Energy (physics)		Force	position

Finding the Area of a Circle

As we will later see, the Fundamental Theorem of Calculus is so surprising at first because integration is defined without any reference to the concept of derivative. Actually, in some sense the concept of integration was discovered[8] earlier than the concept of derivatives as the greeks tried to find the area of geometric figures.

[8] Created could be a more accurate word if you believe that Mathematics is essentially a human creation like Chess or Soccer

Assume that we know that the circumference of a circle of radius R is $C = 2\pi R$. How can we find the area of that circle? The main idea is to approximate the area of the circle by the area of figures which we can calculate easily. For example, we can imagine that we cut the circle into equal slices as the next figure shows.

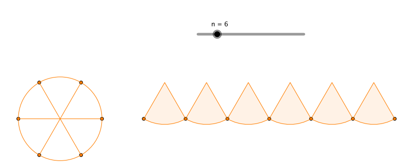


Figure 6: Cutting a circle into slices

The total area of the circle is equal to the sum of the areas of the circular sectors (pizza slices). As the following animation shows, as we cut the circle into more and more slices, the circular sectors look more like triangles. Therefore, we would like to say that we can approximate the area of the circle as the sum of the area of those triangles [9]

$$\text{area circle} \simeq \text{sum area of triangles} \tag{22}$$

where $\simeq$ means "approximately equal to". Now, if we divide the circle into N slices and N is very big [10] then we can make the approximation

$$\begin{aligned} & \text{sum area of triangles} \\ = & \quad N \text{ area of a single triangle} \\ = & \quad N\left(\tfrac{1}{2}\text{base} \cdot \text{height}\right) \\ \simeq & \quad \frac{N}{2}\left(\frac{\text{circumference}}{N}\right)(R) \\ = & \quad \pi R^2 \end{aligned} \tag{23}$$

and thus we have obtained that the area of a circle of radius R is πR^2! Clearly this argument was not a rigorous as the one we used to show that a square encloses the maximum of area among all rectangles; in particular, we never specified what we meant by approximation. That is, we never said what is the *error* we are making when using this approximation. In fact, it took a long time (centuries or decades depending on when you start counting) for all of these ideas to be made precise but in any case it should be surprising that we obtained the correct formula for the area of a circle with so little effort!

[9] the precise statement will be that the *limit* of the sum of the areas of the triangles is the area of the circle

[10] again the precise statement will be that we take $N \longrightarrow \infty$ where N is the number of slices

Contents

Part I

Preliminaries

The Straight Line and the Real Numbers

In Calculus the most important functions will be functions of one variable, which we will typically write as $f(x)$, where both x and $f(x)$ take real values. It is standard practice nowadays to represent the real numbers as points on a straight line, however, it is important to mention at least once how does this construction actually happens.

As a geometric entity, the points on a straight line carry no particular label. This means that there is no particular point on the line that corresponds to the number 0 or 1 or any other number that you may think of. Therefore, the first step in the correspondence between the line and the real numbers is to *choose* a point of the line as the number 0. Traditionally the point chosen is labeled O, where O stands for origin.

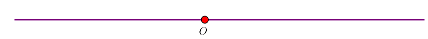


Figure 7: Choice of Origin for the Line

Once this has been done, we declare that the points to the right of 0 correspond to the positive numbers while points to the left of 0 correspond to the negative numbers. After doing this, we need to define the *addition* of two points on the straight line. If we call the points P and Q we define $P+Q$ considering the following cases:

Addition of Points: suppose that P and Q are two points on the straight line and that a choice of origin has already been made. To define $P+Q$ consider the following cases:

- P and Q both positive (that is, to the right of O): identify P and Q with the segments from O to P and from O to Q respectively. The concatenation of these two segments will start at O and end at some point which we identify with $P+Q$.
- P and Q both negative (that is, to the left of O): same procedure as before but now $P+Q$ will be to the left of O
- P and Q of different signs (assume P positive and Q negative for convenience): again we identify P and Q with segments and we compare them to see which one is longer. If P is longer than Q then $P+Q$ will be positive and equal (in length) to the difference between the two segments. If Q is longer than P then $P+Q$ will be negative and equal (in length) to the difference between the two segments.

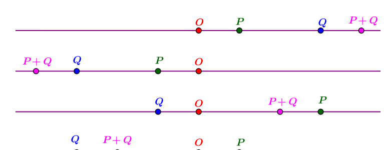


Figure 8: Addition of Points

The following figure illustrates the addition operation. The important thing to notice is that to define the addition between points we just need to identify a point as 0, we do not need to identify a point as 1.

Things become more complicated if we want to define multiplication between points. One plausible option might be the following: to define the addition between points we secretly identified the points with segments starting at 0, therefore, we might take PQ to mean the area of the rectangle whose segments are P and Q. However, there are two problems with this option.

First, the area of a figure is always positive but the multiplication of numbers of opposite signs is negative so it is not clear how to interpret this case as an area. The second objection is that in this construction we have to build a two-dimensional figure, namely, a rectangle, so it is not clear how we would use this rectangle to find a point on the line since we would like to multiply two points in order to get a new point, not a polygon. In fact, if we followed this method then we would be inclined to say that the product of three points PQR would be the volume of some box but then the product of four points $PQRS$ would need to be some sort of "hyper-volume" and so on.

Therefore, we will try to define PQ in such a way that we get a new point and that respects the usual sign tables. As we will see right away, in order to achieve this we need to *choose* (again arbitrarily) some point on the line as the number 1, that is, to define multiplication between points we need to choose a unit length. [11]

Multiplication of Points: Suppose that P and Q are two points on the line (which we call x axis for now). To define PQ do the following steps:

- Choose a y axis, that is, find the perpendicular line to the x axis going through the origin O
- Lay P along the y axis (respecting the sign of P) so that it has coordinates $(0, P)$
- Lay Q along the x axis (respecting the sign of Q) so that it has coordinates $(Q, 0)$
- Draw the segment between $(Q, 0)$ and $(0, 1)$
- Find the line l parallel to the previous segment which passes through the point $(0, P)$
- PQ will correspond to the point of intersection between the line l and the x axis

In most contexts in which Calculus is applied the quantities one deals with have *units*. For example, our variable x could have units of length (like meters, feet, miles), time (like seconds, minutes, hours), mass (like pounds, grams, kilograms), etc. Now, what does a choice of unit mean? It simply means that a standard has been chosen and it is given

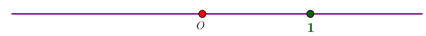


Figure 9: Choice of Origin and Unit Length

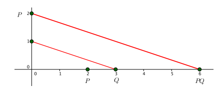


Figure 10: Product with $P = 2$, $Q = 3$

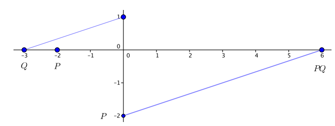


Figure 11: Product with $P = -2$, $Q = -3$

[11] It may not be clear why the following procedure actually corresponds to the multiplication of two numbers. The reason why it works depends on some basic Euclidean Geometry, namely, properties of similar triangles.

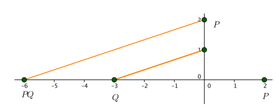


Figure 12: Product with $P = 2$, $Q = -3$

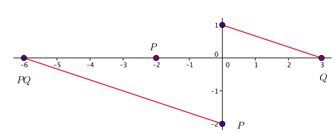


Figure 13: Product with $P = -2$, $Q = 3$

the value 1 by definition. For example, if our units are meters then what we could do is choose some bar and say that its length is *by definition* equal to 1 meter. Then any other length would be compared with this bar in order to determine the numerical value that will be assigned to its length. The point is that to define multiplication between points we had to choose a point as 1 but we didn't have to specify what were the *units* of 1, that is, we can interpret "1" as 1 meter, 1 hour, etc. Therefore, we can represent on our straight line position, area, volume, time, electric charge, mass and any other variable that we want.[12] This we be specially useful because of the following principle:

[12] Assuming that those variables can be modeled on the real numbers

Geometrization Principle: It is possible to translate physical problems into geometric problems

Geometrization Principle
Finding the Slope of a Line
Velocity, Acceleration
Marginal Utility, Marginal Cost
Rates of Change
Finding the Area under the Curve
Position, Velocity
Work, Average Value
"Continuous" Sum

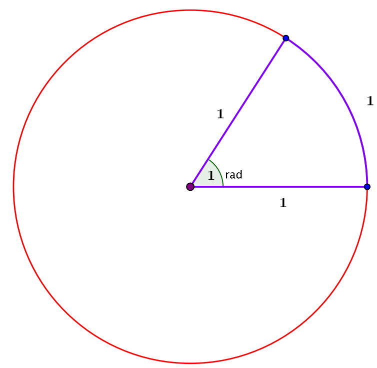


Figure 14: Measuring angles in radians

Another way to apply the geometrization principle is to justify why it is more convenient to work with radians instead of degrees. Suppose that we have chosen a unit length (which for convenience we will call meters) and we build a circle of radius 1 meter. It is typical to measure the angles of the circle either in terms of degrees or in terms in radians.

In order to represent the angles as points on the straight line we consider an angle to be positive if it has a counter-clockwise orientation and we consider it to be negative if it has a clockwise orientation. Since 1 rad corresponds to an arc of length 1 meter we can represent 1 rad as a segment with the same size as 1 meter. Since 1 degree corresponds to an arc of length $\frac{\pi}{180}$ meters we can represent 1 degree as $\frac{\pi}{180}$ meters, as the following figure shows.

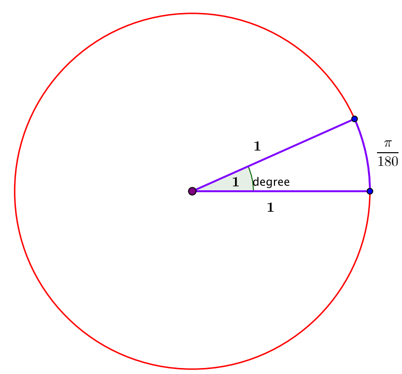


Figure 15: Measuring angles in degrees

It should be noted that if we use radians then the unit for angles coincides with the unit we choose to measure the length of intervals. Therefore, we can forget the distinction between both units and treat them as equals. If we used degrees instead then we would need to introduce always a scaling factor between our unit for angles and our unit for length.

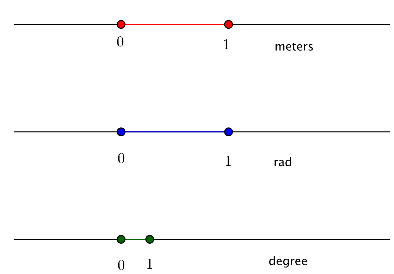


Figure 16: Meters, radians and degrees

Geometry of the Line and the Plane

Now that we have established a correspondence between the real numbers and the straight line we can do analytic geometry. As a matter of notation, the symbol $\mathbb{R}$ will represent the **real numbers** and we will often use the word real line when we want to make explicit our correspondence between numbers and points on the line.

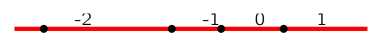


Figure 17: Real Line $\mathbb{R}$

Sometimes we will work not with the entire real line, but just with a portion of it. These parts of the real line are very special and they are knowns as **intervals**. There are eight different kinds of intervals:

1. Open interval (a, b): It consists of all the points between a and b, excluding a and b
2. Closed interval $[a, b]$: It consists of all the points between a and b, including a and b.
3. Half-open interval $(a, b]$: It consists of all the points between a and b, excluding a and including b
4. Half-open interval $[a, b)$: It consists of all the points between a and b, including a and excluding b
5. Open infinite interval (a, ∞): It consists of all the points to the right of a, excluding a
6. Open infinite interval $(-\infty, a)$: It consists of all the points to the left of a, excluding a
7. Closed interval $[a, \infty)$: It consists of all the points to the right of a, including a
8. Closed interval $(-\infty, a]$: It consists of all the points to the left of a, including a

It should be noted that **the symbols ∞, $-\infty$ do not represent numbers**, they are just used in order to have the same kind of notation for all the different types of intervals. Otherwise, if we avoided the symbol ∞, we would need to write instead of (a, ∞) something like "$(a$", which is not as nice as (a, ∞). Also, it might seem a little pedantic to distinguish between open and closed intervals, however, sometimes analyzing the behavior of a function depends on whether the interval is open or closed, so this distinction is a useful one as will be seen later.

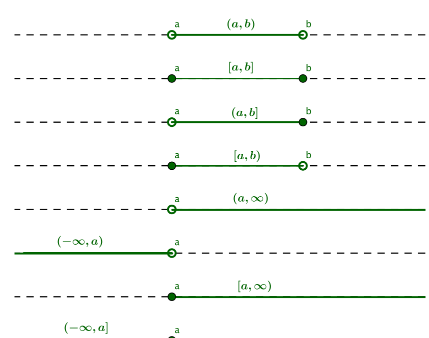


Figure 18: Types of Intervals

Open intervals will also play an important role once we study limits, in particular, if we fix a point p on the real line there is a special class of intervals associated to this point:

- If $p \in \mathbb{R}$ is a point on the real line, a **neighborhood** of p is any open interval containing p
- A **deleted neighborhood** of p is a neighborhood of p with the point p removed from it

For example, if our point is $p = 2$ then some neighborhoods of p are the intervals $(1, 3)$, $(-1, 4)$ and $(0, 3)$ while the corresponding deleted neighborhoods are $(1, 2) \cup (2, 3)$, $(-1, 2) \cup (2, 4)$ and $(0, 2) \cup (2, 3)$. On the other hand, an interval like $(5, 6)$ is not a neighborhood of 2.

Now that we have represented the real numbers as points on a line we can find the distance between them. Before doing this, it is useful to find the distance between a point and the origin. We will use the symbol $d(p, q)$ to denote the distance between the points p and q. Clearly we just need to consider two cases:

- p is to the right of 0, that is, p is a positive number. Then we define $d(p, 0) \equiv p$. For example, $d(3, 0) = 3$.
- p is to the left of 0, that is, p is a negative number. Then we define $d(p, 0) \equiv -p$. For example, $d(-4, 0) = -(-4) = 4$.

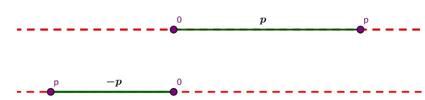


Figure 19: Absolute Value

The distance between the point p and the origin 0 is called the **absolute value of** p and it is generally denoted $|p|$. In other words,

$$|p| = \begin{cases} p & \text{if } p \geq 0 \\ -p & \text{if } p < 0 \end{cases} \tag{24}$$

The following are some properties of the absolute value:

Properties of the Absolute Value: If $p, q \in \mathbb{R}$ are two numbers on the real line then

$$|p| = \begin{cases} p & \text{if } p \geq 0 \\ -p & \text{if } p < 0 \end{cases} \tag{25}$$

- Triangle Inequality: $|p + q| \leq |p| + |q|$
- Symmetry of the absolute value: $|p| = |-p|$
- If $|p| \leq q$ then $-q \leq p \leq q$, that is, p must be inside the interval $[-q, q]$
- If $|p| \geq q$ then $p \geq q$ or $-p \geq q$, that is, p belongs to one of the intervals $[q, \infty)$ or $(-\infty, -q]$

Now that we have defined the absolute value, we can use it to calculate the distance between two points p and q on the real line. We denote the distance $d(p,q)$. For simplicity, we make a few cases:

1. Case p and q positive numbers: if $p \leq q$ then we have $d(p,q) = q - p$. If $q \leq p$ then $d(p,q) = q - p$. We observe that both cases can be written as $d(p,q) = |p - q|$

2. Case p and q negative numbers: we can use the following trick. Since p, q are negative numbers then $-p, -q$ are positive numbers and because $d(p,q) = d(-p,-q)$ we have by the first case that $d(p,q) = |-p-(-q)| = |-(p-q)| = |p-q|$

3. Case p and q are of different signs: for example, if p is positive and q is negative then $d(p,q) = d(p,0) + d(0,q) = p - q$. On the other hand, if q is positive and p is negative then $d(p,q) = d(p,0) + d(0,q) = -p + q$. Again, we see that both cases can be written as $|p-q|$. Therefore, we have found

If p, q are two points on the real line then the distance between p and q is

$$d(p,q) = |p-q| \tag{26}$$

The functions we will study are of one variable, and they can be represented by their graphs, which are curves in the plane with special properties. Therefore, it is useful to recall some facts about the plane.

We will denote the plane (which you can think as an infinite sheet of paper) by $\mathbb{R}^2$. In the same way as we did for the real line, we choose a point on the plane and call it the origin of the plane. This point is represented by $(0,0)$. Once we have chosen the origin, we take two perpendicular lines which intersect at $(0,0)$: one of the lines is called the x axis and the other line is called the y axis. This choice of an origin and two axes is called the **cartesian plane**. The idea is to think of each axis as a copy of $\mathbb{R}$, and use them to assign coordinates to each point p on the plane $\mathbb{R}^2$.

To be more precise, suppose that $p \in \mathbb{R}^2$, that is, p is a point on the plane. Then we can draw two perpendicular lines that pass through p and which intersect the x axis and y axis in a perpendicular way. The intersection of the lines with each respective axis gives a pair of numbers which are called the coordinates of point p. By an abuse of notation, we use the same name for the coordinates of the point as for the axis of the cartesian plane so we write $p = (x,y)$. Each of the four regions of the cartesian plane are called the four quadrants of the xy plane and are denoted as the following figure shows.

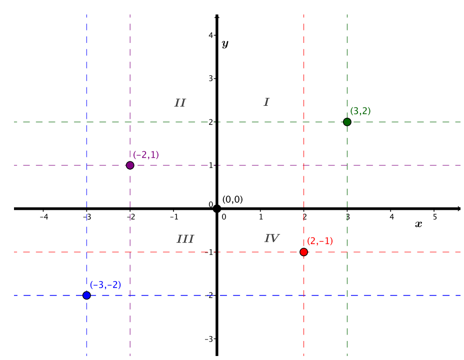


Figure 20: Cartesian plane xy

One of the advantages of using the cartesian plane is that it gives a way to treat algebraically many problems in geometry. For example, consider a straight line in the xy plane. Unless the line is the y axis or parallel to the y axis, it will intersect the y axis only at a point. We call this point $(0,b)$. Now, suppose that θ is the smallest angle between the x axis and the line as measured from the x axis to the line in the counterclockwise direction. Let (x, y) be any other point on the line.

By basic trigonometry we have that

$$\tan\theta = \frac{y-b}{x} \tag{27}$$

If we multiply the previous equation by x and send b to the other side of the equation we see that

$$y = x\tan\theta + b \tag{28}$$

Now, $\tan\theta$ is a very important property of the line known as the **slope** , and we use the letter m to denote it, that is

$$m = \tan\theta \tag{29}$$

In this way, we have found what is known as the **Slope-Intercept Form** of the line:

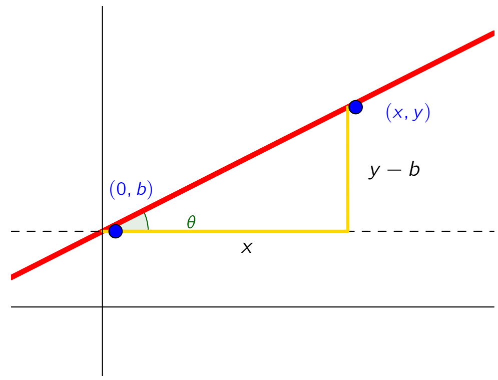


Figure 21: Equation of a Line

Slope-Intercept Form of the Line: Suppose that a straight line intersects the y-axis at the point $(0, b)$ and has a slope m. If (x, y) is any point on the line, we have

$$y = mx + b \tag{30}$$

The following properties are very easy to show so we will just quote them:

Algebraic properties of lines:

- Given two points (x_1, y_1), (x_2, y_2) on the same line l, the slope m of the line can be found as

  $$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{31}$$

  Observe that the slope can be positive, negative or zero.

- Two lines l_1, l_2 are parallel if and only if their respective slopes are the same, that is, $m_1 = m_2$

- Two lines l_1, l_2 are perpendicular if and only if the product of their slopes is -1 , that is $m_1 m_2 = -1$

- When the slope of a line is zero, that is, $m = 0$, we have that the line is horizontal and its equation is $y = b$ where b is the intersection with the y-axis

- When the line is parallel to the y-axis, it is a vertical line and its equation is $x = a$ where a is the intersection with the x-axis. The slope is not defined for vertical lines.

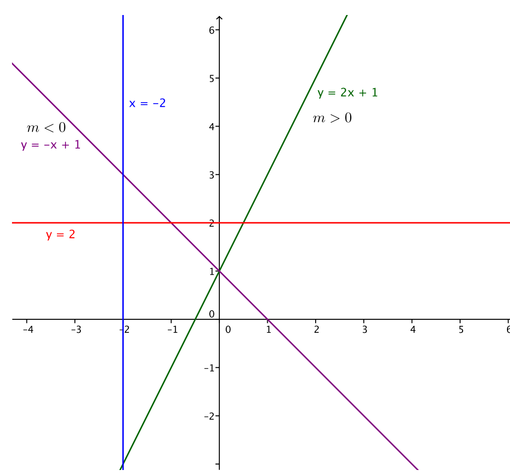


Figure 22: Lines with positive, negative, zero and undefined slopes

Example 1. **Find the slope of the line that passes through points $(4,5)$ and $(3,8)$**

We use formula 31 with $(x_1, y_1) = (4,5)$ and $(x_2, y_2) = (3,8)$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{8-5}{3-4} = -3 \tag{32}$$

It is useful to note that to find the slope it doesn't matter which point we call (x_1, y_1) and which point we call (x_2, y_2). For example, if we reverse the order and write $(x_1, y_1) = (3,8)$ and $(x_2, y_2) = (4,5)$ then by formula 31

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{5-8}{4-3} = -3 \tag{33}$$

Example 2. **Given the equation $y = 4x - 3$ answer the following questions: if x increases by 1 unit, what is the corresponding change in y? If x decreases by 2 units, what is the corresponding change in y?**

To do this problem it is useful to rewrite equation 31 in a different way. If (x_1, y_1), (x_2, y_2) are two points on a line then $y_2 - y_1$ represents the change in the vertical values between the points while $x_2 - x_1$ represents the change in the horizontal values between the points. Historically these changes in values are denoted $\triangle y$ and $\triangle x$, that is,

$$\begin{cases} \triangle y = y_2 - y_1 \\ \triangle x = x_2 - x_1 \end{cases} \tag{34}$$

Here it is important to note that $\triangle x$ or $\triangle y$ does not mean the symbol $\triangle$ times x or y, that is, it is not a multiplication between two variables. Rather, $\triangle x$ and $\triangle y$ are names for new variables and should be considered as a single symbol. Therefore, it is false that $\frac{\triangle y}{\triangle x} = \frac{y}{x}$ because $\triangle$ can't be separated from y or x. With this new notation we have found

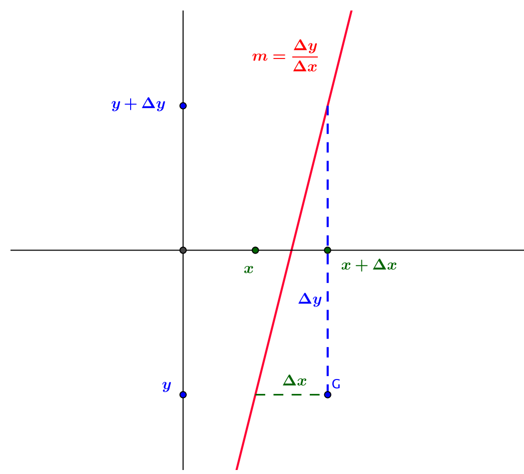


Figure 23: Equation Slope, Delta Notation

Equation for the slope, delta notation: Given two points (x_1, y_1), (x_2, y_2) on the same line l, the horizontal change between the points is denoted $\triangle x$ and the vertical change between the points is denoted $\triangle y$.

$$\begin{cases} \triangle x = x_2 - x_1 \\ \triangle y = y_2 - y_1 \end{cases} \tag{35}$$

With this notation, the slope for the line can be found as

$$m = \frac{\triangle y}{\triangle x} \tag{36}$$

which implies that **the slope of the line is the ratio of the vertical change with respect to the horizontal change.**

With the previous observation, the problem becomes easier to solve. In the first case, the increase of x is by 1 unit so $\triangle x = 1$. Comparing with 30 we can see that the slope for the line is $m = 4$ so by 36 we have

$$4 = \frac{\triangle y}{1} \tag{37}$$

which implies that $\triangle y = 4$. Therefore, the corresponding change in y is an increase of 4 units.

For the second case $\triangle x = -2$ (the negative is because there is a decrease) so we have again by 36 that

$$4 = \frac{\triangle y}{-2} \tag{38}$$

or $\triangle y = -8$. Therefore, there is a decrease of 8 units in y.

Example 3. Given the equation $2x + 3y = 4$, a) is the slope of the line described by this equation positive or negative? b) As x increases in value, does y increase or decrease? c) If x decreases by 2 units, what is the corresponding change in y?

a) In this case we have to rewrite first the equation of the line in the form 30 to identify the slope. Since $2x + 3y = 4$ can be written as $y = \frac{4-2x}{3}$ we see that the slope is $m = -\frac{2}{3}$ so the slope is negative.

b) If x increases in value we have that $\triangle x > 0$ and by equation 36 the only way for m to be negative is if $\triangle y$ is negative, so there is a decrease in y.

c) If $\triangle x = -2$ again by 36 we have

$$-\frac{2}{3} = \frac{\triangle y}{-2} \tag{39}$$

or $\triangle y = \frac{4}{3}$.

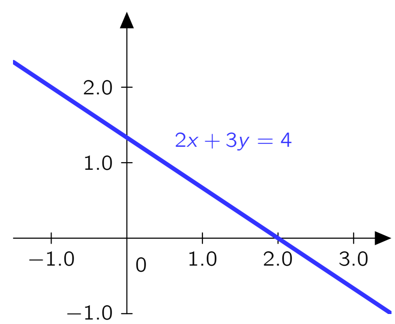


Figure 24: Line $2x + 3y = 4$

Example 4. Determine whether the line going through $A = (1, -2)$, $B = (-3, -10)$ is parallel or not to the line passing through $C = (1, 5)$ and $D = (-1, 1)$

We just need to see whether or not the slopes are the same. Since we have two points for each line we use the formula 31 to find the corresponding slopes. The slope for the first line is

$$m_1 = \frac{\triangle y}{\triangle x} = \frac{-10 - (-2)}{-3 - 1} = \frac{-8}{-4} = 2 \tag{40}$$

The slope for the second line is

$$m_2 = \frac{\triangle y}{\triangle x} = \frac{1 - 5}{-1 - 1} = \frac{-4}{-2} = 2 \tag{41}$$

Since the slopes are the same the lines are parallel.

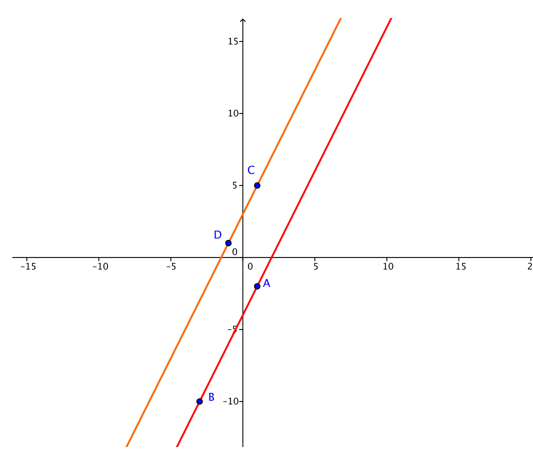


Figure 25: Parallel Lines

Example 5. **Determine whether the line going through $A = (2, 0)$, $B = (1, -2)$ is perpendicular or not to the line passing through $C = (4, 2)$ and $D = (-8, 4)$**

Again we find the corresponding slopes and see whether or not the product give -1. By 31 the first slope is

$$m_1 = \frac{\triangle y}{\triangle x} = \frac{-2-0}{1-2} = \frac{-2}{-1} = 2 \tag{42}$$

The second slope is

$$m_2 = \frac{\triangle y}{\triangle x} = \frac{4-2}{-8-4} = \frac{2}{-12} = -\frac{1}{6} \tag{43}$$

Since $m_1 m_2 = 2\left(-\frac{1}{6}\right) = -\frac{1}{3}$ is not -1 the lines are not perpendicular.

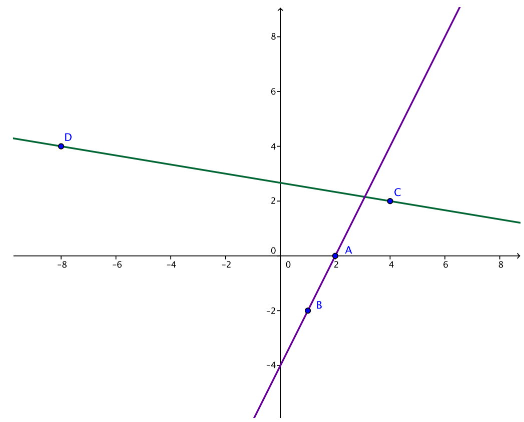


Figure 26: Non perpendicular lines

Example 6. **Find an equation of the line that passes through $(3, -4)$ and has slope $m = 2$**

Call any other point on the line (x, y). Because the slope is 2 we have by formula 31 that

$$2 = \frac{y-(-4)}{x-3} = \frac{y+4}{x-3} \tag{44}$$

Therefore

$$y = 2(x-3) - 4 \tag{45}$$

and the equation for the line is

$$y = 2x - 10 \tag{46}$$

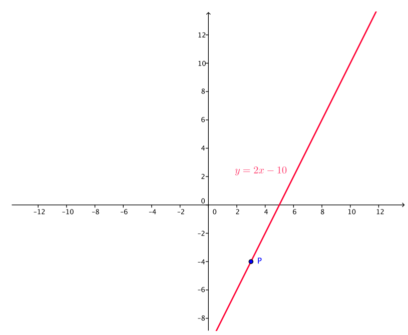


Figure 27: Line through $P = (3, -4)$ with slope 2

Finally, we can use Pythagora's Theorem to find the equation of a circle centered at (a, b) with radius r. As we can see in the figure to the right, if we choose any point (x, y) on the circle we can find a right triangle with legs $x - a$, $y - b$ and hypothenuse r so Pythagora's Theorem says that

$$r^2 = (x-a)^2 + (y-b)^2 \tag{47}$$

and this is the equation of a circle.

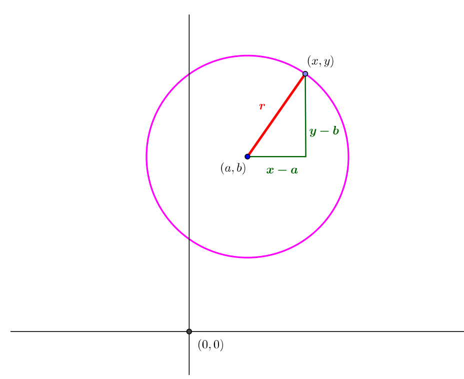


Figure 28: Equation of a circle centered at (a, b) with radius r

The equation of a circle centered at point (a, b) with radius r is

$$r^2 = (x-a)^2 + (y-b)^2 \tag{48}$$

Example 7. Find all possible coordinates of the points that are a distance of 10 units away from the origin and have a y-coordinate equal to -6

The distance of a point (x, y) to the origin is $\sqrt{x^2 + y^2}$ (this follows from Pythagora's Theorem). Therefore, we want $\sqrt{x^2 + y^2} = 10$ or $x^2 + y^2 = 100$. We want also that $y = -6$ so this gives $x^2 + 36 = 100$ or $x^2 = 64$ which gives $x = \pm 8$ so the two points are $(8, -6)$ and $(-8, -6)$.

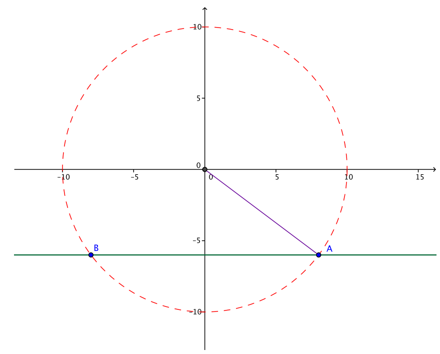


Figure 29: Intersection line $y = -6$ with circle $x^2 + y^2 = 100$

Functions

One objective of Calculus is to relate the changes between different quantities. In order to do this, it is necessary to specify first how these quantities are related, and this is done through a special kind of relationship know as functional relationship.

To be more concrete, suppose that an ecologist is studying a species over a specific period of time. There might be many characteristics that are important for this study, but certainly many would agree that analyzing the number of members of the species over time is among the first things that should be looked at. Therefore, the ecologist will start studying the population of the species, and this quantity will be denoted by P [13]. Clearly at any moment in time, the population has a fixed value but for practical purposes the population is measured only at certain periods, perhaps once every year. Therefore, if t denotes the time (measured in years) that has elapsed since the beginning of the research, the notation $P(t)$ will mean the population P measured at time t.[14]

For example, if $P(0) = 20$ and $P(1) = 30$ then the population had 20 members when the research started and it increased to 30 members after the first year. Therefore, the population P is in a specific relation to the elapsed time t, and the expression $P(t)$ stands for "P is a function of t".

In general, an expression like $f(x)$ means: "the quantity f is a function of the quantity x" and **the word function is used to express the fact that each value of x specifies uniquely a value of f.**

For example, $P(t)$ is called a function because at each particular time t there can only be one value for the population P. That is, we can't have something like $P(1) = 30$ and $P(1) = 40$ because this would imply that after one year the population had 30 and 40 members simultaneously, which is clearly impossible.

On the other hand, suppose that $I(n)$ represents the monthly income I of a family with n members. For example, $I(1)$ could represent the income of a single person and $I(2)$ the income of a married couple. Clearly different individuals have different incomes, so it makes perfect sense to have $I(1) = 2000$ and $I(1) = 3000$ simultaneously (because there are individuals that earn \$2000 and \$3000 a month), which implies that a particular value of n fails to specify uniquely the value of I so the relation $I(n)$ is not a function. Since we will only study functions,

[13] Here P stands for population

[14] Observe that we use a parentheses in $P(t)$ to distinguish it from Pt, which means multiplication between P and t

the notation $f(x)$, $P(t)$,etc will mean from now on that we have a relation that deserves the name of function and the other kinds of relations won't be mentioned again.[15]

Although a function like $P(t)$ is characterized by the fact that a value of t specifies only one value of P, this does not mean that a value of P can only be associated to one value of t. For example, if the value of the population P is 30, then there is nothing odd with having $P(1) = 30$ and $P(2) = 30$, this simply means that the population remained constant during the first and second years, so we can have two different times being paired with the same value for the population. Moreover, if we take the value of the population P to be 100, we might have that there is no time t such that $P(t) = 100$, that is, the population might never have 100 members. Therefore, **it is not required for every value of P to be achieved at some value t, while every value t must achieve some value of P**. This asymmetry is a generic property of functions.

Usually in the notation $P(t)$, the quantity P **is called the dependent variable** and the quantity t **is called the independent variable**. Since these names are very common, we will use them throughout the course, however it is important to observe that they are misleading since it suggests causality, when in fact **the concept of function is completely independent from any notion of cause and effect**. For example, the phrase "independent variable" usually refers to something that can be controlled in an experiment but for our function $P(t)$ certainly humans can't control the flow of time, and the fact that $P(1) = 30$ does not mean that time itself was the cause for the number of the population, instead the size of the population will be caused by many different factors and $P(t)$ simply means that at any time there is a well defined size for the population. Continuing with tradition, we will stick with these names despite the confusion they may cause so $P(t)$ will mean

$$\underset{\text{dependent variable}}{\underset{\downarrow}{P}}\overset{\text{independent variable}}{\overset{\uparrow}{(t)}} \tag{49}$$

The following vocabulary describes some basic properties of functions and will facilitate their study throughout the course.

[15] The moral of the story is that one should keep in mind that a function is a very specific kind of relationship between quantities and because the methods that will be developed throughout the course only apply to functions one should be verify first that one is working with a function

Suppose that $P(t)$ represents a function:

- P is the **dependent variable or output**
- t is the **independent variable or input**
- The **domain** of the function consists in the values that the input can take
- The **codomain** of the function consists in the values that the output P can take
- The **range** of the function consists in the values that the output P actually takes
- The function is called **injective** (or one-to-one) if different values of t specify different values of P
- The function is called **surjective** (or onto) if its codomain and range agree
- The function is called **bijective** (or invertible) if it is injective and surjective

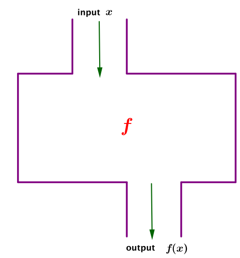


Figure 30: Function as a Machine

Now that we have introduced the concept of function, the next step is to devise strategies to understand them better. For example, let's continue with our function $P(t)$ where P is the population of the species and t is the time elapsed. If our ecologist made 5 measurements, a useful way to represent the data might be something like the table on the right.

t	$P(t)$
0	20
1	30
2	25
3	30
4	40

(50)

Table 4: Size of a Population

On the left column the values that the independent variable take are represented, and to the right of each value of t we write the value measured $P(t)$ for the population at that time.

In this particular example, only a finite number of measurements were made. However, in Calculus we are mostly interested in functions for which the domain of the function takes values on the entire real line or at least on some intervals of the real line. In the case of $P(t)$ this would imply having knowledge of the population at every moment in time between the beginning of the research and the four years it lasted. In practice this is certainly impossible, but we can imagine that instead of making measurements each year, the ecologist starts making measurements of the population every month, week or day, thereby reducing the interval of time between the distinct values. It is a very useful idealization from a mathematical perspective to assume that somehow we have knowledge of $P(t)$ for every t in an entire interval because in this way the techniques of calculus can be applied.

Since we will assume that our functions are defined over the real line or an interval of the real line we can use a visual representation for the function called the graph of the function.

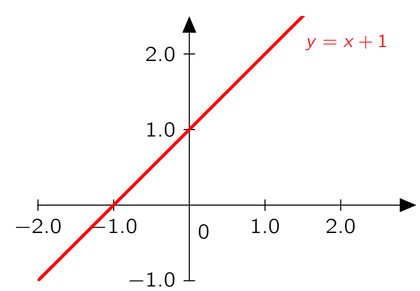
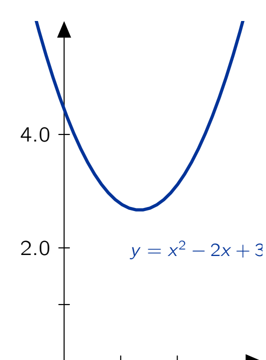


Figure 31: Function $y = f(x) = x + 1$



For this we switch notations and write $f(x)$ for the function. The graph of the function is represented on the xy plane and it is constructed as follows:

1. Pick a value x of the independent variable.
2. Find the corresponding value $f(x)$
3. Call $y = f(x)$ and indicate the point $(x, y) = (x, f(x))$ in the xy plane.
4. Repeat this procedure for every possible value of x and plot all the points $(x, f(x))$ in the xy plane. This will produce a curve in the xy plane called the graph of the function $f(x)$, which is sometimes denoted for emphasis as $y = f(x)$.

The figures on the right are some of the functions that will be studied throughout the course.

In terms of its graph of a function we can observe the following:

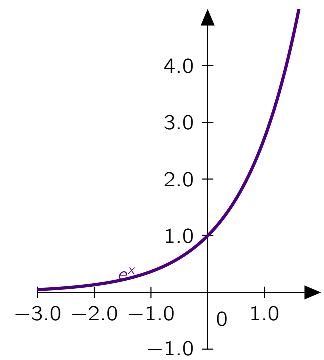


Figure 33: Function $y = f(x) = e^x$

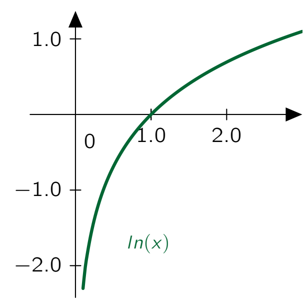


Figure 34: Function $y = f(x) = \ln x$

For a function $y = f(x)$ we have:

- The values of the independent variable are represented in the x axis
- The values of the dependent variable are represented in the y axis
- Vertical Line Test: a vertical line can intersect the graph of the function *at most* one point
- Horizontal Line Test: a function is injective if and only if a horizontal line can intersect the graph of the function at most one point

---

Example 8. Determine which of the following curves can be the graph of a function

The circle is not the graph of any function because it fails the vertical line test: there are plenty of vertical lines which intersect the circle at more than one point

The weird curve on the right is a function because it satisfies the vertical line test: observe that at the point in which there could be two possible values the dots indicates which is the value that is under consideration. This shows in particular that a curve can be the graph of a function even if it has "jumps"

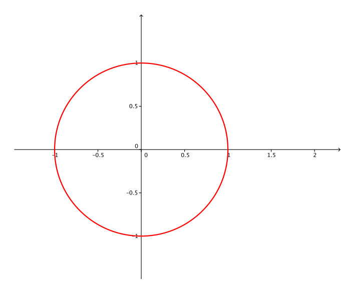
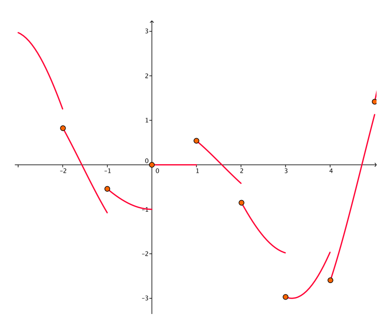


Figure 35: Circle



Just as we can add, subtract, divide, multiply numbers we can define similar operations for functions in a natural way. From a practical point of view, the major algebraic manipulations between functions work the same as they did with numbers, so there is no point in spending a lot of time in this, we will just mention the most important operations and how they look in terms of the graphs of the corresponding functions.

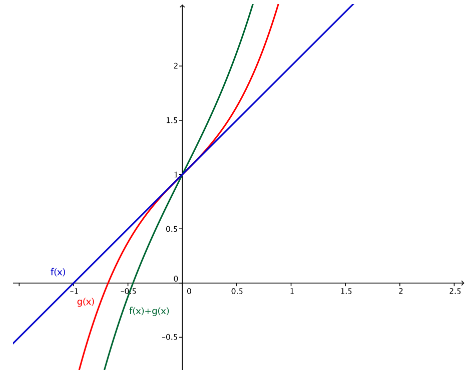


Figure 37: Sum of Functions $f(x) = x + 1$, $g(x) = x^3 + x + 1$, $f(x) + g(x) = x^3 + 2x + 1$

Given two functions $f(x)$ and $g(x)$ it is possible to perform the following algebraic operations:

- Add (or subtract) the functions to produce a new function denoted $f(x) + g(x)$ (or $f(x) - g(x)$)
- Multiply the functions to produce a new function denoted $f(x)g(x)$
- If $g(x)$ is not zero then we can divide the functions to produce a new function denoted $\frac{f(x)}{g(x)}$

Example 9. Let $f(x)$ be a function with domain $[-3, \infty)$ and $g(x)$ a function with domain $(-\infty, 6]$. Find the domain of $f(x) + g(x)$

In order to evaluate $f(x) + g(x)$ we need to evaluate $f(x)$ and $g(x)$ simultaneously. Now, this implies that the domain of $f(x) + g(x)$ is the intersection of the domains of $f(x)$ and $g(x)$. The domain of $f(x)$ consists in the points x with $-3 \leq x$ and the domain of $g(x)$ consists of the points x such that $x \leq 6$. Therefore, the intersection consists in the points x such that $-3 \leq x \leq 6$ which is the closed interval $[-3, 6]$

Suppose that x is a real number and that n is a positive integer, that is, n can be 1, 2,3, etc. Then we can start taking powers of x, that is, multiply x with itself n times. We denote this operation as x^n. In other words

$$x^n \equiv \underbrace{x \cdot x \cdot \dots \cdot x}_{n-\text{times}} \tag{51}$$

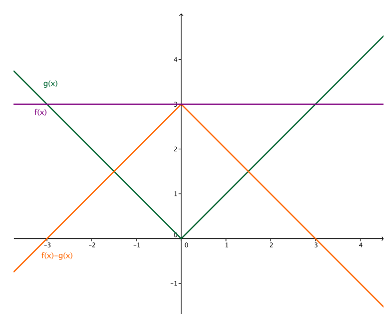


Figure 38: Subtraction functions $f(x) = 3$, $g(x) = |x|$, $f(x) - g(x) = 3 - |x|$

For example, $5^3 = 5 \cdot 5 \cdot 5 = 25 \cdot 5 = 125$. Now, we would also like to define x^0 and x^{-n}. In order to avoid making a long detour we will just give the definition for these two new cases:

$$x^0 \equiv 1 \tag{52}$$

$$x^{-n} \equiv \frac{1}{x^n} \qquad x \neq 0 \tag{53}$$

Moreover, we can define $x^{\frac{1}{n}}$ in the following way: $x^{\frac{1}{n}}$ is the number that satisfies $\left(x^{\frac{1}{n}}\right)^n = x$. It is important to be careful with this definition: for example, $(-2)^{\frac{1}{2}}$ would have to be a number such that $\left((-2)^{\frac{1}{2}}\right)^2 = -2$. However, the square of any number is always positive so the previous equation is contradictory. Also, $x^{\frac{1}{n}}$ might have more than one candidates, for example, 3 and -3 both satisfy

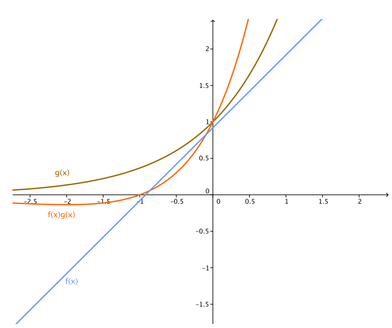


Figure 39: Product of functions $f(x) = x + 1$, $g(x) = e^x$, $f(x)g(x) = xe^x + e^x$

$3^2 = (-3)^2 = 9$ so we would have the choice $9^{\frac{1}{2}} = 3$ or $9^{\frac{1}{2}} = -3$. Fortunately, these are the only things that can go wrong with our definition:

- If n is an even integer and $x \geq 0$ we can define the n-th root of x, denoted $x^{\frac{1}{n}}$ or $\sqrt[n]{x}$ to be the number with the property that $(\sqrt[n]{x})^n = x$. Moreover, $\sqrt[n]{x}$ will always be taken to be positive.
- If n is an odd integer and x *any* real number, we can the n-th root of x, denoted $x^{\frac{1}{n}}$ or $\sqrt[n]{x}$ to be the number with the property that $(\sqrt[n]{x})^n = x$. Moreover, $\sqrt[n]{x}$ is positive if x is positive and $\sqrt[n]{x}$ is negative if x is negative.

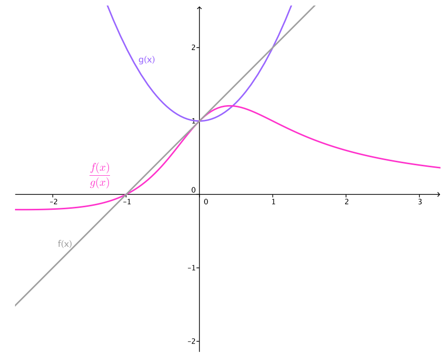


Figure 40: Division of Functions $f(x) = x+1$, $g(x) = x^2+1$, $\frac{f(x)}{g(x)} = \frac{x+1}{x^2+1}$

$$\begin{gathered} a^m \cdot a^n = a^{m+n} \\ \frac{a^m}{a^n} = a^{m-n} \\ (a^m)^n = a^{mn} \\ (ab)^n = a^n b^n \\ \left(\frac{a}{b}\right)^n = \frac{a^n}{b^n} \\ a^{\frac{m}{n}} = \sqrt[n]{a^m} = (\sqrt[n]{a})^m \\ a^{-\frac{m}{n}} = \frac{1}{a^{\frac{m}{n}}} \end{gathered} \tag{54}$$

Table 5: Laws of Exponents

Example 10. Find the domain of the function $s(t) = \frac{\sqrt{t-1}}{t^2-2t-3}$

To find the domain of $s(t)$ we need to determine when does the expression $s(t)$ make sense. To determine this we first focus on the numerator

$$\sqrt{t-1} \tag{55}$$

Since the square root of a number makes sense only when the number is positive or zero this means that

$$\sqrt{t-1} \text{ requires } t-1 \geq 0 \tag{56}$$

On the other hand, the denominator can be problematic whenever it becomes zero, so we need to solve the equation

$$t^2 - 2t - 3 = 0 \tag{57}$$

we can factorize the polynomial as

$$(t-3)(t+1) = 0 \tag{58}$$

which gives the values for

$$t = 3 \text{ or } t = -1 \tag{59}$$

Therefore, t must be different from 3 and -1 and also must satisfy $t \geq 1$ which implies that the domain of $s(t)$ is $[1, \infty) - \{3\}$.

Example 11. Find the domain of $h(t) = \frac{t^2-4t-21}{t^4+81} - \sqrt{200+2x}$

We will determine when does $\frac{t^2-4t-21}{t^4+81}$ and $\sqrt{200+2x}$ make sense separately. The numerator $t^2-4t-21$ always makes sense and the denominator t^4+81 can have a problem only when the denominator becomes zero, however, t^4+81 is always positive so in fact $\frac{t^2-4t-21}{t^4+81}$ is well defined on the entire real line.

On the other hand, $\sqrt{200+2x}$ makes sense only when $200+2x \geq 0$ which gives $100+x \geq 0$ or $x \geq -100$. Therefore, the domain of $h(t)$ is the interval $[-100, \infty)$

Polynomials and Rational Functions

A polynomial is a function $p(x)$ which can be written as sums of powers of x multiplied by some coefficients. Example of polynomials are $p_1(x) = x^2 + x + 3$, $p_2(x) = x^4 - 3x$, $p_3(x) = -\sqrt{2}x^3 + 2x + 1$, etc.

Perhaps the most important polynomials for this course are the quadratic polynomials (or parabolas) ,that is, polynomials of the form

$$p(x) = ax^2 + bx + c \tag{60}$$

where $a \neq 0$ and b, c are real numbers. If $a > 0$ the parabola 60 is called concave up , if $a < 0$ the parabola 60 is called concave down.

To analyze polynomials effectively the following identities are useful:

- **Difference of squares:**

$$a^2 - b^2 = (a - b)(a + b) \tag{61}$$

- **Binomial Formulas:**

$$\begin{cases} (a+b)^2 = a^2 + 2ab + b^2 \\ (a-b)^2 = a^2 - 2ab + b^2 \end{cases} \tag{62}$$

- **Sum of cubes:**

$$a^3 + b^3 = (a + b)\left(a^2 - ab + b^2\right) \tag{63}$$

- **Difference of cubes:**

$$a^3 - b^3 = (a - b)\left(a^2 + ab + b^2\right) \tag{64}$$

With the previous identities, we can write $ax^2 + bx + c$ in a conve-

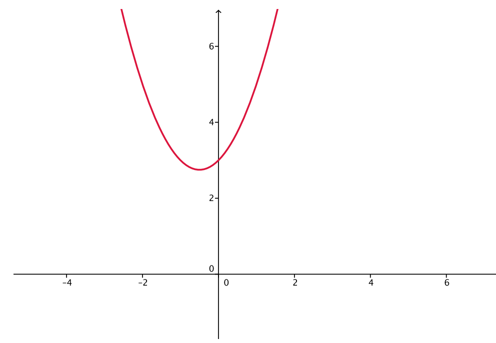


Figure 41: Polynomial $y = x^2 + x + 3$

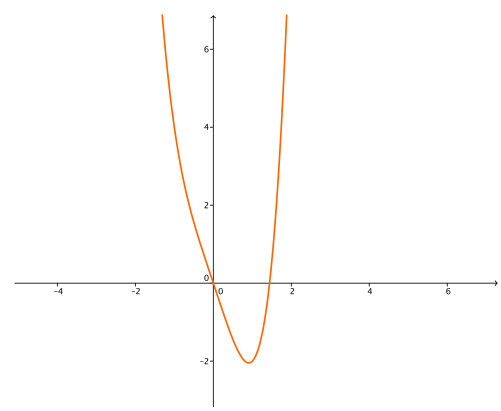


Figure 42: Polynomial $y = x^4 - 3x$

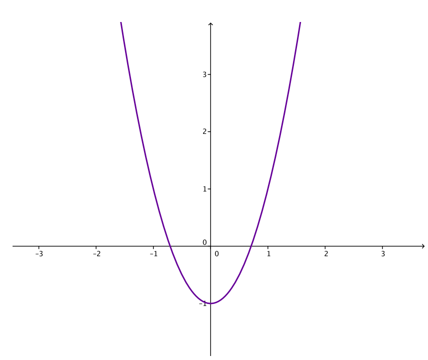


Figure 43: Concave up parabola $y = 2x^2 - 1$

nient form. First of all, it is clear that

$$ax^2+bx+c = ax^2+bx+c+\underbrace{\frac{b^2}{4a}-\frac{b^2}{4a}}_{0} = ax^2+bx+\frac{b^2}{4a}+c-\frac{b^2}{4a} = \frac{4a^2x^2+4abx+b^2}{4a}+c-\frac{b^2}{4a} \quad (65)$$

By the binomial identity we can write

$$4a^2x^2+4abx+b^2 = (2ax+b)^2 \quad (66)$$

and substituting in 65 we have that

$$ax^2+bx+c = \frac{(2ax+b)^2}{4a}+c-\frac{b^2}{4a} = \frac{(2ax+b)^2+4ac-b^2}{4a} \quad (67)$$

The technique of adding and subtracting $\frac{b^2}{4a}$ is known as completing squares. [16]

[16] If you want to see why it is called "completing squares" click here.

Completing squares: If you have the polynomial $ax^2 + bx + c$ then adding and subtracting $\frac{b^2}{4a}$ the polynomial becomes

$$a^2x+bx+c = \frac{(2ax+b)^2+4ac-b^2}{4a} \quad (68)$$

Why is it useful to complete squares? Suppose that we have the polynomial $p(x) = ax^2+bx+c$. We know that the graph of the polynomial is a curve and in the previous examples of polynomials we have seen cases in which the graph intersects the x axis and cases in which it does not. The values for which the curve intersects the x axis are known as roots of the polynomial and they must satisfy $p(x) = 0$, that is, $ax^2+bx+c = 0$. To see what conditions a quadratic polynomial must satisfy so that it can have roots, we can complete squares and write $p(x)$ as

$$p(x) = \frac{(2ax+b)^2+4ac-b^2}{4a} \quad (69)$$

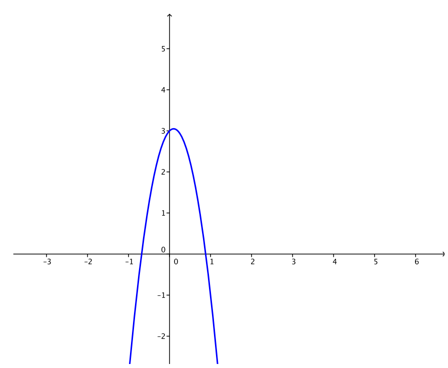


Figure 44: Concave down parabola $y=-5x^2+x+3$

Now, $p(x)$ has a root if and only if $p(x)=0$, which means that

$$0 = \frac{(2ax+b)^2+4ac-b^2}{4a} \quad (70)$$

multiplying both equations by $4a$ and sending $4ac-b^2$ to the other side of the equation we arrive at

$$(2ax+b)^2 = b^2-4ac \quad (71)$$

Following tradition b^2-4ac is called the discriminant of the polynomial and denote it by $\triangle = b^2-4ac$ [17]. Analyzing the cases in which the discriminant is negative, zero or positive we obtain the following result:

[17] This is an unfortunate notation for Calculus since $\triangle$ will reappear for completely different purposes but it hardly causes confusion

Roots of the quadratic polynomial $p(x) = ax^2 + bx + c$: a root of $p(x)$ is a value of x such that $p(x) = 0$. To find if such value exists or not, we compute the discriminant **discriminant** $\triangle = b^2 - 4ac$ and make the following cases:

- Case **discriminant** $\triangle < 0$: looking at 71 we see that there can be no roots because $(2ax+b)^2$ is never negative. Therefore, the graph of $p(x)$ is always above or below the x axis.
- Case **discriminant** $\triangle = 0$: looking at 71 we see that we will have $2ax + b = 0$ so there is only one root which we call
$$r = -\frac{b}{2a} \tag{72}$$
Moreover, we can factorize $ax^2 + bx + c$ as
$$ax^2 + bx + c = a(x-r)^2 = a\left(x + \frac{b}{2a}\right)^2 \tag{73}$$
- Case **discriminant** $\triangle > 0$: looking at 71 it can be seen that there are two roots which we call
$$r_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad r_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{74}$$
Moreover, we can factorize $ax^2 + bx + c$ as
$$ax^2 + bx + c = a(x - r_1)(x - r_2) \tag{75}$$

Example 12. Simplify $\frac{2a^2-3ab-9b^2}{2ab^2+3b^3}$
We need to factorize the numerator and denominator hoping that there will be some cancelations. The denominator seems more easy to work with since we can factorize b^2
$$2ab^2 + 3b^3 = b^2(2a + 3b) \tag{76}$$
Now, it would be useful if $2a + 3b$ appears in the factorization of $2a^2 - 3ab - 9b^2$. By inspection we can see that
$$2a^2 - 3ab - 9b^2 = (2a + 3b)(a - 3b) \tag{77}$$
so
$$\frac{2a^2 - 3ab - 9b^2}{2ab^2 + 3b^3} = \frac{(2a+3b)(a-3b)}{b^2(2a+3b)} = \frac{a-3b}{b^2} \tag{78}$$

Example 13. Simplify $\frac{x^3+2x^2-3x}{-2x^2-x+3}$
In the numerator we can factorize an x
$$x^3 + 2x^2 - 3x = x\left(x^2 + 2x - 3\right) \tag{79}$$

Again, by the method of inspection (or by finding the roots with the formula 74) we see that

$$x(x^2+2x-3)=x\,(x+3)\,(x-1) \tag{80}$$

Similarly, by inspection we can see that

$$-2x^2-x+3=(2x+3)\,(-x+1)=-(2x+3)(x-1) \tag{81}$$

In this way

$$\frac{x^3+2x^2-3x}{-2x^2-x+3}=\frac{x(x+3)(x-1)}{-(2x+3)(x-1)}=-\frac{x(x+3)}{2x+3} \tag{82}$$

Example 14. Simplify $\frac{4}{x^2-9}-\frac{5}{x^2-6x+9}$

By the formula for difference of squares we can see that the first term is

$$\frac{4}{x^2-9}=\frac{4}{x^2-3^2}=\frac{4}{(x-3)(x+3)} \tag{83}$$

Also, by the formula for the binomial

$$\frac{5}{x^2-6x+9}=\frac{5}{(x-3)^2} \tag{84}$$

In this way we can factorize $x-3$ in the numerator

$$\frac{4}{x^2-9}-\frac{5}{x^2-6x+9}=\frac{4}{(x-3)(x+3)}-\frac{5}{(x-3)^2}=\frac{1}{x-3}\left(\frac{4}{x+3}-\frac{5}{x-3}\right) \tag{85}$$

Now we use $(x-3)(x+3)$ as common denominator

$$\frac{1}{x-3}\left(\frac{4}{x+3}-\frac{5}{x-3}\right)=\frac{1}{x-3}\left(\frac{4(x-3)-5(x+3)}{(x-3)(x+3)}\right)=\frac{1}{x-3}\left(\frac{4x-12-5x-15}{(x-3)(x+3)}\right)=\frac{1}{x-3}\left(\frac{-x-27}{(x-3)(x+3)}\right) \tag{86}$$

The numerator is $-x-27=-\,(x+27)$ therefore the expression simplifies into

$$\frac{1}{x-3}\left(\frac{-(x+27)}{(x-3)(x+3)}\right)=-\frac{x+27}{(x-3)^2\,(x+3)} \tag{87}$$

Example 15. Simplify $\frac{\frac{1}{x}+\frac{1}{y}}{1-\frac{1}{xy}}$

First we work with the numerator

$$\frac{1}{x}+\frac{1}{y}=\frac{y+x}{xy} \tag{88}$$

As for the denominator we have

$$1-\frac{1}{xy}=\frac{xy-1}{xy} \tag{89}$$

In this way

$$\frac{\frac{1}{x}+\frac{1}{y}}{1-\frac{1}{xy}}=\frac{\frac{y+x}{xy}}{\frac{xy-1}{xy}}=\frac{xy\,(y+x)}{xy\,(xy-1)}=\frac{x+y}{xy-1} \tag{90}$$

Example 16. Rationalize the denominator of $\frac{\sqrt{a}+\sqrt{b}}{\sqrt{a}-\sqrt{b}}$

The purpose of rationalizing the denominator is to eliminate the square roots that appear in it. To do this we try to use the difference of squares $a^2 - b^2 = (a-b)(a+b)$. The version we are interested in is

$$a - b = \left(\sqrt{a} - \sqrt{b}\right)\left(\sqrt{a} + \sqrt{b}\right) \tag{91}$$

. Therefore, to eliminate $\sqrt{a} - \sqrt{b}$ we would need to multiply it by $\sqrt{a} + \sqrt{b}$. In this way

$$\frac{\sqrt{a}+\sqrt{b}}{\sqrt{a}-\sqrt{b}} = \frac{\sqrt{a}+\sqrt{b}}{\sqrt{a}-\sqrt{b}} \cdot \frac{\sqrt{a}+\sqrt{b}}{\sqrt{a}+\sqrt{b}} = \frac{\left(\sqrt{a}+\sqrt{b}\right)^2}{a-b} = \frac{a+b+2\sqrt{ab}}{a-b} \tag{92}$$

Example 17. Find the values of x that satisfy the inequalities $-6 < x - 2 < 4$

In this example and the next ones we need to remember the following key properties of the inequalities:

Properties of inequalities: suppose that $a < b$ and c is a real number.

- Addition preserves inequalities: if $a < b$ then $a + c < b + c$
- Subtraction preserves inequalities: if $a < b$ then $a - c < b - c$
- Multiplication and division by a positive number preserves inequalities: if $a < b$ and $c > 0$ then $ac < bc$ and $\frac{a}{c} < \frac{b}{c}$
- Multiplication and division by a negative number reverses inequalities: if $a < b$ and $c < 0$ then $ac > bc$ and $\frac{a}{c} > \frac{b}{c}$

With the previous properties it is very easy to solve the inequality $-6 < x - 2 < 4$. We can add 2 to every side of the inequality while preserving it

$$-6 + 2 < x - 2 + 2 < 4 + 2 \tag{93}$$

which gives

$$-4 < x < 6 \tag{94}$$

therefore x must belong to the interval $(-4, 6)$

Example 18. Find the values of x that satisfy the inequality $(2x-3)(x-1) \geq 0$

To analyze this inequality we remember that the product of two positive numbers or two negative numbers is positive while the product of a positive number and a negative number is negative.

Therefore, $(2x-3)(x-1) \geq 0$ can happen in two different ways:

- $2x - 3 \geq 0$ and $x - 1 \geq 0$: the inequality $2x - 3 \geq 0$ is the same as $2x \geq 3$ or $x \geq \frac{3}{2}$. The inequality $x - 1 \geq 0$ is the same as $x \geq 1$. Since x must satisfy both inequalities at the same time we conclude that $x \geq \frac{3}{2}$
- $2x - 3 \leq 0$ and $x - 1 \leq 0$: the inequality $2x - 3 \leq 0$ is the same as $2x \leq 3$ or $x \leq \frac{3}{2}$. The inequality $x - 1 \leq 0$ is the same as $x \leq 1$. Since x must satisfy both inequalities at the same time we conclude that $x \leq 1$

So x must belong to the interval $[\frac{3}{2}, \infty)$ or the interval $(-\infty, 1]$

Example 19. Find the values of x that satisfy the inequality $\frac{2x-1}{x+2} \leq 4$
First we subtract 4 on both sides of the inequality to have

$$\frac{2x-1}{x+2} - 4 \leq 0 \tag{95}$$

using common denominator $x + 2$ we have

$$\frac{2x-1-4(x+2)}{x+2} \leq 0 \tag{96}$$

which is the same as

$$\frac{-2x-9}{x+2} \leq 0 \tag{97}$$

multiplying both sides by -1 we reverse the inequality

$$\frac{2x-9}{x+2} \geq 0 \tag{98}$$

Again we have the following cases:

- $2x - 9 \geq 0$ and $x + 2 > 0$: the inequality $2x - 9 \geq 0$ gives $2x \geq 9$ or $x \geq \frac{9}{2}$. The inequality $x + 2 > 0$ gives $x > -2$. Therefore we must have $x \geq \frac{9}{2}$
- $2x - 9 \leq 0$ and $x + 2 < 0$: we just need to reverse the previous inequalities to see that $x \leq \frac{9}{2}$ and $x < -2$. Therefore we must have $x < -2$ (observe that $x = -2$ is not allowed because the denominator would be undefined in this case).

So x must belong to the interval $[\frac{9}{2}, \infty)$ or the interval $(-\infty, -2)$.

Example 20. Factorize $x^4 - 16$ and $x^3 + 216$
For $x^4 - 16$ we use the difference of squares

$$x^4 - 16 = \left(x^2\right)^2 - 4^2 = \left(x^2 - 4\right)\left(x^2 + 4\right) = (x-2)(x+2)\left(x^2 + 4\right) \tag{99}$$

For $x^3 + 216$ we use the sum of cubes

$$x^3 + 216 = x^3 + (6)^3 = (x+6)\left(x^2 - 6x + 36\right) \tag{100}$$

Example 21. Solve $|7-2x| < 9$

By the properties of the absolute value $|7-2x| < 9$ is the same as the inequalities

$$-9 < 7-2x < 9 \tag{101}$$

Subtracting 7 on all sides we have

$$-16 < -2x < 2 \tag{102}$$

dividing by -2 we have

$$8 > x > -1 \tag{103}$$

so x belongs to the interval $(-1, 8)$

In case we want to solve inequalities involving polynomials or quotient of polynomials there is a more direct method to do it.[18]

[18] To justify this method we need to use the intermediate value theorem which will be explained later in the course

—To solve the inequality $p(x) \geq 0$, $p(x) > 0$, $p(x) \leq 0$ or $p(x) < 0$ where $p(x)$ is a polynomial do the following steps:

1. find the roots of the polynomial $p(x)$
2. draw a picture of the real line and indicate the roots of $p(x)$ on the line. The line will be divided into intervals whose endpoints are the roots of $p(x)$
3. evaluate $p(x)$ in an arbitrary point of each interval to find the sign of $p(x)$

—To solve the inequality $\frac{p(x)}{q(x)} \geq 0$, $\frac{p(x)}{q(x)} > 0$, $\frac{p(x)}{q(x)} \leq 0$ or $\frac{p(x)}{q(x)} < 0$ where $p(x)$, $q(x)$ are polynomials do the following steps:

1. find the roots of the polynomial $p(x)$ and $q(x)$
2. draw a picture of the real line and indicate the roots of $p(x)$ and $q(x)$ on the line. The line will be divided into intervals whose endpoints are the roots of $p(x)$ and $q(x)$. However, since the roots of $q(x)$ do not belong to the domain of $\frac{p(x)}{q(x)}$ use an uncolored dot to indicate the zeros of $q(x)$
3. evaluate $\frac{p(x)}{q(x)}$ in an arbitrary point of each interval to find the sign of $\frac{p(x)}{q(x)}$

Example 22. Solve the inequality $x^2 - 8x - 9 \geq 0$

Call $p(x) = x^2 - 8x - 9$. We follow the previous steps:

1) Since

$$x^2 - 8x - 9 = (x-9)(x+1) \tag{104}$$

Figure 45: Solving $x^2 - 8x - 9 \geq 0$

the roots of $p(x)$ are 9 and -1.

2) We indicate the roots of the polynomial as the figure shows. We get the intervals $(-\infty,-1)$, $(-1,9)$ and $(9,\infty)$.

3) We evaluate $p(x)$ on each interval. We find that $p(x)$ is positive on $(-\infty,-1)$, $(9,\infty)$ and $p(x)$ is negative on $(-1,9)$. Since the inequality is greater or *equal* to 0 we also include the roots of $p(x)$ so x must belong to $(-\infty,-1]\cup[9,\infty)$.

Example 23. Solve the inequality $\frac{x^2-2x-3}{x+7}<0$

Call $p(x)=x^2-2x-3$ and $q(x)=x+7$. We follow the previous steps:

1) Since

$$x^2-2x-3=(x-3)(x+1) \tag{105}$$

the roots of $p(x)$ are 3 and -1 and the root of $q(x)$ is -7.

2) We indicate the roots of the polynomials as the figure shows. We get the intervals $(-\infty,-7)$, $(-7,-1)$, $(-1,3)$ and $(3,\infty)$.

3) We evaluate $\frac{p(x)}{q(x)}$ on each interval. We find that $\frac{p(x)}{q(x)}$ is positive on$(-7,-1)$, $(3,\infty)$ and $p(x)$ is negative on $(-\infty,-7)$, $(-1,3)$. Since the inequality is *strictly* smaller than 0 we exclude the roots of $p(x)$ so x must belong to $(-\infty,-7)\cup(-1,3)$.

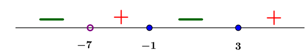


Figure 46: Solving $\frac{x^2-2x-3}{x+7}<0$

Trigonometric Functions

Another family of functions that will be extremely important are the trigonometric functions. The reason why trigonometric functions are so important is that they are used to study periodic phenomena, for example, the motion of a pendulum, a spring, pressure waves moving through air, electromagnetic waves, etc.

There are two ways in which the trigonometric functions can be introduced: by triangle trigonometry and by circle trigonometry. If we use triangle trigonometry we start with an acute angle θ [19] and consider a right triangle with angle θ.

[19] As explained before we assume θ is measured in radians

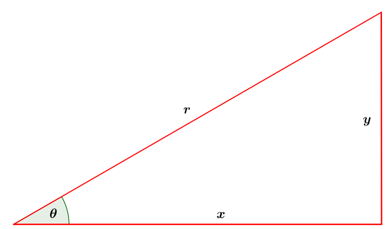


Figure 47: Right triangle with angle θ

The trigonometric functions are defined as

$$\begin{array}{ccc} \cos\theta \equiv \frac{x}{r} & \sin\theta \equiv \frac{y}{r} & \tan\theta \equiv \frac{y}{x} \\ \sec\theta \equiv \frac{r}{x} & \csc\theta \equiv \frac{r}{y} & \cot\theta \equiv \frac{x}{y} \end{array} \tag{106}$$

If we use circle trigonometry the trigonometric functions can be represented as the following figure shows.

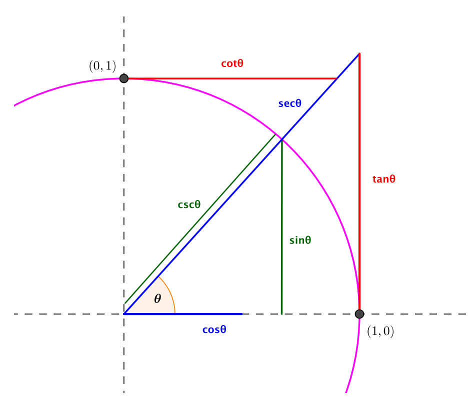


Figure 48: Trigonometric Relationships in the Unit Circle

As we will see later, the calculus techniques will allow us to find the graphs of the trigonometric functions. For now, it suffices to recall some of their most important properties:

Properties of Sine and Cosine:

- $\cos^2\theta + \sin^2\theta = 1$
- $-1 \leq \sin\theta \leq 1$ and $-1 \leq \cos\theta \leq 1$
- $\cos(\theta) = \cos(-\theta)$, $\sin(\theta) = -\sin(-\theta)$
- $\cos(\theta) = 0$ if and only if $\theta = (2k+1)\frac{\pi}{2}$ where k is any integer
- $\sin(\theta) = 0$ if and only if $\theta = k\pi$ where k is any integer

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\ \cos(\alpha-\beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta \\ \sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta \\ \sin(\alpha-\beta) &= \sin\alpha\cos\beta - \cos\alpha\sin\beta \end{aligned} \tag{107}$$

The trigonometric functions also have inverse functions; however, we have to restrict the domain of the functions so that becomes injective and therefore posses a inverse. Normally we take the following domains for the trigonometric functions and their inverses[20]:

[20] Sometimes the inverse of $\sin x$ is denoted $\sin^{-1} x$ and similarly for the other functions, however, this notation should be avoided because $\sin^{-1} x$ might be confused with $\frac{1}{\sin x}$, which is not the inverse function of $\sin x$

Trigonometric Function	**Domain**	**Inverse Function**	**Domain**
$\sin x$	$-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$	$\arcsin x$	$-1 \leq x \leq 1$
$\cos x$	$0 \leq x \leq \pi$	$\arccos x$	$-1 \leq x \leq 1$
$\tan x$	$-\frac{\pi}{2} < x < \frac{\pi}{2}$	$\arctan x$	$\mathbb{R}$
$\csc x$	$(0, \pi/2] \cup (\pi, 3\pi/2]$	$\mathrm{arccsc} x$	$\|x\| \geq 1$
$\sec x$	$[0, \pi/2) \cup [\pi, 3\pi/2)$	$\mathrm{arcsec} x$	$\|x\| \geq 1$
$\cot x$	$(0, \pi)$	$\mathrm{arccot} x$	$\mathbb{R}$

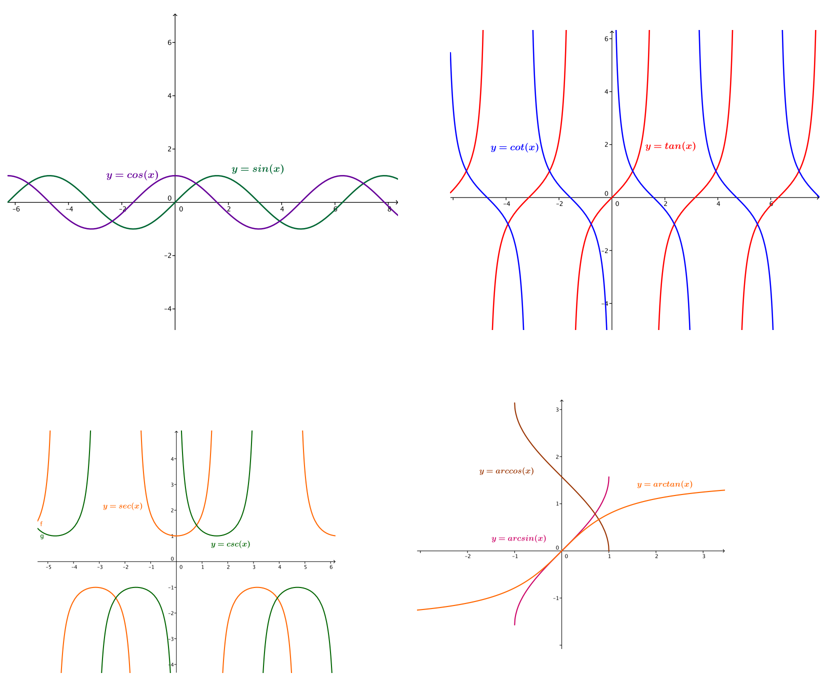


Figure 49: Some Trigonometric Functions

Part II

Limits

Introduction to Limits and some Properties

As mentioned earlier, there are many situations in which we need to approximate a quantity (like finding $\sqrt{2}$) by a sequence of values that get closer and closer to our desired quantity without ever being able to "get to" the desired value (in the example of $\sqrt{2}$ this corresponds to the table of approximations we showed at the beginning that never finish but they become better approximations with each successive step).

An example that resembles more what we will be doing in this course is the following: consider the function

$$f(x) = \frac{x^2 - 4}{x - 2} \tag{108}$$

Up to this point we have not been very worried about identifying the domains of the functions, however, in Calculus knowing the domain of a function is critical.

Therefore, to find the domain of 108 we need to ask when does the expression $\frac{x^2-4}{x-2}$ make sense? Clearly the only problem that may show up is when the denominator $x - 2$ becomes 0, which is when $x = 2$, because we would be forced to make sense out of

$$f(2) \stackrel{?}{=} \frac{2^2 - 4}{2 - 2} = \frac{0}{0} \tag{109}$$

and this is undefined since division by 0 is not allowed. Therefore, $f(x)$ is not defined at $x = 2$ and we will say that the domain of $f(x)$ is $\mathbb{R} - \{2\}$, that is, all the points in the real line except the point $x = 2$.

Now, if we had not paid attention to the domain of $f(x)$ we may have used the formula for the difference of squares

$$\frac{x^2 - 4}{x - 2} = \frac{(x-2)(x+2)}{x - 2} = x + 2 \tag{110}$$

and someone would have said that $f(2)$ has to be $2 + 2 = 4$. Again, this analysis is correct except that we can't cancel $x - 2$ when $x = 2$ so this method can't be used at the point where we are having problems. To make this more precise let's define a new function

$$g(x) = x + 2 \tag{111}$$

We can see that $g(x)$ is defined on the entire real line $\mathbb{R}$ and by the calculation in 110 we know that if $x \neq 2$ then $f(x) = g(x)$. Therefore,

$f(x)$ and $g(x)$ are two functions that agree on $\mathbb{R} - \{2\}$ but they have different domains so as a general rule we will consider two functions to be different if they have different domains even when they agree on their common domain.

If we plot the function $f(x)$ we will see something as the following figure shows. Here the red dot is used to remind us that we have not defined $f(x)$ at the point $x = 2$.

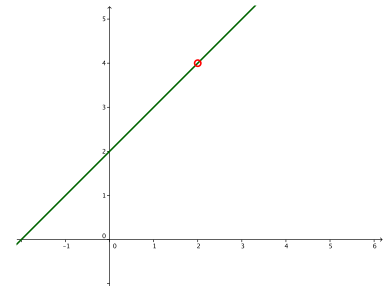


Figure 50: Graph of $f(x) = \frac{x^2-4}{x-2}$

As we will see shortly, one of the purposes of the concept of a limit is to analyze what happens to the behavior of $f(x)$ when $x = 2$ even when $f(x)$ is not defined there. To be more precise, $\lim_{x \longrightarrow 2} \frac{x^2-4}{x-2}$ means the following: what is the behavior of the function $\frac{x^2-4}{x-2}$ as the value of x gets closer and closer to 2 without ever being 2?

In other words, we start with a value $x \neq 2$ and study $\frac{x^2-4}{x-2}$. Because of the previous calculations for $x \neq 2$ we have $\frac{x^2-4}{x-2} = x + 2$ so we need to analyze the behavior of $x + 2$ as x gets closer and closer to 2, without being equal to 2. Clearly as x gets closer to 2, $x + 2$ gets closer to 4 so we will say that

$$\lim_{x \longrightarrow 2} \left(\frac{x^2 - 4}{x - 2} \right) = \lim_{x \longrightarrow 2} (x + 2) = 4 \tag{112}$$

Again the crucial point here is that to study a limit at a specific point we don't care about the value of the function at that point or even if the function is defined at that point. Also, the limit at a point only depends on the behavior of the function near that point and it does not matter how the function behaves away from that point.

Consider the following two figures. In the first figure we take $f(x) = \frac{x^2-4}{x-2}$ and since the function $f(x)$ is not defined at $x = 2$ we decide arbitrarily to make its value 2 when $x = 2$. Since we are changing the domain of the function because it is now defined at $x = 2$ we can use a new name for this function, for example, call it $F(x)$ so that

$$F(x) = \begin{cases} \frac{x^2-4}{x-2} & \text{if } x \neq 2 \\ 2 & \text{if } x = 2 \end{cases} \tag{113}$$

The new function $F(x)$ behaves exactly the same as $f(x)$ when $x \neq 2$ so the first property for the limit says that

$$\lim_{x \longrightarrow 2} F(x) = \lim_{x \longrightarrow 2} \frac{x^2 - 4}{x - 2} = 4 \tag{114}$$

because the limit does not care about what happens with the function at the point where we are trying to evaluate the limit.

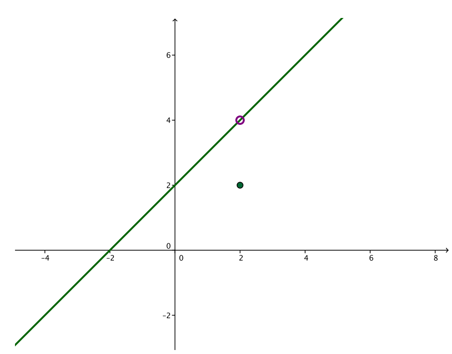


Figure 51: Limit at a point is independent of the value of the function at that point

In the second example, we modified the function away from the point $x = 2$ but didn't change it around $x = 2$. If $h(x)$ is the new function then we can take it to be

$$h(x) = \begin{cases} 2 & \text{if } x \leq 0 \\ \frac{x^2-4}{x-2} & \text{if } 0 < x < 3 \text{ and } x \neq 2 \\ 5 & \text{if } 3 \leq x \end{cases} \tag{115}$$

and the properties for limits say that

$$\lim_{x \longrightarrow 2} h(x) = \lim_{x \longrightarrow 2} f(x) = 4 \tag{116}$$

From a practical point of view, we can create a table with the values of the function to see if there is a pattern that help us determine what the limit will be. For example, if we use our original function $f(x) = \frac{x^2-4}{x-2}$ and we want to find $\lim_{x \longrightarrow 2} \frac{x^2-4}{x-2}$ our table would look like

values of x	values of $f(x) = \frac{x^2-4}{x-2}$
2.1	4.1
2.01	4.01
2.001	4.001
2.0001	4.0001
2	???
1.9999	3.9999
1.999	3.999
1.99	3.99
1.9	3.9

Observe that we can't fill the table when $x = 2$ because the function is not defined at that point. However, we can see that for any other value of x, if x gets closer to 2 then $f(x)$ is getting closer to 4 which suggests that $\lim_{x \longrightarrow 2} \frac{x^2-4}{x-2} = 4$ and this is in fact the case. This method is not flawless since in principle to have complete certainty our table would have to be evaluated at an infinite number of points so this can only be used to suggest what the limit should be instead of proving what the limit actually is.

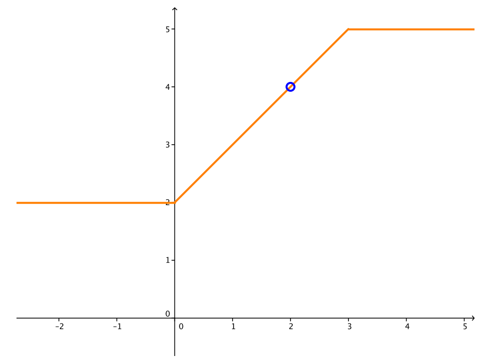


Figure 52: Limit at a point is independent of the value of the function at that point

As far as this course is concerned, most of the limits that will be studied can be obtained by applying a set of rules that specify how the limits behave, so obtaining limits will become somewhat mechanical. Nevertheless, it is important to note that most of the heuristic arguments that will be given here can be improved to give a more satisfactory account of why a limit exists and has a particular value. In this way we have:

Suppose that $f(x)$ is a function which may or may not be defined at the point a. Then we say that L is the limit of $f(x)$ as x approaches a, and we write

$$\lim_{x \longrightarrow a} f(x) = L \tag{117}$$

if the values of $f(x)$ can be taken as close as we want to L by taking a sufficiently close number x to a.

Under this "definition", $\lim_{x \longrightarrow 2} \frac{x^2-4}{x-2}$ is 4 because if we take $L = 4$ then $f(x) = \frac{x^2-4}{x-2}$ is as close as we want to 4 by choosing a sufficiently close value x to 2.

Example 24. Find the limit $\lim_{x\longrightarrow 1}\frac{x-1}{x^3+x^2-2x}$

First of all, if we call $f(x)=\frac{x-1}{x^3+x^2-2x}$ we can see that $f(x)$ is not defined at $x=1$ because the denominator becomes $1^3+1^2-2\cdot 1=0$ and division by 0 is not possible. Therefore we are in a similar situation to our first example. To deal with it we write the denominator in a more convenient way

$$x^3+x^2-2x = x\left(x^2+x-2\right) = x\left(x+2\right)\left(x-1\right) \tag{118}$$

In this way

$$\frac{x-1}{x^3+x^2-2x} = \frac{x-1}{x\left(x+2\right)\left(x-1\right)} \tag{119}$$

Now we would like to cancel the term $x-1$ in 119. We can only do this if $x\neq 1$. Since we are studying the limit as x approaches 1, we are never at the point 1 so we are allowed to cancel $x-1$ and so 119 becomes

$$\frac{1}{x(x+2)} \tag{120}$$

Therefore

$$\lim_{x\longrightarrow 1}\frac{x-1}{x^3+x^2-2x} = \lim_{x\longrightarrow 1}\frac{1}{x(x+2)} \tag{121}$$

The point of doing the previous algebraic manipulation is that $\frac{1}{x(x+2)}$ is defined when $x=1$. In fact, when $x=1$ we have $\frac{1}{1\cdot(1+2)}=\frac{1}{3}$ and as we will see later this will imply that the limit must be $\frac{1}{3}$. For now we will be happy with some "visual" and "numerical" evidence. In any case, we have

$$\lim_{x\longrightarrow 1}\frac{x-1}{x^3+x^2-2x} = \lim_{x\longrightarrow 1}\frac{1}{x(x+2)} = \frac{1}{3} \tag{122}$$

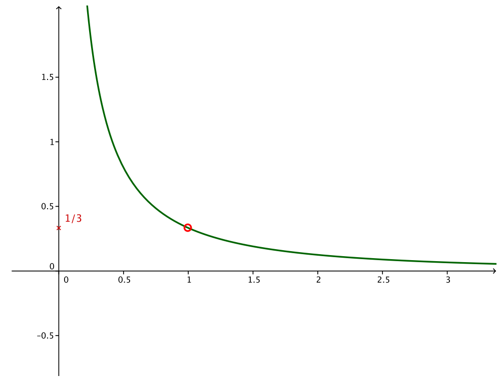


Figure 53: $y=\frac{x-1}{x^3+x^2-2x}$

values of x	values of $f(x)=\frac{x-1}{x^3+x^2-2x}$
1.1	$\frac{1}{3.41}$
1.01	$\frac{1}{3.0401}$
1.001	$\frac{1}{3.004001}$
1.0001	$\frac{1}{3.00040001}$
1	???
0.9999	$\frac{1}{2.99960001}$
0.999	$\frac{1}{2.996001}$
0.99	$\frac{1}{2.9601}$
0.9	$\frac{1}{2.61}$

Table 6: Numerical Approximation $\lim_{x\longrightarrow 1}\frac{x-1}{x^3+x^2-2x}$

Limits for polynomials and radicals:

- **Limits for polynomials:** suppose that $f(x)$ is a polynomial. For example, $f(x) = x^3 - 5x + 2$, $f(x) = x - 2$ or $f(x) = x^2 + 2$. Then $f(x)$ is defined everywhere and if $a \in \mathbb{R}$ then $\lim_{x \longrightarrow a} f(x) = f(a)$, that is, to find a limit for a polynomial we can simply evaluate. For example,

$$\begin{aligned} \lim_{x\longrightarrow 1} \left(x^3 - 5x + 2\right) &= \left(1^3 - 5\cdot 1 + 2\right) = -2 \\ \lim_{x\longrightarrow -3} (x-2) &= (-3-2) = -5 \\ \lim_{x\longrightarrow 0} \left(x^2 + 2\right) &= \left(0^2 + 2\right) = 2 \end{aligned} \tag{123}$$

- **Limits for radicals of polynomials:** suppose that $f(x)$ is the root of some polynomial. For example, $f(x) = \sqrt{x^2+1}$, $f(x) = \sqrt[3]{x^3 - 2x + 1}$, $f(x) = \sqrt[5]{3x-1}$. Then $\lim_{x\longrightarrow a} f(x) = f(a)$ provided that a belongs to the domain of $f(x)$. For example,

$$\begin{aligned} \lim_{x\longrightarrow -1} \sqrt{x^2+1} &= \sqrt{(-1)^2+1} = \sqrt{2} \\ \lim_{x\longrightarrow 2} \sqrt[3]{x^3 - 2x + 1} &= \sqrt[3]{2^3 - 2\cdot 2 + 1} = \sqrt[3]{5} \\ \lim_{x\longrightarrow \frac{1}{2}} \sqrt[5]{3x-1} &= \sqrt[5]{3\cdot\frac{1}{2} - 1} = \sqrt[5]{\frac{1}{2}} = \frac{1}{\sqrt[5]{2}} \end{aligned} \tag{124}$$

The previous examples show that in many cases it is possible to find a limit by evaluating the function at that point. Therefore, it may seem that we can avoid talking about limits and be happy with evaluating the function at a point. However, the concept of limit is more general since we already saw cases for which $f(a)$ does not make sense yet the limit is well defined. Most of the interesting limits fall under the latter situation and before working more examples we need some additional results on how to find limits via basic algebraic operations.

Algebraic Operations with Limits:
Suppose that $f(x)$ and $g(x)$ are two functions for which

$$\lim_{x \longrightarrow a} f(x) = L_1 \qquad \lim_{x \longrightarrow a} g(x) = L_2 \tag{125}$$

Then

- Constants can be taken outside a limit: if $c \in \mathbb{R}$ is a constant then

$$\lim_{x \longrightarrow a} cf(x) = c \lim_{x \longrightarrow a} f(x) = cL_1 \tag{126}$$

  for example, since $\lim_{x \longrightarrow 2} \left(x^2+1\right) = 5$ then

$$\lim_{x \longrightarrow 2} 2\left(x^2+1\right) = 2 \lim_{x \longrightarrow 2} \left(x^2+1\right) = 10 \tag{127}$$

  .

- The limit of a sum is the sum of the limits:

$$\lim_{x \longrightarrow a} (f(x) \pm g(x)) = \lim_{x \longrightarrow a} f(x) \pm \lim_{x \longrightarrow a} g(x) = L_1 \pm L_2 \tag{128}$$

  for example, since $\lim_{x \longrightarrow 2} \sqrt{x+3} = \sqrt{5}$ then

$$\lim_{x \longrightarrow 2} \left(x^2+1+\sqrt{x+3}\right) = \lim_{x \longrightarrow 2} \left(x^2+1\right) + \lim_{x \longrightarrow 2} \sqrt{x+3} = 5 + \sqrt{5} \tag{129}$$

- The limit of a product is the product of the limits:

$$\lim_{x \longrightarrow a} f(x)g(x) = \left(\lim_{x \longrightarrow a} f(x)\right)\left(\lim_{x \longrightarrow a} g(x)\right) = L_1 L_2 \tag{130}$$

  For example,

$$\lim_{x \longrightarrow 2} \left[\left(x^2+1\right)\sqrt{x+3}\right] = \left(\lim_{x \longrightarrow 2} (x^2+1)\right)\left(\lim_{x \longrightarrow 2} \sqrt{x+3}\right) = 5\sqrt{5} \tag{131}$$

- The limit of a power is the power of a limit:

$$\lim_{x \longrightarrow a} (f(x))^r = \left[\lim_{x \longrightarrow a} f(x)\right]^r = L^r \tag{132}$$

  provided L^r makes sense. For example,

$$\lim_{x \longrightarrow 2} \left(x^2+1\right)^{-\frac{1}{2}} = \left[\lim_{x \longrightarrow 2} x^2+1\right]^{-\frac{1}{2}} = [5]^{-\frac{1}{2}} = \frac{1}{\sqrt{5}} \tag{133}$$

- The limit of a quotient is the quotient of the limits provided the limit of the denominator is not zero:

$$\lim_{x \longrightarrow a} \left(\frac{f(x)}{g(x)}\right) = \frac{\lim_{x \longrightarrow a} f(x)}{\lim_{x \longrightarrow a} g(x)} = \frac{L_1}{L_2} \qquad \text{provided } L_2 \neq 0 \tag{134}$$

  For example,

$$\lim_{x \longrightarrow 2} \left(\frac{x^2+1}{\sqrt{x+3}}\right) = \frac{\lim_{x \longrightarrow 2} \left(x^2+1\right)}{\lim_{x \longrightarrow 2} \sqrt{x+3}} = \frac{5}{\sqrt{5}} \tag{135}$$

Example 25. Find $\lim_{x \longrightarrow -5} \frac{x^2-25}{x+5}$

Again, the function $f(x) = \frac{x^2-25}{x+5}$ is not defined at $x = -5$ so before applying the limit techniques we have to rewrite $f(x)$ in a more convenient expression. Using the formula for difference of squares we can write the numerator as

$$x^2 - 25 = (x-5)(x+5) \tag{136}$$

Therefore $f(x)$ becomes

$$\frac{x^2-25}{x+5} = \frac{(x-5)(x+5)}{x+5} = x-5 \tag{137}$$

since we can cancel the term $x+5$ by a similar argument to the ones given previously. Therefore

$$\lim_{x \longrightarrow -5} \frac{x^2-25}{x+5} = \lim_{x \longrightarrow -5} (x-5) \tag{138}$$

Now, $x-5$ is a polynomial so it is defined everywhere and by the rules given for polynomials we can simply evaluate the limit at $x = -5$

$$\lim_{x \longrightarrow -5} (x-5) = -5-5 = -10 \tag{139}$$

and in this way $\lim_{x \longrightarrow -5} \frac{x^2-25}{x+5} = -10$

Example 26. Find $\lim_{z \longrightarrow 2} \frac{z^3-8}{z-2}$

Notice in this example that instead of using the variable x we are using the variable z. The point of using different notations is to realize that the symbol used for a variable from a mathematical point of view is arbitrary although in practice some notation is better than the other (for example, it is better to use t to represent time than use the variable θ or x). Therefore, from a mathematical point of view the following limits are the same

$$\lim_{z \longrightarrow 2} \frac{z^3-8}{z-2} = \lim_{x \longrightarrow 2} \frac{x^3-8}{x-2} = \lim_{\phi \longrightarrow 2} \frac{\phi^3-8}{\phi-2} = \lim_{\bullet \longrightarrow 2} \frac{\bullet^3-8}{\bullet-2} = \lim_{\star \longrightarrow 2} \frac{\star^3-8}{\star-2} \tag{140}$$

To find the limit we can rewrite the expression $\frac{z^3-8}{z-2}$ by factorizing $z^3 - 8$ using the difference of cube formula

$$z^3 - 8 = z^3 - 2^3 = (z-2)(z^2+2z+4) \tag{141}$$

Therefore

$$\lim_{z \longrightarrow 2} \frac{z^3-8}{z-2} = \lim_{z \longrightarrow 2} \frac{(z-2)\left(z^2+2z+4\right)}{z-2} = \lim_{z \longrightarrow 2} \left(z^2+2z+4\right) \tag{142}$$

Since z^2+2z+4 is a polynomial we can evaluate the polynomial at $z = 2$

$$\lim_{z \longrightarrow 2} \left(z^2+2z+4\right) = 2^2 + 2 \cdot 2 + 4 = 12 \tag{143}$$

Example 27. Find $\lim_{x\longrightarrow 1}\frac{\sqrt{x}-1}{x-1}$

In this case we need to rationalize the numerator $\sqrt{x}-1$

$$\sqrt{x}-1=\left(\sqrt{x}-1\right)\cdot\left(\frac{\sqrt{x}+1}{\sqrt{x}+1}\right)=\frac{\left(\sqrt{x}-1\right)\left(\sqrt{x}+1\right)}{\sqrt{x}+1}=\frac{\left(\sqrt{x}\right)^2-1^2}{\sqrt{x}+1}=\frac{x-1}{\sqrt{x}+1} \tag{144}$$

Therefore the fraction becomes

$$\frac{\sqrt{x-1}}{x-1}=\frac{x-1}{\left(\sqrt{x}+1\right)\left(x-1\right)} \tag{145}$$

and the limit is

$$\lim_{x\longrightarrow 1}\frac{\sqrt{x}-1}{x-1}=\lim_{x\longrightarrow 1}\frac{1}{x-1}\cdot\left(\frac{x-1}{\sqrt{x}+1}\right)=\lim_{x\longrightarrow 1}\frac{1}{\sqrt{x}+1} \tag{146}$$

Now, $\sqrt{x}$ is defined at $x=1$ so by the properties of limits the limit of the denominator is

$$\lim_{x\longrightarrow 1}\left(\sqrt{x}+1\right)=\lim_{x\longrightarrow 1}\sqrt{x}+\lim_{x\longrightarrow 1}1=\sqrt{1}+1=2 \tag{147}$$

and the limit of $\frac{1}{\sqrt{x}+1}$ is the quotient of the limits

$$\lim_{x\longrightarrow 1}\frac{1}{\sqrt{x}+1}=\frac{\lim_{x\longrightarrow 1}1}{\lim_{x\longrightarrow 1}\left(\sqrt{x}+1\right)}=\frac{1}{2} \tag{148}$$

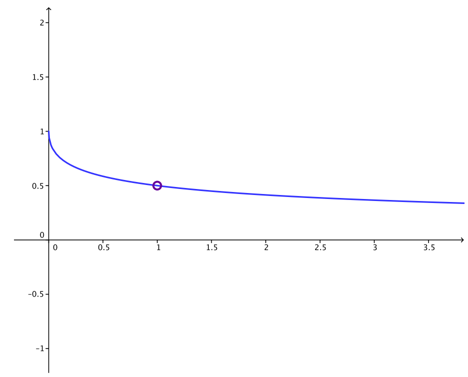


Figure 54: $f(x)=\frac{\sqrt{x}-1}{x-1}$

One-Sided Limits, Infinite Limits and Limits at Infinity

We have seen that to find $\lim_{x \longrightarrow a} f(x)$ we need to analyze the behavior of $f(x)$ near $x = a$. In the previous examples the limit has always existed, however, there are many case in which this is not the case. Consider a function whose graph looks like the following figure.

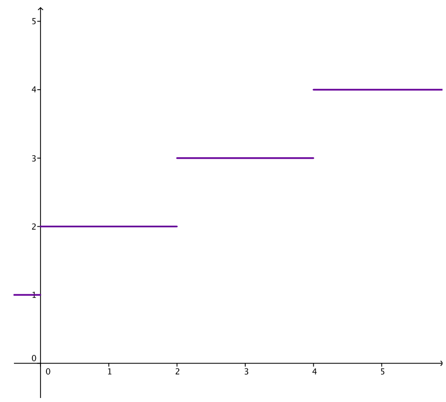


Figure 55: Function without limit when t is an even integer

This function could be used to represent the size of a population in time, since the size of a population remains constant for a certain period of time until an individual dies or is born. If we call this function $P(t)$ then we can see that $P(t)$ is constant on the intervals $(0,2)$, $(2,4)$, $(4,6)$, etc, and as time goes by a new member is added at the end of each interval. If we want we can write

$$P(t) = \begin{cases} 2 & 0 < t < 2 \\ 3 & 2 < t < 4 \\ 4 & 4 < t < 6 \\ \vdots & \\ n+2 & 2n < t < 2n+2 \end{cases} \tag{149}$$

Now we can see that in the definition of 149 there is some ambiguity with the value of $P(t)$ when t is an even integer because at that instant a new member is born so $P(t)$ suddenly changes.

Now, what happens when we try to find $\lim_{t \longrightarrow t_0} P(t)$ where t_0 is a specific value of time? If t_0 is not even, then t_0 belongs to an interval in which $P(t)$ is constant therefore $\lim_{t \longrightarrow t_0} P(t)$ will be that constant value. For example, $\lim_{t \longrightarrow 2.001} P(t) = 3$ and $\lim_{t \longrightarrow \sqrt{2}} P(t) = 2$.

However, when t_0 is even the situation is entirely different, in fact in this case the limit does not exist! To see this we can create a numerical table with the behavior of $P(t)$ near 2.

values of t	values of $P(t)$
2.1	3
2.01	3
2.001	3
2.0001	3
2	???
1.9999	2
1.999	2
1.99	2
1.9	2

Table 7: Values of Population

In contrast to the previous numerical tables, $P(t)$ takes only two values near $t_0 = 2$ and they are not getting closer as t becomes closer to 2. More precisely, if we choose a time larger than 2 then $P(t) = 3$ while if we choose a time smaller than 2 then $P(t) = 2$ so the limit can't exist when t_0 is 2.

However, not everything is lost since the fact that the behavior of $P(t)$ can be separated into two cases allows us to talk about a partial notion of limit, called one-sided limits.

To be more precise, any point in the real line can be approached from two directions, from the right or from the left. If we approach the point

t_0 from the right and there is a limit when we consider only points to the right of t_0 then we will say that the **right hand limit** exists and we denote it $\lim_{t\longrightarrow t_0^+} P(t)$. For example, $\lim_{t\longrightarrow 2^+} P(t) = 3$. If we approach the point t_0 from the left and there is a limit when we consider only points to the left of t_0 then we will say that the **left hand limit** exists and we denote it $\lim_{t\longrightarrow t_0^-} P(t)$. For example, $\lim_{t\longrightarrow 2^-} P(t) = 2$. The left hand limit and the right hand limit are called **one-sided limits**.

Therefore, if the limit does not exist, the one-sided limits are some form of consolation prize that gives information about the behavior of the function. However, in another sense they are more than a consolation prize because if they agree then the limit must exist

Suppose that $f(x)$ is a function and $a \in \mathbb{R}$ a point that may or may not belong to the domain of $f(x)$. Then the limit

$$\lim_{x\longrightarrow a} f(x) \tag{150}$$

exists if and only if both one-sided limits

$$\lim_{x\longrightarrow a^-} f(x) \qquad \lim_{x\longrightarrow a^+} f(x) \tag{151}$$

are the same.

With this criteria, given that $\lim_{t\longrightarrow 2^+} P(t) = 3$ and $\lim_{t\longrightarrow 2^-} P(t) = 2$ are different we can conclude that $\lim_{t\longrightarrow 2} P(t)$ does not exist.

Example 28. Determine if $\lim_{x\longrightarrow 0} f(x)$ exists where $f(x) = \begin{cases} -1 & \text{if } x \leq 0 \\ x-1 & \text{if } x > 0 \end{cases}$

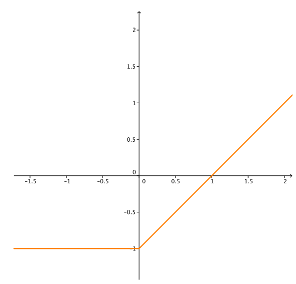


Figure 56: Example lateral limits

We will use the criteria that the limit exists if and only if both one-sided limits exist and are the same.

Since $f(x)$ is defined by parts using one-sided limits is very convenient. Looking at the graph of the function, or the definition of $f(x)$ we can see that

$$\lim_{x\longrightarrow 0^-} f(x) = \lim_{x\longrightarrow 0^-} -1 = -1 \tag{152}$$

on the other hand

$$\lim_{x\longrightarrow 0^+} f(x) = \lim_{x\longrightarrow 0^+} (x-1) = \lim_{x\longrightarrow 0^+} x - \lim_{x\longrightarrow 0^+} 1 = 0 - 1 = -1 \tag{153}$$

Since the one-sided limits are the same the limit exists so

$$\lim_{x\longrightarrow 0} f(x) = -1 \tag{154}$$

Example 29. Determine if $\lim_{x\longrightarrow 1} f(x)$ exists where $f(x) = \begin{cases} -2x+4 & \text{if } x < 1 \\ 4 & \text{if } x = 1 \\ \frac{1}{10}x^2 + 1.9 & \text{if } x > 1 \end{cases}$

Again we use one-sided limits

$$\lim_{x\longrightarrow 1^-} f(x) = \lim_{x\longrightarrow 1^-} (-2x+4) = -2\left(\lim_{x\longrightarrow 1^-} x\right) + 4 = 2 \tag{155}$$

$$\lim_{x\longrightarrow 1^+} f(x) = \lim_{x\longrightarrow 1^+}\left(\frac{1}{10}x^2+1.9\right) = \frac{1}{10}\left(\lim_{x\longrightarrow 1^+} x^2\right)+1.9 = \frac{1}{10}+1.9 = 2 \tag{156}$$

Because the one-sided limits agree we conclude that the limit exists so

$$\lim_{x\longrightarrow 1} f(x) = 2 \tag{157}$$

We can see in this case that the limit is different from $f(1)$ which serves to illustrate the point that the limit "does not care" about the value of the function at a point, only about its behavior near that point. The functions for which the limit agrees with the value of the function at the point will be studied later in the course.

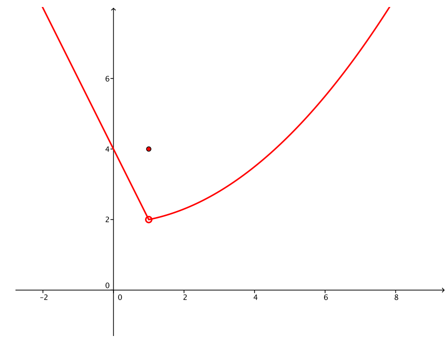


Figure 57: Example Lateral Limits

Example 30. Find $\lim_{t\longrightarrow 2}\frac{|3t-6|}{t-2}$

Observe that by the properties of the absolute value

$$|3t-6| = |3(t-2)| = 3\,|t-2| \tag{158}$$

Now we use the definition of the absolute value

$$|t-2| = \begin{cases} t-2 & \text{if } t-2\geq 0 \\ -(t-2) & \text{if } t-2<0 \end{cases} = \begin{cases} t-2 & \text{if } t\geq 2 \\ -(t-2) & \text{if } t<2 \end{cases} \tag{159}$$

Therefore, to find the limit it is easier to work with the one-sided limits

$$\lim_{t\longrightarrow 2^-}\frac{3\,|t-2|}{t-2} = \lim_{t\longrightarrow 2^-}\frac{-3\,(t-2)}{t-2} = \lim_{t\longrightarrow 2^-} -3 = -3 \tag{160}$$

$$\lim_{t\longrightarrow 2^+}\frac{3\,|t-2|}{t-2} = \lim_{t\longrightarrow 2^+}\frac{3\,(t-2)}{t-2} = \lim_{t\longrightarrow 2^-} 3 = 3 \tag{161}$$

Since the one-sided limits are different $\lim_{t\longrightarrow 2}\frac{|3t-6|}{t-2}$ does not exist.

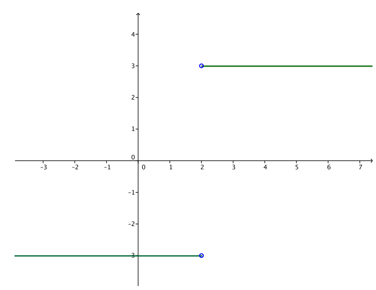


Figure 58: Example Lateral Limit

So far we have implicitly assumed that the limit of a function is a real number. However, in some situations it is useful to extend the notion of what a limit can be and introduce infinite limits as possible values of a limit. For example, consider $f(x) = \frac{1}{x^2}$. We can see that $f(x)$ is not defined for $x = 0$ but the function has a different behavior near 0 from the ones we have seen so far. This is clear if we take a look at its graph and the numerical table.

In this situation we will say that

$$\lim_{x\longrightarrow 0}\frac{1}{x^2} = \infty \tag{162}$$

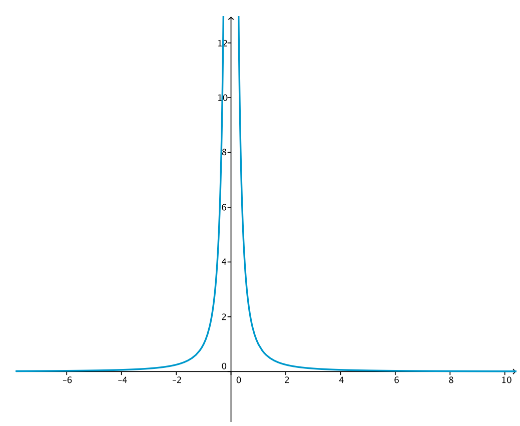


Figure 59: Behavior of $f(x) = \frac{1}{x^2}$ near 0

where the symbol "∞", called infinity, means that as x approaches 0, $\frac{1}{x^2}$ keeps getting bigger without restrictions. It must be emphasized that the symbol ∞ does not stand for a number, it is just a convenient way of saying that $\frac{1}{x^2}$ increases as much as we want near 0.

In the same way we can introduce the symbol $-\infty$ to represent the fact that the value of a function decreases as much as we want. For example, if $f(x) = -\frac{1}{x^2}$ we would have

$$\lim_{x \longrightarrow 0} -\frac{1}{x^2} = -\infty \tag{163}$$

since the behavior of $-\frac{1}{x^2}$ is exactly the same as the one of $\frac{1}{x^2}$ except that the positive values become negative values.

values of x	values of $f(x)$
0.1	100
0.01	10000
0.001	1000000
0.0001	100000000
0	???
-0.0001	100000000
-0.001	1000000
-0.01	10000
-0.1	100

Table 8: Numerical Table $f(x) = \frac{1}{x^2}$

Also, it may be the case that one lateral limit is ∞ while the other is $-\infty$. This happens in the case of $f(x) = \frac{1}{x^2-1}$ where we have

$$\lim_{x \longrightarrow -1^-} \frac{1}{x^2-1} = \infty \qquad \lim_{x \longrightarrow -1^+} \frac{1}{x^2-1} = -\infty \tag{164}$$

$$\lim_{x \longrightarrow 1^-} \frac{1}{x^2-1} = -\infty \qquad \lim_{x \longrightarrow 1^+} \frac{1}{x^2-1} = \infty \tag{165}$$

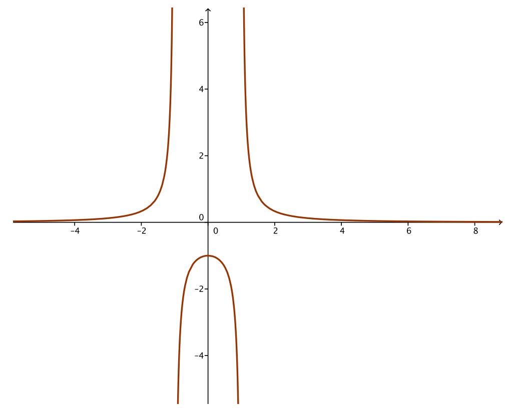


Figure 60: Behavior of $f(x) = \frac{1}{x^2-1}$ at $x = 1$ and $x = -1$

Although ∞ and $-\infty$ are not numbers there are some manipulations that can be performed with those symbols which behave as numbers:

Symbolic Manipulation of ∞ and $-\infty$:

- $\infty + \infty = \infty$
- $-\infty - \infty = -\infty$
- $\infty \cdot \infty = \infty$
- $\infty \cdot (-\infty) = (-\infty) \cdot \infty = -\infty$
- $c + \infty = \infty$ and $c - \infty = -\infty$ where c is a constant
- $c\infty = \infty$, $c(-\infty) = -\infty$ if $c > 0$ and $c\infty = -\infty$, $c(-\infty) = \infty$ if $c < 0$
- $\frac{c}{\infty} = 0$ and $\frac{c}{-\infty} = 0$ where c is a constant different from 0
- $\infty - \infty$ or $-\infty + \infty$ are not defined
- $\frac{\infty}{\infty}$, $\frac{-\infty}{\infty}$ and $\frac{\infty}{-\infty}$ are not defined

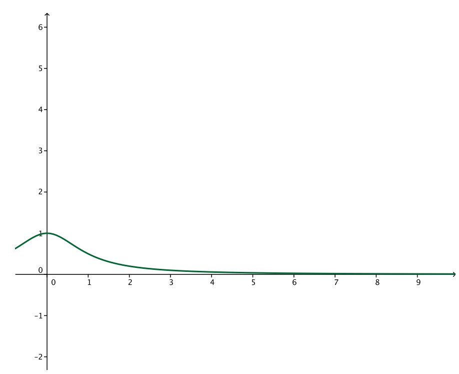


Figure 61: Graph of $f(x) = \frac{1}{x^2+1}$

Up to this point, we have analyzed the behavior of a function near a point of the real line, however, it is also possible to analyze the behavior of a function as we "go to infinity". If we consider our variable as being time, then this kind of limit represents the asymptotic behavior (long term behavior) of a physical system.

For example, consider $f(x) = \frac{1}{x^2+1}$. If we look at the graph of $f(x)$ we can see that as we take larger and larger values of x, or to put it

values of x	values of $f(x)$
1	0.5
10	0.00990099...
100	0.00009999...
1000	0.000000999999...

Table 9: Numerical Approximation $\lim_{x \longrightarrow \infty} \frac{1}{x^2+1}$

another way, as we move farther and farther to the right, the values of $f(x)$ approach 0 as much as we want. This can also be seen from a numerical table for the values of $f(x)$ so we write

$$\lim_{x\longrightarrow\infty} \frac{1}{x^2+1} = 0 \tag{166}$$

In this case $x \longrightarrow \infty$ means that we are moving farther and farther to the right. In a similar manner, we can consider a function like $f(x) = -x$. If we look at the graph of $f(x)$ we can see that as we move farther and farther to the left $f(x)$ keeps increasing.

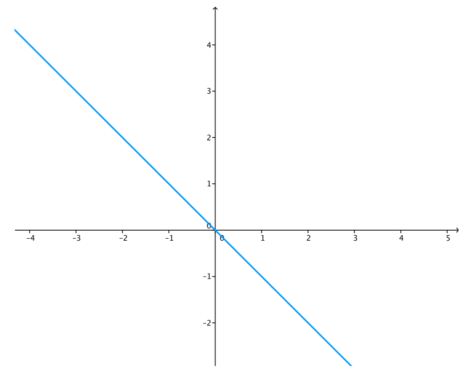


Figure 62: Graph of $f(x) = -x$

In this way we write

$$\lim_{x\longrightarrow-\infty} -x = \infty \tag{167}$$

values of x	values of $f(x)$
-1	1
-10	10
-100	100
-1000	1000

Table 10: Numerical Table

This example shows that a limit at infinity can be infinite!

Example 31. Find $\lim_{x\longrightarrow\infty} \frac{3x+2}{x-5}$

If we try to evaluate directly we would end up with something like

$$\frac{3\infty+2}{\infty-5} = \frac{\infty}{\infty} \tag{168}$$

which is not defined. Therefore we need to try a different strategy.

In general, to study a limit at infinity which is a quotient of two polynomials we factorize the highest power of x in the numerator and the denominator.

For example, $\frac{3x+2}{x-5}$ is a quotient of two polynomials so and the higher power of x in the numerator and denominator is x. Therefore

$$\frac{3x+2}{x-5} = \frac{x\left(3+\frac{2}{x}\right)}{x\left(1-\frac{5}{x}\right)} = \frac{3+\frac{2}{x}}{1-\frac{5}{x}} \tag{169}$$

where we can cancel the x because as x goes to infinity x won't be 0. The purpose of rewriting the expression in this way is that the numerator and denominator now have a well defined value as x goes to infinity.

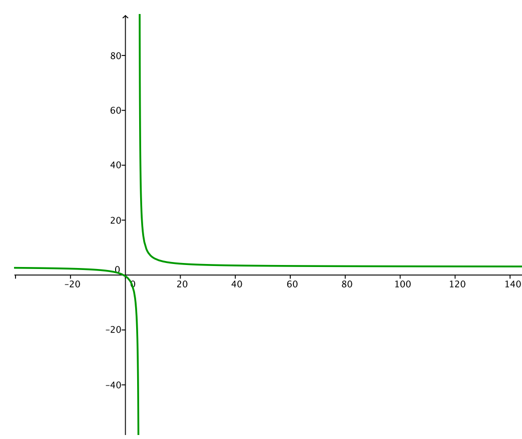


Figure 63: Graph of $f(x) = \frac{3x+2}{x-5}$

$$\lim_{x\longrightarrow\infty}\left(3+\frac{2}{x}\right) = 3 + 2\lim_{x\longrightarrow\infty}\left(\frac{1}{x}\right) = 3+2\cdot 0 = 3 \tag{170}$$

$$\lim_{x\longrightarrow\infty}\left(1-\frac{5}{x}\right) = 1 - 5\lim_{x\longrightarrow\infty}\left(\frac{1}{x}\right) = 1-5\cdot 0 = 1 \tag{171}$$

Therefore,

$$\lim_{x\longrightarrow\infty}\frac{3x+2}{x-5} = \lim_{x\longrightarrow\infty}\left(\frac{3+\frac{2}{x}}{1-\frac{5}{x}}\right) = \frac{\lim_{x\longrightarrow\infty}\left(3+\frac{2}{x}\right)}{\lim_{x\longrightarrow\infty}\left(1-\frac{5}{x}\right)} = \frac{3}{1} = 3 \tag{172}$$

Example 32. Find $\lim_{x\longrightarrow-\infty}\frac{4x^2-1}{x+2}$

Again this is the limit of two polynomials. The highest power of x in the numerator is x^2 and the highest power of x in the denominator is x so

$$\frac{4x^2-1}{x+2}=\frac{x^2\left(4-\frac{1}{x^2}\right)}{x\left(1+\frac{2}{x}\right)}=x\left(\frac{4-\frac{1}{x^2}}{1+\frac{2}{x}}\right) \tag{173}$$

Now,

$$\lim_{x\longrightarrow-\infty}\frac{4-\frac{1}{x^2}}{1+\frac{2}{x}}=\frac{\lim_{x\longrightarrow-\infty}\left(4-\frac{1}{x^2}\right)}{\lim_{x\longrightarrow-\infty}\left(1+\frac{2}{x}\right)}=\frac{4}{1}=4 \tag{174}$$

Therefore, using the rules for multiplying infinity times a constant we have

$$\begin{aligned}\lim_{x\longrightarrow-\infty}\tfrac{4x^2-1}{x+2}&=\lim_{x\longrightarrow-\infty}x\left(\tfrac{4-\frac{1}{x^2}}{1+\frac{2}{x}}\right)\\=\left(\lim_{x\longrightarrow-\infty}x\right)\left(\lim_{x\longrightarrow-\infty}\left(\tfrac{4-\frac{1}{x^2}}{1+\frac{2}{x}}\right)\right)&=(-\infty)\cdot4=-\infty\end{aligned} \tag{175}$$

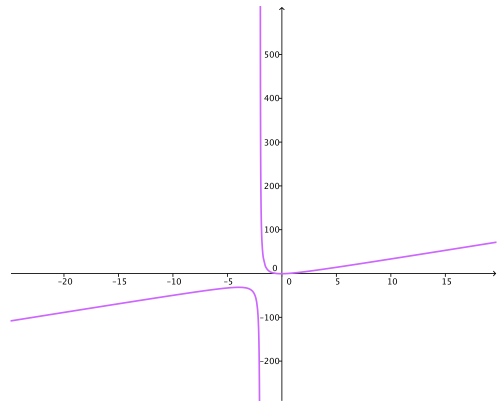


Figure 64: Graph $f(x)=\frac{4x^2-1}{x+2}$

Example 33. Find $\lim_{x\longrightarrow\infty}\frac{x^5-x^3+x-1}{x^6+2x^2+1}$

The highest power of x in the numerator is x^5 and the highest power of x in the denominator is x^6 so

$$\frac{x^5-x^3+x-1}{x^6+2x^2+1}=\frac{x^5\left(1-\frac{1}{x^2}+\frac{1}{x^4}-\frac{1}{x}\right)}{x^6\left(1+\frac{2}{x^4}+\frac{1}{x^6}\right)}=\frac{1-\frac{1}{x^2}+\frac{1}{x^4}-\frac{1}{x^5}}{x\left(1+\frac{2}{x^4}+\frac{1}{x^6}\right)}=\left(\frac{1}{x}\right)\underbrace{\left(\frac{1-\frac{1}{x^2}+\frac{1}{x^4}-\frac{1}{x^5}}{1+\frac{2}{x^4}+\frac{1}{x^6}}\right)}_{(1)} \tag{176}$$

Now, the limit of (1) is

$$\lim_{x\longrightarrow\infty}\frac{1-\frac{1}{x^2}+\frac{1}{x^4}-\frac{1}{x^5}}{1+\frac{2}{x^4}+\frac{1}{x^6}}=\frac{1-0+0-0}{1+0+0}=1 \tag{177}$$

so the entire limit is

$$\lim_{x\longrightarrow\infty}\frac{x^5-x^3+x-1}{x^6+2x^2+1}=\left(\lim_{x\longrightarrow\infty}\frac{1}{x}\right)\left(\lim_{x\longrightarrow\infty}\frac{1-\frac{1}{x^2}+\frac{1}{x^4}-\frac{1}{x^5}}{1+\frac{2}{x^4}+\frac{1}{x^6}}\right)=0\cdot1=0 \tag{178}$$

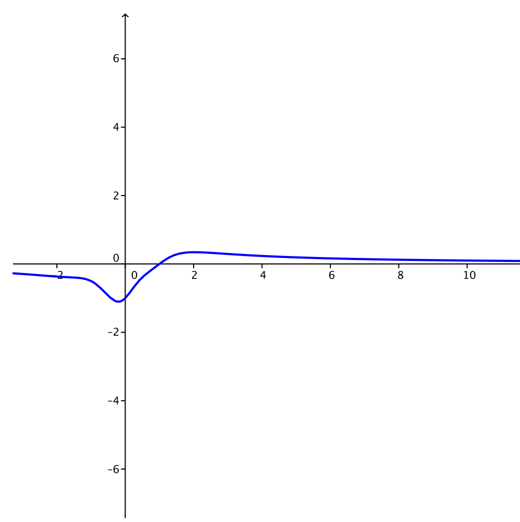


Figure 65: Graph of $f(x)=\frac{x^5-x^3+x-1}{x^6+2x^2+1}$

Example 34. Find $\lim_{x\longrightarrow\infty}\frac{\sqrt{x^2+1}}{3x-4}$

In this case we proceed as before, that is, we factorize the highest power in the numerator inside the square root

$$\sqrt{x^2+1}=\sqrt{x^2\left(1+\frac{1}{x^2}\right)}=x\sqrt{1+\frac{1}{x^2}} \tag{179}$$

and for the denominator we have

$$3x - 4 = x\left(3 - \frac{4}{x}\right) \tag{180}$$

Therefore we obtain

$$\lim_{x\longrightarrow\infty} \frac{\sqrt{x^2+1}}{3x-4} = \lim_{x\longrightarrow\infty} \frac{x\sqrt{1+\frac{1}{x^2}}}{x\left(3-\frac{4}{x}\right)} = \frac{\lim_{x\longrightarrow\infty}\sqrt{1+\frac{1}{x^2}}}{\lim_{x\longrightarrow\infty}\left(3-\frac{4}{x}\right)} = \frac{1}{3} \tag{181}$$

Example 35. Find $\lim_{x\longrightarrow-\infty} \frac{2x+1}{\sqrt{x^2-3}}$

This time we have to be more careful when we take x^2 outside the square root since x is taking negative values. We must use the property

$$\sqrt{x^2} = |x| \tag{182}$$

so that

$$\sqrt{x^2-3} = \sqrt{x^2\left(1-\frac{3}{x^2}\right)} = |x|\sqrt{1-\frac{3}{x^2}} = -x\sqrt{1-\frac{3}{x^2}} \tag{183}$$

where we used in the last step the fact that we only care about negative values of x. Therefore

$$\lim_{x\longrightarrow-\infty} \frac{2x+1}{\sqrt{x^2-3}} = \lim_{x\longrightarrow-\infty} \frac{x\left(2+\frac{1}{x}\right)}{-x\sqrt{1-\frac{3}{x^2}}} = \lim_{x\longrightarrow-\infty} -\frac{2+\frac{1}{x}}{\sqrt{1-\frac{3}{x^2}}} = -2 \tag{184}$$

Defining Limits and the Squeeze Theorem

Now that we have found some examples of limits it is time to give a more precise definition of what is meant by a limit. This limit definition makes it possible to show the properties of limits that have been used so far and it will allow us to prove the Squeeze Theorem, which turns to be very useful to limits which can't be found with the algebraic techniques we used in previous examples.

Suppose that one believes that

$$\lim_{x \longrightarrow a} f(x) = L \tag{185}$$

and we want to give some criteria that guarantees that a limit does in fact exist. First of all, if the limit does in fact exist, then 185 can be rewritten as

$$\lim_{x \longrightarrow a} (f(x) - L) = 0 \tag{186}$$

If we think of the values of $f(x)$ as our output and L as our "measured" value, then $f(x) - L$ represents some sort of error and 186 says that the error is shrinking to 0 as x approaches a. Given that 186 is equivalent to

$$\lim_{x \longrightarrow a} (L - f(x)) = 0 \tag{187}$$

it should not matter in which order we take the subtraction of L with $f(x)$ so we should work with $|f(x) - L|$. Also, saying that x approaches a is the same as saying that $x - a$ is approaching 0 and if we consider again $x - a$ as a sort of error between the input x measured and the actual input a we want to measure then we should work with $|x - a|$.

It is natural to think that we care about the error $|f(x) - L|$ we make in our measurement, so the limit definition will require that no matter which accuracy we desire in our output it is always possible to achieve it by making sufficiently precise measurements of our input.

Definition of Limit: we say that

$$\lim_{x \longrightarrow a} f(x) = L \tag{188}$$

if and only if

> For every number $\varepsilon > 0$ there exists a number δ with the following property: for any x that satisfies $0 < |x - a| < \delta$ we have $|f(x) - L| < \varepsilon$

Some remarks are in order:

1. Here ε is the error in the output we are willing to tolerate. Since the statement is for every $\varepsilon > 0$ this means that we have the intention of keep decreasing the error made in our output measurements, however, we don't try to reduce the error instantaneously to 0, which is why we don't allow $\varepsilon = 0$. Therefore, there must always be some error allowed although it can be made as small as we want.

2. We can think of ε as an error chosen *a priori* (before the measurement) while δ is an error chosen *a posteriori* which will guarantee that as long as the input x we measured is within an error δ of a then our error in the output will be less than the error ε we were willing to tolerate.

3. Observe that we don't allow x to be equal to a because we have seen situations in which the function f is not defined at a: in fact, one of the reasons limits are interesting is because of this situation.

4. Observe that we allow the possibility for $f(x)$ to be equal to L; for example, this is the case when $f(x)$ is a constant function like $f(x) = 3$.

5. In general the number δ that is found will depend on a, ε (and obviously $f(x)$) so we can write $\delta = \delta(a, \varepsilon, f(x))$ to indicate explicitly its dependence on these quantities. [21]

[21] The notation $\delta = \delta(a, \varepsilon, f(x))$ suggests that δ should be thought of as some *function* of a, ϵ and $f(x)$. This idea could be made more precise, however, for our purposes it suffices to think that for many of the concrete functions we will deal with, δ can be written as some expression involving ϵ and the point a.

Example 36. Show, using the epsilon-delta definition that $\lim_{x\longrightarrow 2}(-3x+1) = -5$

Here we have $a = 2$, $L = -5$ and $f(x) = -3x+1$. Before doing the complete proof that the limit is -5 let's start with the case in which $\varepsilon = 1$, that is, the error we are willing to tolerate is 1.

To find the value δ let's find first what is the relationship between the error in the output and the error in the input

$$\begin{aligned} |f(x) - L| &= |-3x+1-(-5)| \\ &= |-3x+6| \\ &= |-3(x-2)| \\ &= 3|x-2| \end{aligned} \tag{189}$$

so the error in the output is three times the error in the input. Since we want $|f(x) - L| < 1$ we need to guarantee that $3|x-2| < 1$ or $|x-2| < \frac{1}{3}$, that is, our error in the input must be less than $\frac{1}{3}$. Any number less or equal to $\frac{1}{3}$ will work, for example, $\delta = \frac{1}{6}$ or $\delta = \frac{1}{3}$.

Now, we need to do this not just for $\varepsilon = 1$, but for $\varepsilon = \frac{1}{2}$, $\varepsilon = \frac{1}{9}$, etc. Since there are infinitely many numbers for which this has to be verified we no longer fix a specific value of ε but work with it as a variable. The calculation in 189 still works so we now have to guarantee that

$$3|x-2| < \varepsilon \tag{190}$$

Which is the same as

$$|x-2| < \frac{\varepsilon}{3} \tag{191}$$

If we take $\delta = \frac{\varepsilon}{3}$ given that $|x-a|$ will be less than δ we can see that 191 will be satisfied so we have shown that $\lim_{x \longrightarrow 2} (-3x+1) = -5$

Example 37. Show, using the epsilon-delta definition that $\lim_{x \longrightarrow 1} x^2 = 1$

We fix an $\varepsilon > 0$ and consider the output error

$$\begin{aligned} |f(x)-L| &= \left|x^2-1\right| \\ &= |(x-1)(x+1)| \\ &= |x-1|\,|x+1| \end{aligned} \tag{192}$$

In this case the relationship between the input error and the output error is more complicated, since now they don't differ by a constant factor. To solve this problem, we will do something which seems more complicated but actually makes everything work out better: we are going to control $|f(x)-L|$ by an auxiliary quantity and this auxiliary quantity will then be made smaller than ε.

To see how this is done, observe that since x approaches 1, x will eventually be sufficiently close to 1 so in fact we may assume that x is positive. This allows us to write 192 as

$$|f(x)-L| = |x-1|\,(x+1) \tag{193}$$

Moreover, we can even ask for x to satisfy the inequality

$$\frac{1}{2} < x < \frac{3}{2} \tag{194}$$

If we do this then

$$\frac{3}{2} < x+1 < \frac{5}{2} \tag{195}$$

and so on the region $\frac{1}{2} < x < \frac{3}{2}$ 193 can be controlled by

$$|f(x)-L| < \frac{5}{2}\,|x-1| \tag{196}$$

The term $\frac{5}{2}\,|x-1|$ is the auxiliary quantity we were trying to find since this one is more easily related to the input error. Therefore, it suffices to guarantee

$$\frac{5}{2}\,|x-1| < \varepsilon \tag{197}$$

which is the same as

$$|x-1| < \frac{2}{5}\varepsilon \tag{198}$$

Clearly $\delta = \frac{2\varepsilon}{5}$ will do the trick [22] and we have shown that $\lim_{x \longrightarrow 1} x^2 = 1$.

[22] Strictly speaking we should write $\delta = \min\left(\frac{2\varepsilon}{5}, \frac{1}{2}\right)$ to guarantee that x belongs to the region $\frac{1}{2} < x < \frac{3}{2}$ but sometime it can be omitted if the context makes it clear

The last example also shows that to prove a limit it is convenient to bound a function by another function for which it is easier to show that a limit exists. A precise version of this technique is the so called Squeeze Theorem: [23]

[23] it is also known as the pinching theorem or the sandwich theorem

Squeeze Theorem: Suppose that $f(x), g(x), h(x)$ are three functions defined on a deleted neighborhood of a that satisfy

$$f(x) \leq g(x) \leq h(x) \tag{199}$$

on such a deleted neighborhood. In addition suppose that

$$\lim_{x \longrightarrow a} f(x) = \lim_{x \longrightarrow a} h(x) = L \tag{200}$$

Then

$$\lim_{x \longrightarrow a} g(x) = L \tag{201}$$

Proof. The existence of the same limit for $f(x)$ and $h(x)$ imply that

1. For every number $\varepsilon > 0$ there exists a number $\delta = \delta(a, \varepsilon, f(x))$ with the following property: for any x that satisfies $0 < |x - a| < \delta$ we have $|f(x) - L| < \varepsilon$

2. For every number $\varepsilon > 0$ there exists a number $\delta = \delta(a, \varepsilon, h(x))$ with the following property: for any x that satisfies $0 < |x - a| < \delta$ we have $|h(x) - L| < \varepsilon$

We need to show

3. For every number $\varepsilon > 0$ there exists a number $\delta = \delta(a, \varepsilon, g(x))$ with the following property: for any x that satisfies $0 < |x - a| < \delta$ we have $|g(x) - L| < \varepsilon$

Therefore, take $\varepsilon > 0$. We need to relate $|g(x) - L|$ to $|f(x) - L|$ and $|h(x) - L|$. By the "squeezing" hypothesis 199 we have that

$$f(x) - L \leq g(x) - L \leq h(x) - L \tag{202}$$

and when we take absolute values in the previous inequalities we can say that $|g(x) - L| \leq |h(x) - L|$ or $|g(x) - L| \leq |f(x) - L|$ [24]. In any case, we can guarantee that

[24] To see this consider for example what happens if $-3 \leq g(x) - L \leq 1$ or $2 \leq g(x) - L \leq 5$

$$|g(x) - L| \leq \max\{|f(x) - L|, |h(x) - L|\} \tag{203}$$

If we use 1. we can find $\delta_1 = \delta_1(a, \varepsilon, f(x))$ such that $|f(x) - L| < \varepsilon$. If we use 2. we can find $\delta_2 = \delta_2(a, \varepsilon, h(x))$ such that $|h(x) - L| < \varepsilon$. [25] To show 3. we just need to take

[25] The only thing we have to be careful about is that there is a different delta for each function.

$$\delta = \delta_3 = \min\{\delta_1, \delta_2\} \tag{204}$$

and by 203 the result follows. $\square$

Example 38. Show, using the Squeeze Theorem that

$$\lim_{x\longrightarrow 0} \sin x = 0$$
$$\lim_{x\longrightarrow 0} (1-\cos x) = 0 \qquad (205)$$

If we consider the following figure it can be seen that for x inside a sufficiently small neighborhood of 0 we have [26]

$$0 \leq |\sin x| \leq \sqrt{2}\sqrt{1-\cos(x)} \leq |x| \qquad (206)$$

Therefore, we have the two inequalities [27]

$$0 \leq |\sin x| \leq |x|$$
$$0 \leq 1-\cos(x) \leq \frac{x^2}{2} \qquad (207)$$

[26] Here we use the fact that $\sin(x)$ is an odd function, that is, $\sin(-x) = -\sin(x)$ while $\cos(x)$ is an even function, that is, $\cos(x) = \cos(-x)$

[27] For the second inequality we use that $x^2 = |x|^2$ and that $0 \leq a \leq b$ if and only if $0 \leq a^2 \leq b^2$

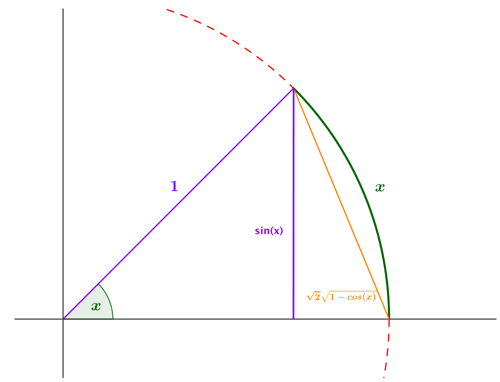


Figure 66: Example Squeeze Theorem

We apply the Squeeze Theorem twice. For the first inequality we use $f(x) = 0$, $h(x) = |x|$, $g(x) = |\sin x|$ and the fact that $\lim_{x\longrightarrow 0} 0 = \lim_{x\longrightarrow 0} |x| = 0$ to conclude that

$$\lim_{x\longrightarrow 0} |\sin x| = 0 \qquad (208)$$

Now, it is very easy to show that

$\lim_{x\longrightarrow a} f(x) = 0$ if and only if $\lim_{x\longrightarrow a} |f(x)| = 0$

so it follows that

$$\lim_{x\longrightarrow 0} \sin x = 0 \qquad (209)$$

For the second application of the Squeeze Theorem we take $f(x) = 0$, $h(x) = \frac{x^2}{2}$, $g(x) = 1-\cos(x)$ and the fact that $\lim_{x\longrightarrow 0} 0 = \lim_{x\longrightarrow 0} \frac{x^2}{2} = 0$ to conclude that

$$\lim_{x\longrightarrow 0} (1-\cos(x)) = 0 \qquad (210)$$

Example 39. Use the Squeeze Theorem to show that

$$\lim_{x\longrightarrow 0} \frac{\sin x}{x} = 1 \qquad (211)$$

and use this limit to show that

$$\lim_{x\longrightarrow 0} \frac{1-\cos x}{x} = 0 \qquad (212)$$

We now use the following figure and the fact that the area of a circular sector of angle θ is $\frac{r^2\theta}{2}$. For x in a sufficiently small deleted neighborhood of 0 we have that

$$\text{area}\,(OAB) \leq \text{area}\,(ODB) \leq \text{area}\,(ODC) \qquad (213)$$

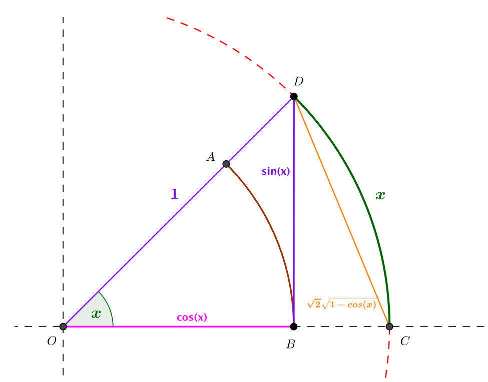


Figure 67: Example Squeeze Theorem

which gives

$$(\cos x)^2 \frac{|x|}{2} \leq \frac{|\sin x|\,(\cos x)}{2} \leq \frac{|x|}{2} \tag{214}$$

If we divide the inequality by $(\cos x)\,\frac{|x|}{2}$ that

$$(\cos x) \leq \frac{|\sin x|}{|x|} \leq \frac{1}{\cos x} \tag{215}$$

It is very easy to check that $\frac{|\sin x|}{|x|} = \frac{\sin x}{x}$ so the inequality becomes

$$\cos x \leq \frac{\sin x}{x} \leq \frac{1}{\cos x} \tag{216}$$

To apply the Squeeze Theorem we use $f(x) = \cos x$, $h(x) = \frac{1}{\cos x}$, $g(x) = \frac{\sin x}{x}$ and the fact that $\lim_{x \longrightarrow 0} \cos x = \lim_{x \longrightarrow 0} \frac{1}{\cos x} = 1$ to conclude that [28] [29]

$$\lim_{x \longrightarrow 0} \frac{\sin x}{x} = 1 \tag{217}$$

To find $\lim_{x \longrightarrow 0} \frac{1-\cos x}{x}$ we multiply by the conjugate of $1 - \cos x$

$$\begin{aligned}
\lim_{x \longrightarrow 0} \frac{1-\cos x}{x} &= \lim_{x \longrightarrow 0} \left(\frac{1-\cos x}{x}\right)\left(\frac{1+\cos x}{1+\cos x}\right) \\
&= \lim_{x \longrightarrow 0} \left(\frac{1-\cos^2 x}{x}\right)\left(\frac{1}{1+\cos x}\right) \\
&= \lim_{x \longrightarrow 0} \left(\frac{\sin^2 x}{x}\right)\left(\frac{1}{1+\cos x}\right) \\
&= \left(\lim_{x \longrightarrow 0} \frac{\sin x}{x}\right) \frac{(\lim_{x \longrightarrow 0} \sin x)}{\lim_{x \longrightarrow 0}(1+\cos x)} \\
&= 1 \cdot \frac{0}{2} \\
&= 0
\end{aligned} \tag{218}$$

Therefore,

$$\lim_{x \longrightarrow 0} \frac{1-\cos x}{x} = 0 \tag{219}$$

[28] Observe that up to this point the limits were found basically by doing some algebraic manipulation but that is not the case here. This shows that the Squeeze theorem can be a very powerful techinique

[29] We used the previous limit $\lim_{x \longrightarrow 0} (1 - \cos x) = 0$ to conclude that $\lim_{x \longrightarrow 0} \cos(x) = 1$

More Trigonometric Limits

The following examples show how to use the Squeeze Theorem and the limits

$$\begin{aligned} \lim_{x\longrightarrow 0} \frac{\sin x}{x} &= 1 \\ \lim_{x\longrightarrow 0} \frac{1-\cos x}{x} &= 0 \end{aligned} \tag{220}$$

to find certain trigonometric limits.

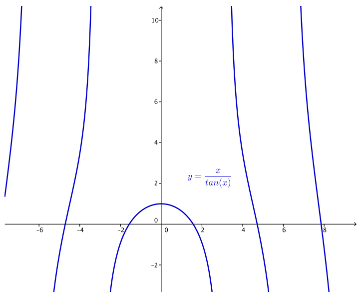


Figure 68: $f(x) = \frac{x}{\tan x}$

Example 40. Find $\lim_{x\longrightarrow 0} \frac{x}{\tan x}$

$$\begin{aligned} \lim_{x\longrightarrow 0} \frac{x}{\tan x} &= \lim_{x\longrightarrow 0} \frac{x\cos x}{\sin x} \\ &= \lim_{x\longrightarrow 0} \frac{\cos x}{\frac{\sin x}{x}} \\ &= \frac{\lim_{x\longrightarrow 0}\cos x}{\lim_{x\longrightarrow 0}\left(\frac{\sin x}{x}\right)} \\ &= \frac{1}{1} \\ &= 1 \end{aligned} \tag{221}$$

Example 41. Find $\lim_{\theta\longrightarrow \frac{\pi}{2}} \frac{\cos^2\theta}{1-\sin\theta}$

This limit only uses the identity $\sin^2\theta + \cos^2\theta = 1$ and the difference of square formulas

$$\begin{aligned} \lim_{\theta\longrightarrow \frac{\pi}{2}} \frac{\cos^2\theta}{1-\sin\theta} &= \lim_{\theta\longrightarrow \frac{\pi}{2}} \frac{1-\sin^2\theta}{1-\sin\theta} \\ &= \lim_{\theta\longrightarrow \frac{\pi}{2}} \frac{(1-\sin\theta)(1+\sin\theta)}{1-\sin\theta} \\ &= \lim_{\theta\longrightarrow \frac{\pi}{2}} (1+\sin\theta) \\ &= 1+1 \\ &= 2 \end{aligned} \tag{222}$$

Example 42. Find $\lim_{x\longrightarrow 0} \frac{\sin(3x)}{x}$

Here we make a change of variables: call $u = 3x$. Then as $x \longrightarrow 0$ we have $u \longrightarrow 0$ so

$$\begin{aligned} \lim_{x\longrightarrow 0} \frac{\sin(3x)}{x} &= \lim_{u\longrightarrow 0} \frac{\sin u}{\frac{u}{3}} \\ &= \lim_{u\longrightarrow 0} 3\frac{\sin u}{u} \\ &= 3\lim_{u\longrightarrow 0} \frac{\sin u}{u} \\ &= 3\cdot 1 \\ &= 3 \end{aligned} \tag{223}$$

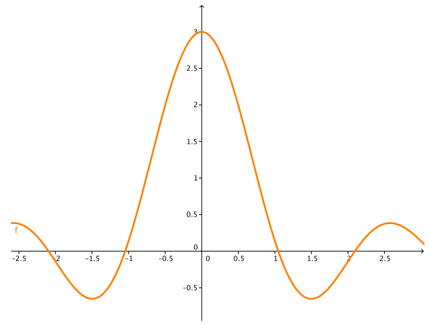


Figure 69: $f(x) = \frac{\sin(3x)}{x}$

Example 43. Find $\lim_{t\longrightarrow 0} \frac{1-\cos t}{\sin t}$

We multiply and divide by t

$$\begin{aligned} \lim_{t\longrightarrow 0} \frac{1-\cos t}{\sin t} &= \lim_{t\longrightarrow 0} \frac{\frac{1-\cos t}{t}}{\frac{\sin t}{t}} \\ &= \frac{\lim_{t\longrightarrow 0}\left(\frac{1-\cos t}{t}\right)}{\lim_{t\longrightarrow 0}\left(\frac{\sin t}{t}\right)} \\ &= \frac{0}{1} \\ &= 0 \end{aligned} \tag{224}$$

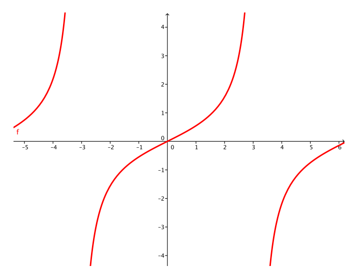


Figure 70: $f(t) = \frac{1-\cos t}{\sin t}$

Example 44. Find $\lim_{x\longrightarrow 0} x\sin\left(\frac{1}{x}\right)$

For any value of its argument, we know that

$$-1 \leq \sin\left(\frac{1}{x}\right) \leq 1 \tag{225}$$

Multiplying both sides of the inequality by x we conclude that

$$\begin{aligned} -x \leq x\sin\left(\tfrac{1}{x}\right) \leq x \quad &\text{for} \quad x > 0 \\ -x \geq x\sin\left(\tfrac{1}{x}\right) \geq x \quad &\text{for} \quad x < 0 \end{aligned} \tag{226}$$

Regardless of the sign of x, since $\lim_{x\longrightarrow 0} x = \lim_{x\longrightarrow 0} -x = 0$ we can see using the Squeeze Theorem that

$$\lim_{x\longrightarrow 0} x\sin\left(\frac{1}{x}\right) = 0 \tag{227}$$

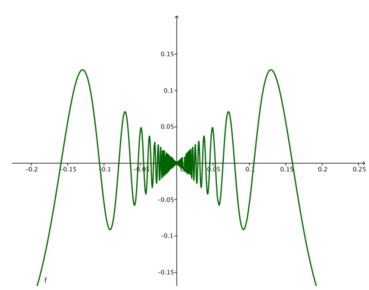


Figure 71: $f(x) = x\sin\left(\frac{1}{x}\right)$

Example 45. Find $\lim_{x\longrightarrow\infty} \frac{2-\cos x}{x+3}$

For any value of its argument, we have

$$-1 \leq -\cos x \leq 1 \tag{228}$$

Therefore, if we add 2 to the inequality and then divide by $x+3$, which is positive since x approaches infinity, we have

$$\frac{1}{x+3} \leq \frac{2-\cos x}{x+3} \leq \frac{3}{x+3} \tag{229}$$

We can use the Squeeze Theorem, which is also valid for limits at infinity so

$$\lim_{x \longrightarrow \infty} \frac{1}{x+3} \leq \lim_{x \longrightarrow \infty} \frac{2-\cos x}{x+3} \leq \lim_{x \longrightarrow \infty} \frac{3}{x+3} \tag{230}$$

Therefore

$$\lim_{x \longrightarrow \infty} \frac{2-\cos x}{x+3} = 0 \tag{231}$$

Example 46. Find $\lim_{x \longrightarrow -\infty} \frac{x \arctan x}{\sqrt{16x^2+6}}$

Using that $\sqrt{x^2} = -x$ for x negative we find:

$$\begin{aligned}
\lim_{x \longrightarrow -\infty} \frac{x \arctan x}{\sqrt{16x^2+6}} &= \lim_{x \longrightarrow -\infty} \frac{x \arctan x}{\sqrt{x^2\left(16+\frac{6}{x^2}\right)}} \\
&= \lim_{x \longrightarrow -\infty} \frac{x \arctan x}{(-x)\sqrt{\left(16+\frac{6}{x^2}\right)}} \\
&= \lim_{x \longrightarrow -\infty} -\frac{\arctan x}{\sqrt{\left(16+\frac{6}{x^2}\right)}} \\
&= -\frac{\lim_{x \longrightarrow -\infty} \arctan x}{\lim_{x \longrightarrow -\infty} \sqrt{\left(16+\frac{6}{x^2}\right)}} \\
&= -\frac{1}{4} \lim_{x \longrightarrow -\infty} \arctan x
\end{aligned} \tag{232}$$

Therefore, we just need to find $\lim_{x \longrightarrow -\infty} \arctan x$. Call $y = \arctan x$. Then $\tan y = x$ or

$$\frac{\sin y}{\cos y} = x \tag{233}$$

By definition of $\arctan x$, $y \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and since $\sin y$ is bounded between -1 and 1 the only way for x to approach $-\infty$ is if $\cos y$ approaches 0 from the left so we must have y approaches $-\frac{\pi}{2}$. Therefore, $\lim_{x \longrightarrow -\infty} \arctan x = -\frac{\pi}{2}$ and so

$$\lim_{x \longrightarrow -\infty} \frac{x \arctan x}{\sqrt{16x^2+6}} = \frac{\pi}{8} \tag{234}$$

Exponential and Logarithmic Functions

The Number e

Suppose you deposit \$1200 on a savings account earning interest at the rate of 100% per year (this is a very generous bank). After one year you will receive \$1200 in interest so you will end up with \$2400. Now, let's suppose that the bank decides to divide the 100% interest equally into each month of the year, therefore, it will give you $\frac{100}{12}$% of your money after each month. If we call M_0 the initial money (so $M_0 = 1200$) and M_1 the money we have after the first month since we earned \$100 in interest in the first month we have that

$$M_1 = M_0 + \frac{1}{12}M_0 = \left(1+\frac{1}{12}\right) M_0 = \left(1+\frac{1}{12}\right) 1200 = 1300 \quad (235)$$

To make things interesting, let's assume that for the second month, the bank will give you another $\frac{100}{12}$% of interest based on the money at the end of the first month, not the initial money (this is call compounded interest and will be discussed later in greater detail). So we are going to use M_1 to calculate the interest and the total money we have at the end of the second month is

$$M_2 = \left(1+\frac{1}{12}\right) M_1 = \left(1+\frac{1}{12}\right)^2 M_0 = \frac{4225}{3} \simeq 1408 \quad (236)$$

Now the pattern is clear so at the end of the year we will end up with

$$M_{12} = \left(1+\frac{1}{12}\right)^{12} M_0 \simeq 3135 \quad (237)$$

so under this method we made around 735 extra dollars.

What would happen if the bank decides to divide the 100% interest in days rather than months? We can see that we can use the same formula as in 237, but we only need to change $\frac{1}{12}$ by $\frac{1}{365}$

$$M_{365} = \left(1+\frac{1}{365}\right)^{365} M_0 \simeq 3257 \quad (238)$$

so under this interest calculation we made more money but the difference is not as big as before. Now the natural question becomes, what happens if the bank decides to divide the year n times? (for example, $n = 12$ would correspond to months, $n = 365$ to days and

$n = 31,557,600$ to seconds). Then we simply would change 237 into

$$M_n = \left(1 + \frac{1}{n}\right)^n M_0 \tag{239}$$

we have already seen that if we increase n from 1 to 12 or from 12 to 365 then M_n becomes bigger. However, we have also seen that the difference earned from 12 to 365 is not as big as the difference earned from 1 to 12 so it is not clear which of these two tendencies will win as n becomes larger and larger. From what we have seen so far the question we are asking is whether or not

$$\lim_{n \longrightarrow \infty} \left(1 + \frac{1}{n}\right)^n \tag{240}$$

exists. The answer is that the limit 240 exists and the value of the limit turns out to be one of the most important limits in all of mathematics!

The number e is the limit

$$e = \lim_{n \longrightarrow \infty} \left(1 + \frac{1}{n}\right)^n \tag{241}$$

approximately, $e \simeq 2.71828$.

The Exponential Function e^x

In the first chapter we discussed the power of a number, for example, 2^3, 5^{-2}, $\pi^{\frac{1}{2}}$. In the case of a negative number, we might happen that its powers are not always defined, for example, $(-5)^{\frac{1}{2}}$ is not defined. Therefore, we will focus on numbers of the form b^x with $b > 0$ so that b^x always make sense regardless of the value of the power. Moreover, we will exclude the case $b = 1$ because $1^x = 1$ always so its behavior is not interesting.

If b is a positive number different from 1, the exponential function with base b is

$$f(x) = b^x \tag{242}$$

The number x is called the exponent.

- The domain of b^x is the real line $\mathbb{R}$
- The range of b^x is the set $(0, \infty)$
- The case in which the base is e will be extremely important and it is denoted e^x or $\exp(x)$

Just as a reminder we have the following properties for the exponential functions

$$b^x \cdot b^y = b^{x+y} \qquad (b^x)^y = b^{xy} \qquad \left(\frac{a}{b}\right)^x = \frac{a^x}{b^x} \qquad \frac{b^x}{b^y} = b^{x-y} \qquad (ab)^x = a^x b^x \tag{243}$$

The following examples illustrate that the behavior of b^x depends on the base b being greater or smaller than 1.

From the examples we observe the following properties of the exponential

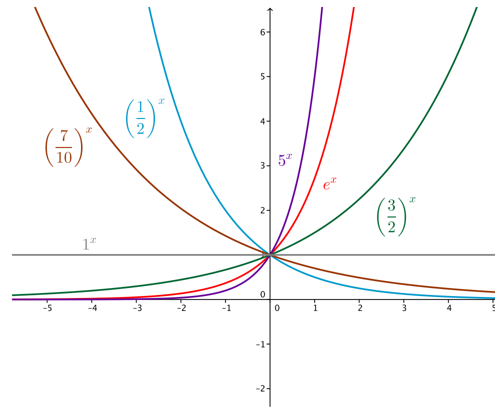


Figure 72: Example Exponential Functions

- All the exponentials intersect the y axis at the point $(0, 1)$
- If $b > 1$ the exponential function b^x is increasing. Moreover, $\lim_{x\longrightarrow\infty} b^x = \infty$ and $\lim_{x\longrightarrow-\infty} b^x = 0$
- If $0 < b < 1$ the exponential function b^x is decreasing. Moreover, $\lim_{x\longrightarrow\infty} b^x = 0$ and $\lim_{x\longrightarrow-\infty} b^x = 0$

Example 47. Solve the equation $9^{x+1} = \frac{1}{27^x}$

The key idea is to write 9 and 27 as powers of 3 so that the equation becomes

$$\left(3^2\right)^{x+1} = \frac{1}{\left(3^3\right)^x} \tag{244}$$

By the properties of the exponential this is the same as

$$3^{2x+2} = \frac{1}{3^{3x}} \tag{245}$$

and we can send 3^{3x} to multiply on the left side of the equation

$$3^{2x+2} \cdot 3^{3x} = 1 \tag{246}$$

Again, using properties of the exponential we get

$$3^{5x+2} = 1 \tag{247}$$

The only way for this equation to be true is if $5x + 2 = 0$ (because the exponential intersects the y axis only when the exponent is 0) so we obtain

$$x = -\frac{2}{5} \tag{248}$$

Problem 48. Solve the equation $4^{3x+2} = \left(\frac{1}{16}\right)^x$

We can write this equation as

$$4^{3x+2} = \left(\frac{1}{4^2}\right)^x \tag{249}$$

which again is equivalent to

$$4^{3x+2} \cdot 4^{2x} = 1 \tag{250}$$

and therefore we need to solve

$$5x + 2 = 1 \tag{251}$$

so x must be

$$x = -\frac{1}{5} \tag{252}$$

The Logarithmic Function

Consider the function $y = 2^x$. If we start with a value x in the domain (for example, $x = 1$ or $x = 2$) we have a value y associated to it (for example, $y = 2$ or $y = 4$). From the graph of the function it is clear that this association is a one to one correspondence, that is, a value x specifies y and y also specifies x (in the notation at the beginning of the course, this is the same as saying that $y = 2^x$ is injective).

Now, if we start with a positive number y, by the previous remark we can find only one x such that $2^x = y$. To emphasize the idea that x is now a function of y we write $x = \log_2 y$, so

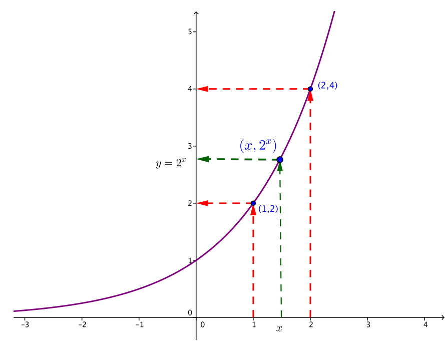


Figure 73: A value x specifies a value $y = 2^x$

- If y is a positive number ($y > 0$) the number x such that $2^x = y$ is called the logarithm (in base 2) of y and we write $x = \log_2 y$. So

$$x = \log_2 y \text{ if and only if } y = 2^x \tag{253}$$

- More generally, if y is a positive number and b is a positive number different from 1, the (unique) number x such that $b^x = y$ is called the logarithm (in base b) of y and we write $x = \log_b y$. So

$$x = \log_b y \text{ if and only if } y = b^x \tag{254}$$

- When $b = e$ the logarithm is also called the natural logarithm and we write $\ln y$. So

$$x = \ln y \text{ if and only if } y = e^x \tag{255}$$

- When $b = 10$ we usually write $\log y$ instead of $\log_{10} y$

We recall the familiar rules of the logarithm (which follow from the rules of the exponential function)

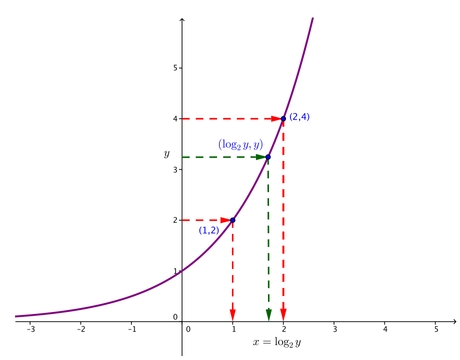


Figure 74: A value y specifies a value $x = \log_2 y$

Rules of Logarithms: If m, n are positive numbers then

- $\log_b(mn) = \log_b m + \log_b n$
- $\log_b\left(\frac{m}{n}\right) = \log_b m - \log_b n$
- $\log_b m^n = n \log_b m$
- $\log_b 1 = 0$
- $\log_b b = 1$
- $b^{\log_b x} = x$

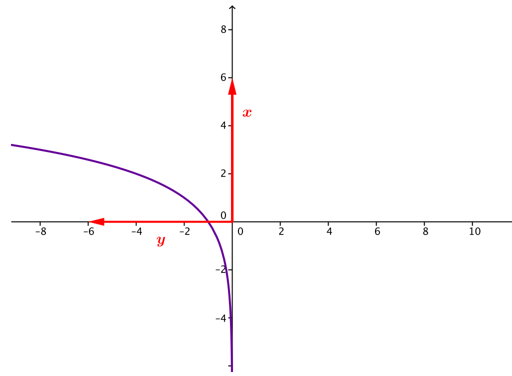


Figure 75: Obtaining the logarithm from the exponential function

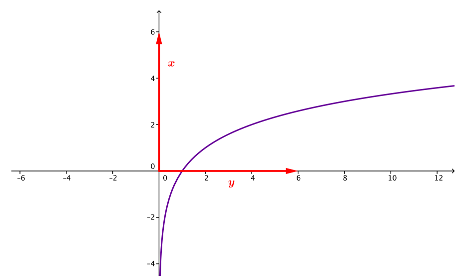


Figure 76: Obtaining the logarithm from the exponential function

Example 49. Use the properties of logarithms to simplify the following expression $\ln\left(te^{-t^2}\right)$

We use the property that the logarithm turns products into sums

$$\ln\left(te^{-t^2}\right) = \ln(t) + \ln\left(e^{-t^2}\right) \tag{256}$$

Now we use the property that the exponent of a number can be taken outside the logarithm

$$\ln\left(e^{-t^2}\right) = -t^2 \ln e = -t^2 \tag{257}$$

where we used in the last equality the property $\ln e = 1$. Therefore,

$$\ln\left(te^{-t^2}\right) = \ln t - t^2 \tag{258}$$

Example 50. Use the properties of logarithms to simplify $\log\left(\frac{\sqrt{u+1}}{u^2+1}\right)$

We use the property that the logarithm turns fractions into subtraction

$$\log \frac{\sqrt{u+1}}{u^2+1} = \log(u+1)^{\frac{1}{2}} - \log\left(u^2+1\right) = \frac{1}{2}\log(u+1) - \log\left(u^2+1\right) \tag{259}$$

Just as we defined the function $y(x) = b^x$, we can define the function $y(x) = \log_b x$. For example, if $b = 2$ then $y(1) = 0$, $y(2) = 1$ and $y(4) = 2$. As with the exponential functions, the general behavior of $\log_b x$ depends on whether or not $b > 1$

From the examples we can see the following properties of the logarithmic functions

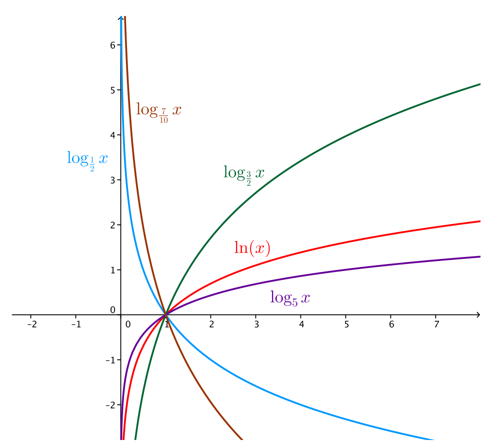


Figure 77: Example Logarithmic Functions

- The domain of $y = \log_b x$ is the interval $(0, \infty)$
- The range of $y = \log_b x$ is $\mathbb{R}$
- All the logarithms intersect the x axis at the point $(1, 0)$
- If $b > 1$ the logarithmic function $\log_b x$ is increasing. Moreover, $\lim_{x \longrightarrow \infty} \log_b x = \infty$ and $\lim_{x \longrightarrow 0^+} \log_b x = -\infty$
- If $0 < b < 1$ the logarithmic function $\log_b x$ is decreasing. Moreover, $\lim_{x \longrightarrow \infty} \log_b x = -\infty$ and $\lim_{x \longrightarrow 0^+} \log_b x = \infty$

Example 51. Solve the equation $\frac{1}{3^{4x}} = 9^{2x-4}$

Using that $9 = 3^2$ the equation $\frac{1}{3^{4x}} = 9^{2x-4}$ is equivalent to

$$3^{-4x} = 3^{4x-8} \tag{260}$$

we can equate the exponents so

$$-4x = 4x - 8 \tag{261}$$

which gives $x = 1$.

Example 52. Solve the equation $2 \ln (x + 3) = \ln (2x + 14)$

By the properties of the logarithm

$$2 \ln (x + 3) = \ln (x + 3)^2 \tag{262}$$

so we need to solve

$$\ln (x + 3)^2 = \ln (2x + 14) \tag{263}$$

we can set equal the argument of the logarithms

$$(x + 3)^2 = 2x + 14 \tag{264}$$

and the solutions of the previous quadratic equation are $x = -5$ and $x = 1$. However, $x = -5$ does not work because $\ln (-5 + 3) = \ln (-2)$ is not defined so $x = 1$ is the only solution. [30]

Example 53. Solve the equation $\log \left(10x^2 + 30x\right) = 2$

Since the logarithm is base 10 we can exponentiate both sides by 10 so that

$$10^{\log\left(10x^2+30x\right)} = 10^2 \tag{265}$$

and by the cancellation property this is the same as

$$10x^2 + 30x = 100 \tag{266}$$

[30] This is a reminder than when you solve an equation using an algebraic procedure (like taking log or squaring an equation) you must check that your solution(s) solves the equation you started with. For example, the equation $x^2 = -1$ has no solution but someone might mistakingly think it has a solution because he/she could have squared both sides of the equation to obtain $x^4 = 1$ and this equation has solutions $x = \pm 1$.

Therefore we need to solve

$$x^2 + 3x - 10 = 0 \tag{267}$$

which gives $x = 2$ and $x = -5$. However, we also need $10x^2 + 30x > 0$ because $\log\left(10x^2 + 30x\right)$ is well defined only when the argument is positive. Therefore only $x = 2$ works.

Example 54. Solve the equation $\ln(x) - \ln(x-1) = \frac{1}{2}$

By the properties of the logarithm this equation is the same as

$$\ln\left(\frac{x}{x-1}\right) = \frac{1}{2} \tag{268}$$

We can exponentiate both sides of the equation

$$e^{\ln\left(\frac{x}{x-1}\right)} = e^{\frac{1}{2}} \tag{269}$$

And by the cancellation property

$$\frac{x}{x-1} = e^{\frac{1}{2}} \tag{270}$$

which is the same as

$$x = \sqrt{e}x - \sqrt{e} \tag{271}$$

And therefore

$$x = \frac{\sqrt{e}}{\sqrt{e}-1} \tag{272}$$

Example 55. Solve the equation $x^x = \sqrt{x}$

Writing $\sqrt{x} = x^{\frac{1}{2}}$ and takin natural logarithm on both sides

$$x \ln x = \frac{1}{2} \ln x \tag{273}$$

which gives

$$\left(x - \frac{1}{2}\right) \ln x = 0 \tag{274}$$

therefore $x = \frac{1}{2}$ or $x = 1$ are the two solutions.

Example 56. Find $\lim_{x \longrightarrow 2} \frac{xe^x - 2e^x}{x^2 - x - 2}$

We have that

$$\begin{aligned} \lim_{x \longrightarrow 2} \frac{xe^x - 2e^x}{x^2 - x - 2} &= \lim_{x \longrightarrow 2} \frac{e^x(x-2)}{(x-2)(x+1)} \\ &= \lim_{x \longrightarrow 2} \frac{e^x}{x+1} \\ &= \frac{e^2}{3} \end{aligned} \tag{275}$$

Hyperbolic Functions

Another important class of functions are the hyperbolic functions. As we will see in a moment they have similar properties to the ones the trigonometric functions share, so the names $\sinh x$, $\cosh x$, etc. are used to denote them. [31]

We will introduce them in such a way that makes their geometric definitions analogous. To define the trigonometric functions we used the unit circle $x^2 + y^2 = 1$. If Q is a point in the unit circle then it is easy to see that we can write Q as $Q = (x, y) = (\cos\theta, \sin\theta)$, that is, the trigonometric functions cosine and sine can be used to specify the points on the unit circle.

[31] If you know complex numbers then there is a closer relationship between them, namely, $\cos x = \cosh ix$ and $\sin x = -i \sinh ix$

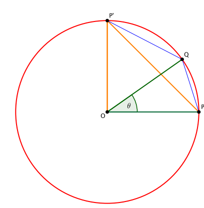


Figure 78: Angle and Area

In a similar way, if we start with the unit hyperbola $x^2 - y^2 = 1$ we want to say that if we choose a point Q on the hyperbola then we can write it as $Q = (x, y) = (\cosh t, \sinh t)$ where the variable t, in analogy with the variable θ, will be called a hyperbolic angle.

To see how this is done, recall that the variable θ has the (defining) property that $s = r\theta$, where r is the radius of a circle and s is the length of the corresponding arc. However, it is also possible to relate θ to certain areas of a circle of radius r. Comparing with the figure at the right, it is easy to check that

$$\text{area sector } OPQ = \tfrac{1}{2}r^2\theta \qquad \text{area triangle } OPP' = \tfrac{1}{2}r^2$$
$$\text{area triangle } OQP' = \tfrac{1}{2}r^2\cos\theta \qquad \text{area triangle } OPQ = \tfrac{1}{2}r^2\sin\theta \tag{276}$$

Therefore, θ, $\sin\theta$, $\cos\theta$ *could* have been defined as

$$\begin{aligned}
\theta &\equiv \frac{\text{area sector } OPQ}{\text{area triangle } OPP'} \\
\cos\theta &\equiv \frac{\text{area triangle } OQP'}{\text{area triangle } OPP'} \\
\sin\theta &\equiv \frac{\text{area triangle } OPQ}{\text{area triangle } OPP'}
\end{aligned} \tag{277}$$

Now, to define the hyperbolic functions in a similar way we will also need the conjugate hyperbola $y^2 - x^2 = 1$. In this case we take $P = (1, 0)$,

$P' = (0,1)$ and $Q = (x,y)$. We define

$$
\begin{aligned}
t &\equiv \frac{\text{area sector } OPQ}{\text{area triangle } OPP'} \\
\cosh t &\equiv \frac{\text{area triangle } OQP'}{\text{area triangle } OPP'} \qquad (278) \\
\sinh t &\equiv \frac{\text{area triangle } OPQ}{\text{area triangle } OPP'}
\end{aligned}
$$

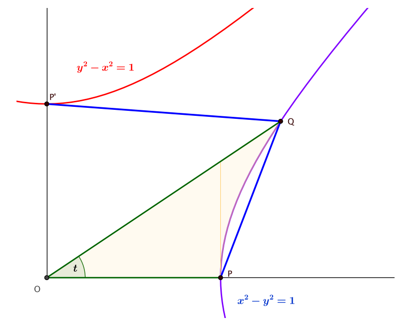


Figure 79: Definition hyperbolic functions

Using analytic geometry it is easy to see that

$$
\begin{aligned}
\text{area triangle } OQP' &= \tfrac{1}{2}x \\
\text{area triangle } OPQ &= \tfrac{1}{2}y \qquad (279) \\
\text{area triangle } OPP' &= \tfrac{1}{2}
\end{aligned}
$$

It is also possible to calculate the area of the sector OPQ, however, we would need the theory of integration to do this in an efficient way. Therefore, we will just write the result [32]

$$\text{area sector } OPQ = \frac{1}{2}\ln(x+y) \qquad (280)$$

[32] some algebra is necessary to get this answer but we won't show it here

Therefore,

$$
\begin{aligned}
t &= \ln(x+y) \\
\cosh t &= x \qquad (281) \\
\sinh t &= y
\end{aligned}
$$

If we look at the figure, we represented t as a sort of "angle". However, this should not be taken too literally since this "hyperbolic angle" has different properties from the "trigonometric angle". For example, while the standard angle θ takes values only between 0 and 2π (or $-\pi$ and π to make it more symmetric), the "hyperbolic angle" takes all possible values (that is, different values of t specify different points on the hyperbola while values of the angle θ differing by 2π represent the same point on the unit circle).

We can also notice that

$$Q = (x,y) = (\cosh t, \sinh t) \qquad (282)$$

which was our first objective. So far we have not written an explicit formula for $\cosh t$ and $\sinh t$ in terms of t. To find such a formula observe that the first equation of 281 can be written as

$$x + y = e^t \qquad (283)$$

so

$$y = e^t - x \qquad (284)$$

Since (x,y) are points on the hyperbola they satisfy the equation $x^2 - y^2 = 1$ and substituting y in this equation we obtain

$$x^2 - \left(e^t - x\right)^2 = 1 \qquad (285)$$

By the difference of square formula we obtain

$$\left(x+e^t-x\right)\left(x-e^t+x\right)=1 \tag{286}$$

which is the same as

$$x=\frac{e^t+e^{-t}}{2} \tag{287}$$

But since $x=\cosh t$ we have actually found that

$$\cosh t=\frac{e^t+e^{-t}}{2} \tag{288}$$

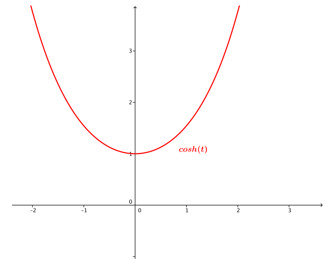


Figure 80: Graph of hyperbolic cosine $\cosh t = \frac{e^t+e^{-t}}{2}$

In a similar way we can show that

$$\sinh t=\frac{e^t-e^{-t}}{2} \tag{289}$$

With these formulas for the hyperbolic sine and cosine it is very easy to prove the following properties:

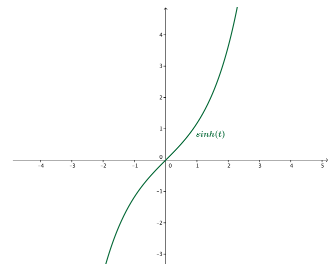


Figure 81: Graph of hyperbolic sine $\sinh t = \frac{e^t-e^{-t}}{2}$

Properties of Hyperbolic Functions:

$$\cosh t=\frac{e^t+e^{-t}}{2} \qquad \sinh t=\frac{e^t-e^{-t}}{2} \tag{290}$$

- In analogy with the trigonometric functions we can define

$$\tanh t \equiv \frac{\sinh t}{\cosh t} \qquad \operatorname{csch} t \equiv \frac{1}{\sinh t} \qquad \operatorname{sech} t \equiv \frac{1}{\cosh t} \qquad \coth t \equiv \frac{\cosh t}{\sinh t} \tag{291}$$

- $\sinh(-t)=-\sinh(t)$ and $\cosh(-t)=\cosh t$
- $(\cosh t)^2-(\sinh t)^2=1$
- $\cosh(a+b)=\cosh a\cosh b+\sinh a\sinh b$
- $\sinh(a+b)=\sinh a\cosh b+\cosh a\sinh b$

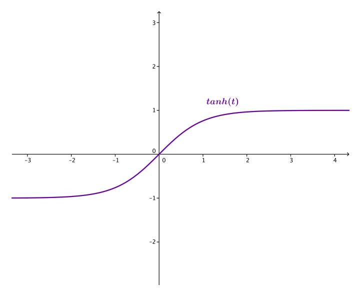


Figure 82: Graph of hyperbolic tangent $\tanh t = \frac{e^t-e^{-t}}{e^t+e^{-t}}$

In a analogous way to the trigonometric functions it is possible, by restricting the domain of the hyperbolic functions, to find inverse functions. Since these are not as important as the trigonometric inverses, we will just state the three most important ones.

Hyperbolic Function	**Domain**	**Inverse Function**	**Domain**
$\sinh t$	$\mathbb{R}$	$\operatorname{arcsinh} t=\ln\left(t+\sqrt{t^2+1}\right)$	$\mathbb{R}$
$\cosh t$	$0\le t$	$\operatorname{arccosh} t=\ln\left(t+\sqrt{t^2-1}\right)$	$1\le t$
$\tanh t$	$\mathbb{R}$	$\operatorname{arctanh} t=\frac{1}{2}\ln\left(\frac{1+t}{1-t}\right)$	$-1<t<1$

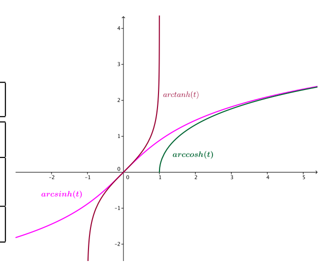


Figure 83: Graphs of $\operatorname{arcsinh} t$, $\operatorname{arccosh} t$, $\operatorname{arctanh} t$

Part III

Continuous Functions

Definition and Properties

We have seen examples in which $\lim_{x \longrightarrow a} f(x)$ exists but it is different from $f(a)$, in fact, the situation can be worse since $f(a)$ might not even exist! However, we have also seen examples for which $\lim_{x \longrightarrow a} f(x) = f(a)$. If we look carefully at the graphs of the functions for which $\lim_{x \longrightarrow a} f(x) = f(a)$ we see that in some sense they are "continuous", that is, they have no holes, gaps, jumps or breaks.

Clearly the functions for which $\lim_{x \longrightarrow a} f(x) = f(a)$ have the nice property that the value of the function at $x = a$ determine the limit and given that the graphs of those functions seem "continuous" we will say that a function is **continuous** if the previous situation holds.

Suppose that $f(x)$ is a function and $a \in \mathbb{R}$ belongs to the domain of $f(x)$. We will say that $f(x)$ is continuous at the point a if

$$\lim_{x \longrightarrow a} f(x) = f(a) \tag{292}$$

Moreover, $f(x)$ is called continuous if it is continuous at every point a of its domain.

It is possible to break up the definition of being continuous into three parts:

Criteria for Continuity: A function $f(x)$ is continuous at point $x = a$ if it satisfies the following conditions

- a must belong to the domain of $f(x)$, that is, $f(a)$ must exist
- the limit $\lim_{x \longrightarrow a} f(x)$ must exist, that is, the one-sided limits $\lim_{x \longrightarrow a^+} f(x)$ and $\lim_{x \longrightarrow a^+} f(x)$ must be the same
- the limit must coincide with the value of the function at the point, $\lim_{x \longrightarrow a} f(x) = f(a)$

We have already seen many examples of continuous functions, in particular, the properties of the limits and polynomials implies that:

Continuity and Its Properties:

- Polynomials are continuous**:** suppose that $f(x)$ is a polynomial. For example, $f(x) = x^3 - 5x + 2$. Then $f(x)$ continuous on the entire real line, which means that $\lim_{x \longrightarrow a} f(x) = f(a)$. For example

$$\lim_{x \longrightarrow 1} \left(x^3 - 5x + 2\right) = \left(1^3 - 5 \cdot 1 + 2\right) = -2 \tag{293}$$

- Radicals of polynomials are continuous: suppose that $f(x)$ is the root of some polynomial. For example, $f(x) = \sqrt{x^2+1}$. Then $f(x)$ is continuous provided a belongs to the domain of $f(x)$. For example,

$$\lim_{x \longrightarrow -1} \sqrt{x^2+1} = \sqrt{(-1)^2+1} = \sqrt{2} \tag{294}$$

- A continuous function times a constant is continuous: If $f(x)$ is continuous and c is a constant then $cf(x)$ is continuous. For example, if, $f(x) = x^2 + 1$ is continuous then $3f(x) = 3x^2 + 3$ is continuous

- The sum of continuous functions is a continuous function: If $f(x)$ and $g(x)$ are continuous functions then $f(x) \pm g(x)$ are continuous functions. For example,$g(x) = \sqrt{x+3}$ is a continuous function so $f(x) + g(x) = x^2 + 1 + \sqrt{x+3}$ is a continuous function

- The product of continuous functions is continuous: If $f(x)$ and $g(x)$ are continuous functions then $f(x)g(x)$ is a continuous functions. For example, $f(x)g(x) = (x^2+1)\left(\sqrt{x+3}\right)$ is a continuous function.

- The power of a continuous function is a continuous function: If $f(x)$ is a continuous function then $[f(x)]^r$ is a continuous function provided $[f(x)]^r$ makes sense. For example, $f(x)^{-\frac{1}{2}} = \frac{1}{\sqrt{x^2+1}}$ is a continuous function.

- The ratio of two continuous functions is continuous at the points where the denominator does not vanish: If $f(x)$ and $g(x)$ are continuous functions then $\frac{f(x)}{g(x)}$ is a continuous function at every point in which $g(x)$ is not 0. For example, $\frac{f(x)}{g(x)} = \frac{x^2+1}{\sqrt{x+3}}$ is continuous at every point in which $\sqrt{x+3}$ is not 0 , that is, $x > -3$ (we also need to exclude values for which $x+3$ is negative).

- Exponential, Logarithmic, Trigonometric and Hyperbolic Functions are continuous at every point of their domain: for example, e^x is a continuous function on $\mathbb{R}$, $\ln(x)$ is a continuous function on $(0, \infty)$, $\sin(x)$ is continuous on $\mathbb{R}$, $\tan x$ is continuous whenever $\cos x$ is different from 0, etc.

Example 57. Let $f(x) = \begin{cases} \frac{10}{x-5} & \text{if } x < 0 \\ x^3+1 & \text{if } x \geq 0 \end{cases}$. a) What is the domain of $f(x)$? b) Show that f is continuous at $x = 2$. c) Show that f is not continuous at $x = 0$.

a) To find the domain of $f(x)$ we see that for $x \geq 0$ the expression $x^3 + 1$ always makes sense, so the domain includes at least the interval $[0, \infty)$. Now, for $x < 0$, the expression $\frac{10}{x-5}$ doesn't make sense when $x - 5 = 0$ or $x = 5$, but 5 is not less than 0 so for $x < 0$, $\frac{10}{x-5}$ is always well defined so $f(x)$ has $\mathbb{R}$ as a domain.

b) To show that $f(x)$ is continuous at 2 we need to show that $\lim_{x\longrightarrow 2} f(x) = f(2)$. Now, $f(2) = 2^3 + 1 = 8 + 1 = 9$ and near 2, $f(x) = x^3 + 1$ so

$$\lim_{x\longrightarrow 2} f(x) = \lim_{x\longrightarrow 2} \left(x^3 + 1\right) = \lim_{x\longrightarrow 2} x^3 + 1 = 2^3 + 1 = 9 \tag{295}$$

so continuity is verified.

c) To show that f is not continuous at $x = 0$ we have to see that $\lim_{x\longrightarrow 0} f(x) \neq f(0)$. Now $f(0) = 0^3 + 1 = 1$. On the other hand, to find $\lim_{x\longrightarrow 0} f(x)$ we will find the one-sided limits

$$\lim_{x\longrightarrow 0^-} f(x) = \lim_{x\longrightarrow 0^-} \frac{10}{x-5} = \frac{10}{0-5} = -2 \tag{296}$$

$$\lim_{x\longrightarrow 0^+} f(x) = \lim_{x\longrightarrow 0^+} \left(x^3 + 1\right) = 0^3 + 1 = 1 \tag{297}$$

Since the one-sided limits are different $\lim_{x\longrightarrow 0} f(x)$ does not exist so it is impossible for $f(x)$ to be continuous at 0.

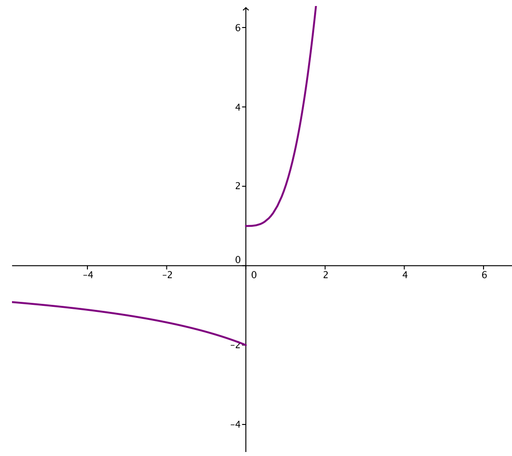


Figure 84: Graph of $f(x)$

Example 58. Let $g(x) = \begin{cases} x+2 & \text{if } x \leq 1 \\ kx^2 & \text{if } x > 1 \end{cases}$. Find the value of k that will make g continuous on $(-\infty, \infty)$

The only point in which x can fail to be continuous is at $x = 1$. Therefore, we need to find a value for k such that $\lim_{x\longrightarrow 1} g(x) = g(1)$. Now, $g(1) = 1 + 2 = 3$. To find $\lim_{x\longrightarrow 1} g(x)$ we will find the one-sided limits

$$\lim_{x\longrightarrow 1^-} g(x) = \lim_{x\longrightarrow 1^-} (x+2) = 1 + 2 = 3 \tag{298}$$

$$\lim_{x\longrightarrow 1^+} g(x) = \lim_{x\longrightarrow 1^+} kx^2 = k \tag{299}$$

We need both one-sided limits to be the same and equal to 3, therefore we must take $k = 3$

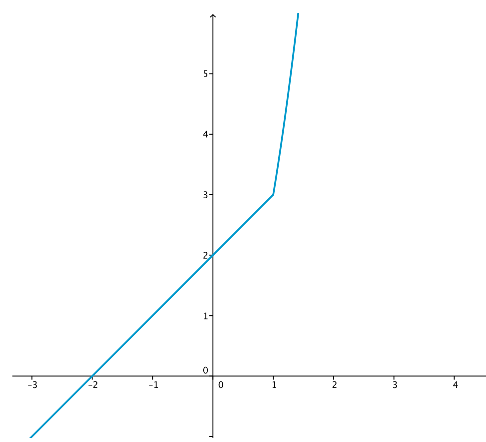


Figure 85: Graph of $f(x)$ with $k = 3$

Example 59. Let $f(x)$ be defined as $f(x) = \begin{cases} \sqrt{3x+3} & \text{if } x < 2 \\ 5 & \text{if } x = 2 \\ \frac{3}{2x-3} & \text{if } x > 2 \end{cases}$. Is $f(x)$ continuous at $x = 2$? If not, is it possible to redefine $f(2)$ so that it becomes continuous?

Given that $f(2) = 5$ we have to check if $\lim_{x \longrightarrow 2} f(x) = 5$. To do this we find the one-sided limits

$$\lim_{x \longrightarrow 2^-} f(x) = \lim_{x \longrightarrow 2^-} \sqrt{3x+3} = \sqrt{6+3} = 3 \tag{300}$$

$$\lim_{x \longrightarrow 2^+} f(x) = \lim_{x \longrightarrow 2^+} \frac{3}{2x-3} = \frac{3}{4-3} = 3 \tag{301}$$

Therefore $\lim_{x \longrightarrow 2} f(x) = 3$ but this is different from 5 so the function is not continuous at $x = 2$.

Although $f(x)$ it is not continuous; if we redefine it by changing $f(2)$ into 3 we can see that $f(x)$ would become continuous.

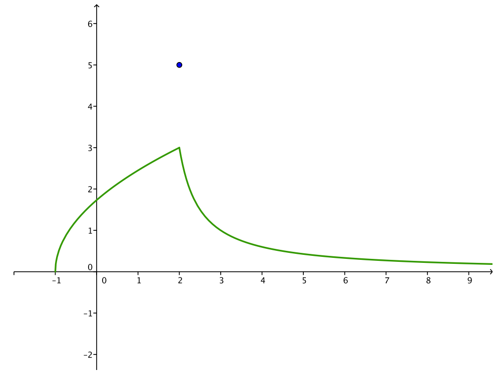


Figure 86: Graph of $f(x)$

Intermediate Value Theorem

Intuitively, a continuous function is a function whose graph has no holes or gaps as we said before. Therefore, it should sound plausible that if a continuous function takes two specific values then it will take all values between those two. The precise content of the previous statement is the Intermediate Value Theorem. [33]

[33] The site mathisfun mentions some interesting applications

Intermediate Value Theorem:
Suppose $f(x)$ is a continuous function on the closed interval $[a, b]$. Then $f(x)$ must attain every value between the values of the function at the endpoints. That is, if M is any number between $f(a)$ and $f(b)$ then there is at least one number c in $[a, b]$ such that $f(c) = M$.

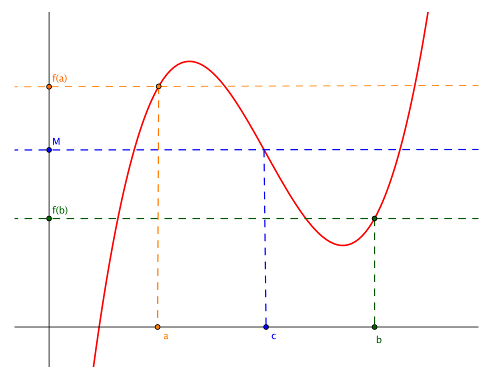


Figure 87: Intermediate Value Theorem

There are two important things to observe about the Intermediate Value Theorem:

- The theorem says that there is at least one point c such that $f(c) = M$. There could in fact be more than one point that works.
- The theorem says that there exists at least one point c, but it does not say how to *find* it. In this sense, the theorem does not provide an algorithm on how to find the point, but only guarantees its existence.

The fact that the intermediate value theorem does not say how to find the point c limits somewhat its power, however, it is still able to draw interesting conclusions about different problems as the following examples show.

Example 60. Prove that $f(x) = x^3 + x + 1$ has a root in $[-1, 0]$
First of all, a root of a polynomial is just a value r such that $f(r) = 0$ (r stands for root). Now, $f(-1) = -1$ and $f(0) = 1$ and because $f(x)$ is a continuous function the Intermediate Value Theorem says that for $M = 0$ there exists a value c such that $f(c) = M = 0$.
Calling $r = c$, then we have found a root for the polynomial. Observe that we have not said what is the value of the root!

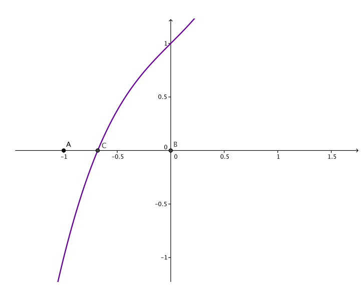


Figure 88: Root for the polynomial $f(x) = x^3 + x + 1$

Example 61. Suppose that the day starts at $t = 0$ and ends at $t = 24$ (the unit of time being hours). Suppose that at $t = 0$ the temperature is 55 degrees and it increases to a high of 85 degrees at $t = 12$ and returns to 55 degrees at $t = 24$. Show that there is at least one time in the morning (that is, between $t = 0$ and $t = 12$) when the temperature is the same as the temperature exactly 12 hours later.

The idea is to assume that the temperature is a continuous function of time so that we can apply the Intermediate Value Theorem. If we call the temperature $T(t)$ then $T(t)$ is defined on the interval $[0, 24]$ and satisfies $T(0) = T(24) = 55$, $T(12) = 85$.

Now, we are trying to find a value of time, call it t_0 such that $0 \leq t_0 \leq 12$ and $T(t_0) = T(t_0 + 12)$. As it stands, we can't apply the Intermediate Value Theorem directly because taking $a = 0$, $b = 24$ the temperatures at the endpoints are the same so we can't use it to find a value M between the endpoints.

However, the condition $T(t_0) = T(t_0 + 12)$ is the same as $T(t_0 + 12) - T(t_0) = 0$ and so this is the same as saying that the difference function $f(t) = T(t + 12) - T(t)$ must be 0 at some time. Now, $f(t)$ measures the difference between two times separated by 12 hours so $f(t)$ is defined only on the interval $[0, 12]$.

We have that $f(0) = T(12) - T(0) = 85 - 55 = 30$ and $f(12) = T(24) - T(0) = 55 - 85 = -30$. Because $M = 0$ lies between -30 and 30, the Intermediate Value Theorem says that there exists a value t_0 such that $f(t_0) = 0$, that is, $T(t_0 + 12) = T(t_0)$ which is what we wanted to show. Observe that we have not said at what time does this happen nor what is the value of the temperature at this time.

With the help of the intermediate value theorem we can prove the existence of square roots and the fact a polynomial of odd degree always has at least one root.

- **Existence of Square Roots:** Suppose that $c > 0$ is a positive number. Then there exists a positive number x such that $x^2 = c$, that is, the number c has a square root.
- **Existence of Roots for Odd Polynomials:** Suppose that $p(x) = a_{2n+1}x^{2n+1} + a_{2n}x^{2n} + a_{2n-1}x^{2n-1} + \cdots + a_1 x + a_0$ is a polynomial whose highest power is odd. Then $p(x)$ has at least a root, that is, the graph of $p(x)$ intersects the x axis at least once.

Proof. a) To show the existence of square roots consider the polynomial $q(x) = x^2 - c$ on the interval $[0, \infty)$. The function is clearly continuous on this interval and we have

$$\begin{aligned} q(0) &= -c && \text{negative} \\ q\left(c + \tfrac{1}{2}\right) &= \left(c + \tfrac{1}{2}\right)^2 - c = c^2 + \tfrac{1}{4} && \text{positive} \end{aligned} \tag{302}$$

Therefore, $q(x)$ starts being negative and achieves positive values. By the intermediate value theorem, $q(x)$ had to be equal to 0 at some point. Call this point r. Then $q(r) = 0$ or $r^2 = c$. Following tradition we write $r = \sqrt{c}$.

b) To show that $p(x)$ intersects the x axis observe that

$$\begin{aligned} \lim_{x\longrightarrow\infty} p(x) = \lim_{x\longrightarrow\infty} x^{2n+1}\left(a_{2n+1} + \frac{a_{2n}}{x} + \cdots + \frac{a_0}{x^{2n+1}}\right) = (\text{sign}(a_{2n+1}))\,\infty \\ \lim_{x\longrightarrow-\infty} p(x) = \lim_{x\longrightarrow\infty} x^{2n+1}\left(a_{2n+1} + \frac{a_{2n}}{x} + \cdots + \frac{a_0}{x^{2n+1}}\right) = (\text{sign}(a_{2n+1}))\,(-\infty) \end{aligned} \tag{303}$$

Since $p(x)$ is a continuous function on $\mathbb{R}$ the previous limits show that $p(x)$ achieves both positive and negative values so by the Intermediate Value Theorem it must be 0 at some point, that is, $p(x)$ has a root. □

With the Intermediate Value Theorem we can prove a special version for periodic functions, which has an important interpretation as an Intermediate Value Theorem for the Circle.

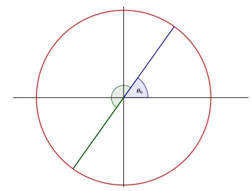


Figure 89: Intermediate Value Theorem for the Circle

- **Intermediate Value Theorem for Periodic Functions:** suppose that $f(\theta)$ is a periodic function continuous function of period 2π, that is, for all $\theta \in \mathbb{R}$, we have

$$f(\theta + 2\pi) = f(\theta) \tag{304}$$

then there is a value $\theta_0 \in [0, \pi]$ such that

$$f(\theta_0 + \pi) = f(\theta_0) \tag{305}$$

- **Intermediate Value Theorem on the Circle**: suppose that $f(\theta)$ is a continuous function defined on the circle. Then there is a pair of antipodal pints for which $f(\theta)$ takes the same value.

Proof. 1. Define the continuous function $g(\theta)$ on $[0, \pi]$ as

$$g(\theta) \equiv f(\theta + \pi) - f(\theta) \tag{306}$$

Observe that

$$\begin{aligned} g(0) = f(\pi) - f(0) \\ g(\pi) = f(2\pi) - f(\pi) = f(0) - f(\pi) = -(f(\pi) - f(0)) \end{aligned} \tag{307}$$

Therefore, $g(\theta)$ changes sign on the interval $[0, \pi]$ and so by the Intermediate Value Theorem we conclude that there is a point θ_0 such that $g(\theta_0) = 0$ or $f(\theta_0 + \pi) - f(\theta_0) = 0$ which is the statement of the intermediate value theorem for periodic functions.

2. A function $f(\theta)$ defined on the circle can be considered as a function defined on $[0, 2\pi]$ with the property that $f(0) = f(2\pi)$. Therefore, we can define $f(\theta)$ on $\mathbb{R}$ in such a way that it is periodic of period 2π. Applying the previous theorem we can find θ_0 such that $f(\theta_0 + \pi) = f(\theta_0)$. Since θ_0 and $\theta_0 + \pi$ represent antipodal points on the circle then the result follows. □

A physical application of the last theorem is as follows: consider the equator of the Earth (which can be approximated by a circle) and the temperature $T(\theta)$ measuring the temperature at points of the equator. Then there is a pair of antipodal points that have the temperature at every instant of time and regardless of how it varies! [34] A similar result is true if we interpret $T(\theta)$ as pressure, wind speed or any other (continuous) physical quantity that can be defined on a circle.

[34] As long as it can be considered a continuous function

The Intermediate Value Theorem can also be used to justify the method for solving inequalities we gave at the beginning. Since all the cases are analogous we will just justify how to solve the inequality $0 \leq \frac{p(x)}{q(x)}$ when $p(x)$ and $q(x)$ are polynomials. First of all, we need to find the domain of the function $\frac{p(x)}{q(x)}$. Clearly it will be $\mathbb{R} - \{\text{roots of } q(x)\}$, because the function can only become undefined when $q(x) = 0$.

Now, the function $\frac{p(x)}{q(x)}$ is continuous in its domain and can only be zero when $p(x) = 0$, that is, when x is a root of $p(x)$. We can think of the domain of $\frac{p(x)}{q(x)}$ as a bunch of disjoint intervals which begin and end at the different zeroes of $q(x)$. Call one of such intervals (a, b). By the Intermediate Value Theorem applied to the interval (a, b) we know that $\frac{p(x)}{q(x)}$ can only change sign (from positive to negative or from negative to positive) by crossing the x axis, that is, by going through a root of $p(x)$. Therefore, the sign of $\frac{p(x)}{q(x)}$ remains the same on each sub-interval (if any) determined by the roots of $p(x)$ so to find when does $0 \leq \frac{p(x)}{q(x)}$ we just need to evaluate the sign of the function and each sub-interval by evaluating at an arbitrary point within that sub-interval and selecting those which give a positive value for $\frac{p(x)}{q(x)}$.

Part IV

Differentiation of Functions

Geometric Interpretation of the Derivative

From elementary geometry we know that a tangent line to a circle is any line that touches only one point of the circle and is perpendicular to the radius from the center of the circle to that point.

Now, a circle is a particular kind of curve and it would be desirable to extend the concept of a tangent line to other curves, like the one shown in the next figure. From a geometrical point of view for a general curve we can't expect that the tangent line will touch the curve only at one point, however, if we "zoom in" we expect that near the tangency point, the line touches the curve only at one point (except for some curves like straight lines which are already the same as their tangent line). In any case, the important property of the tangent line is that it gives us a sense of direction of the curve, that is, if we move along the tangent line we will have a good idea of where the curve is located in space (this will be made more precise later). In any case, even for the circle we see that the tangent line can't be followed for a long time because the discrepancy between the circle and the line increases so we should follow the tangent line only for points close to the point of contact. The fact that we need to look close to the point of contact between the line and the circle suggests that there might be a limit involved.

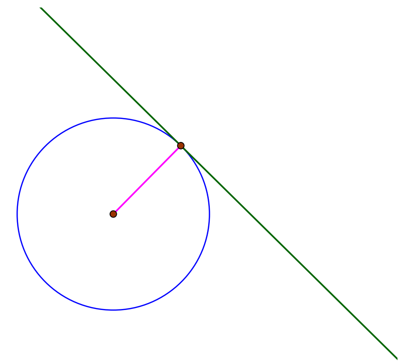


Figure 90: Tangent to a Circle

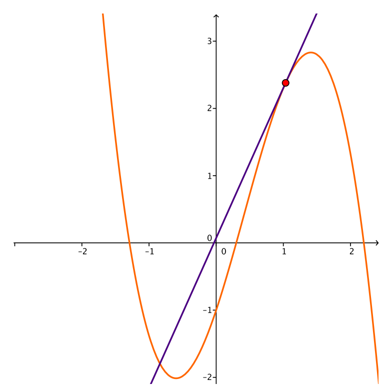


Figure 91: Tangent to a Curve

Before giving the formula on how to find the tangent line for a general curve, we will find the tangent line to a circle with the method that can be generalized to curves. For simplicity, suppose that our circle is $x^2 + y^2 = 1$ and that we want to find the tangent line to the circle at the point $(0, 1)$. From what we already know, we expect the equation of the tangent line to be $y = 1$

The equation of a line is $y = mx + b$. Since we want the line to go through $(0, 1)$ we need $b = 1$ so $y = mx + 1$. Therefore the only thing we need to find is the slope m and we can use the formula

$$m = \frac{\triangle y}{\triangle x} \tag{308}$$

where $\triangle x, \triangle y$ represent the differences between the coordinates of two points on the tangent line. Now, the problem is that we only know one point on the tangent line, which is $(0, 1)$. Therefore, instead of using the tangent line directly, we assume that the tangent line can be found as a "limit" of secant lines.

A secant line is a line between two points on a curve. In particular, we can consider the secant line that passes through $(0,1)$ and another point (x,y) on the circle, the idea being that as (x,y) approaches $(0,1)$, the secant line will approach the tangent line.

For this secant line going through $(0,1)$ and $P=(x,y)$ the slope is

$$m_P = \frac{\triangle y}{\triangle x} = \frac{y-1}{x-0} = \frac{y-1}{x} \tag{309}$$

Now, since $x^2+y^2=1$ and we want (x,y) to approach $(0,1)$ we can take y to be on the upper hemisphere so $y=\sqrt{1-x^2}$ .Substituting in 309 we have

$$m(x) = \frac{\sqrt{1-x^2}-1}{x} \tag{310}$$

where we have changed our notation of the slope to emphasize that it only depends on the coordinate x of the point P. Therefore, our geometric intuition tells us that

$$m = \lim_{x\longrightarrow 0} \frac{\sqrt{1-x^2}-1}{x} \tag{311}$$

so we only need to find the previous limit to find the slope of the tangent line! To find the limit we rationalize the expression

$$\frac{\sqrt{1-x^2}-1}{x} = \left(\frac{\sqrt{1-x^2}-1}{x}\right)\left(\frac{\sqrt{1-x^2}+1}{\sqrt{1-x^2}+1}\right) = \frac{(1-x^2)-1^2}{x\left(\sqrt{1-x^2}+1\right)} = \frac{-x}{\sqrt{1-x^2}+1} \tag{312}$$

Therefore

$$m = \lim_{x\longrightarrow 0} \frac{\sqrt{1-x^2}-1}{x} = \lim_{x\longrightarrow 0}\left(\frac{-x}{\sqrt{1-x^2}+1}\right) = \frac{0}{\sqrt{1}+1} = 0 \tag{313}$$

so the slope of the tangent line should be 0 which is what we already knew! Using this example as our model for finding tangent lines, we have found the following method:

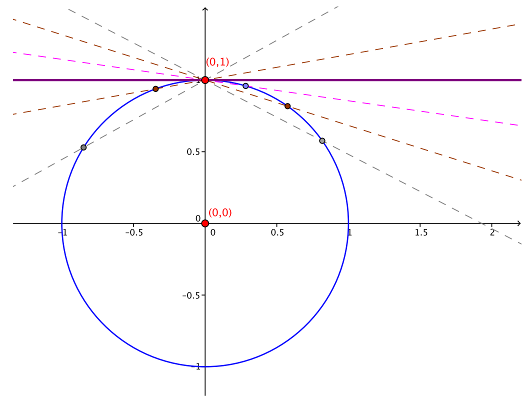


Figure 92: Tangent to a circle as a limit of secant lines

Suppose that $y = f(x)$ is a function and we want to find the tangent line to the curve that the graph of the function represents. The slope of the tangent line is called the **derivative** and to find it we calculate the limit of the slopes of the secant lines approaching the tangent line.

To be more mathematically precise, suppose that $y = f(x)$ is a function and that we want to find the tangent line to the point $(x, f(x))$ which lies on the graph of the function $y = f(x)$ (see next figure). [35] To find the equation of the tangent line the only remaining information needed is the slope of the tangent line. We move a distance h from x and consider the point $P_h = (x+h, f(x+h))$; h can be positive or negative to account for the fact that we can move to the right or the left of x. The secant line through $(x, f(x))$ and $(x+h, f(x+h))$ has slope

[35] And this link

$$m_{P_h} = \frac{\triangle y}{\triangle x} = \frac{f(x+h)-f(x)}{(x+h)-x} = \frac{f(x+h)-f(x)}{h} \tag{314}$$

Therefore, we want the slope of the tangent line to be

$$m = \lim_{h \longrightarrow 0} m_{P_h} = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h} \tag{315}$$

At this point it is important to remember that since m is defined by a limit it may or may not exist (some examples in which the limit does not exist will be discussed later).

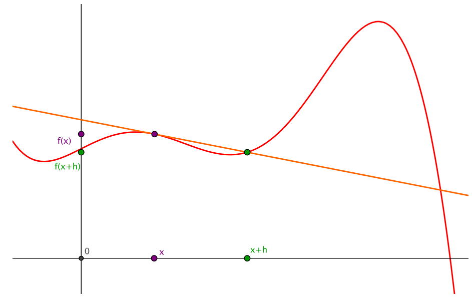


Figure 93: Tangent Line as a Limit of Secant Lines

If $y = f(x)$ is a function then the slope of the tangent line to the point $(x, f(x))$ is called the **derivative** at x and it is denoted $f'(x)$ or $y'(x)$. It can be calculated as

$$y'(x) = f'(x) \equiv \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h} \tag{316}$$

If the previous limit exists then $f(x)$ is called **differentiable at the point** x. We say that $f(x)$ is **differentiable** if it is differentiable at every point of its domain.

Example 62. Find the derivative and the equation of the tangent line for the function $y = x^2 - 2$ at point $x = 1$.

We use the definition 316 for $f(x) = x^2 - 2$.

$$y'(1) = f'(1) = \lim_{h \longrightarrow 0} \frac{f(1+h) - f(1)}{h} = \lim_{h \longrightarrow 0} \frac{\left((1+h)^2 - 2\right) - \left(1^2 - 2\right)}{h} \tag{317}$$

We use the binomial formula to simply the numerator

$$\left((1+h)^2 - 2\right) - (-1) = 1 + 2h + h^2 - 2 + 1 = 2h + h^2 = h(2+h) \tag{318}$$

Therefore 317 becomes

$$y'(1) = \lim_{h \longrightarrow 0} \frac{h(2+h)}{h} = \lim_{h \longrightarrow 0} (2+h) = 2 \tag{319}$$

Now, 2 is the slope of the tangent line so the equation of the tangent line has to be of the form

$$y = 2x + b \tag{320}$$

to find b we use the fact that the line must go through the point $(1, f(1)) = (1, -1)$ so

$$-1 = 2 + b \tag{321}$$

which gives $b = -3$ and the equation of the tangent line is

$$y = 2x - 3 \tag{322}$$

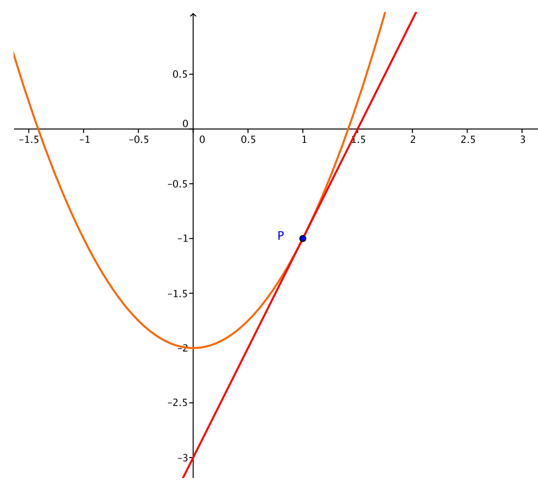


Figure 94: Tangent line to $y = x^2 - 2$ going through $(1, -1)$

Example 63. Find the derivative for the function $y = x^3$ at any point $x \in \mathbb{R}$

If we call $f(x) = x^3$ then by formula 316 we need to consider

$$y'(x) = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \longrightarrow 0} \frac{(x+h)^3 - x^3}{h} \tag{323}$$

The formula for $(x+h)^3$ is $x^3 + 3x^2h + 3xh^2 + h^3$ so

$$(x+h)^3 - x^3 = x^3 + 3x^2h + 3xh^2 + h^3 - x^3 = h\left[3x^2 + 3xh + h^2\right] \tag{324}$$

Therefore 323 becomes

$$y'(x) = \lim_{h \longrightarrow 0} \frac{h\left(3x^2 + 3xh + h^2\right)}{h} = \lim_{h \longrightarrow 0} \left(3x^2 + 3xh + h^2\right) = 3x^2 \tag{325}$$

Therefore, at point x the slope of the tangent line is $3x^2$. As x varies, the slope varies so in fact we can consider the derivative as a new function. The notation used for this new function is simply $y'(x)$ or $f'(x)$ (we will see other notations later) and the idea is that we can obtain a lot of information of $f(x)$ by studying $f'(x)$ as a function in its own right.

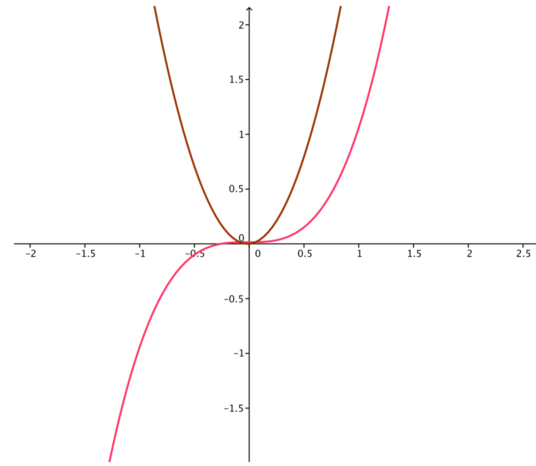


Figure 95: Graph of $f(x) = x^3$ and $f'(x) = 3x^2$

We have seen that from a geometric point of view the derivative is the slope of the tangent line to the curve that represents the graph of a function. If this were its only use probably it would be of interest only to mathematicians, fortunately, it has found many other applications, specially in physics.

During the 16th and 17th centuries, physics underwent a revolution after Aristotle's theory of motion was challenged. In Aristotle's theory the heavens and the earth followed different rules, since the former was made of ether and the latter of the remaining four elements (earth, water, wind and fire). Therefore, there was no reason to believe that the same laws should apply to both heaven and earth, in fact, their description was quite different. In particular, most attempts to understand motion were focused at the motion of the planets and the sun, which were believed to revolve around the Earth.

In this context, when Kepler announced his three laws on planetary motion in which the planets revolved around the sun and not the Earth a new theory of motion was needed to explain why the planets followed ellipses around the Sun and why the motion of the Earth around the Sun is not perceived on a daily basis. Also, Galileo proposed that the laws of nature should be written in mathematical language, which he exemplified by describing the motion of projectiles as following parabolas (in contradiction to Aristotle's account). Therefore, it became necessary to understand concepts such as position, velocity and acceleration.

In particular, one crucial question became, what is velocity? When an object is moving with constant velocity during a period of time then we may use the formula

$$v = \frac{\text{distance travelled}}{\text{time ellapsed}} \tag{326}$$

Now, in most situations the velocity of an object will not be constant. Since the velocity can change from time to time what is important is how to find the *instantaneous* velocity. If we try to use formula 326 for the instantaneous velocity we run into a problem: in any instant no distance is travelled and no time is elapsed so 326 becomes

$$v = \frac{0}{0} \tag{327}$$

which is undefined! However, while studying the concept of limit we have encountered many examples in which we have expressions of the form $\frac{0}{0}$ which become well defined after doing some algebraic manipulation and then taking the limit, which suggests that we may need some limit definition to define the instantaneous velocity correctly.

For concreteness suppose that $x(t)$ represents the position of a particle (like a ball) moving in one fixed direction. If we want to find the velocity $v(t)$ at time t, instead of applying directly 326 we may imagine that we choose some interval of time $\triangle t$ and start measuring the position of the ball at times $t, t + \triangle t$. Now, if $\triangle t$ is a small interval of time, we can use 326 to find the average velocity of the ball on the interval $[t, t + \triangle t]$

$$v_{\text{average}} = \frac{x(t+\triangle t) - x(t)}{(t+\triangle t) - (t)} = \frac{x(t+\triangle t) - x(t)}{\triangle t} \tag{328}$$

Now, if keep taking smaller intervals of time $\triangle t$ then we expect 328 to become a better approximation to the "instantaneous velocity" so we *define* the instantaneous velocity as

$$v(t) = \lim_{\triangle t \longrightarrow 0} v_{\text{average}} = \lim_{\triangle t \longrightarrow 0} \frac{x(t + \triangle t) - x(t)}{\triangle t} \tag{329}$$

Comparing with formula 316 we can see that this is exactly the same if we take $h = \triangle t$ so we have found the following.

If $x(t)$ represents the position of a particle moving in time in one direction then it's instantaneous velocity is the derivative of $x(t)$,

$$v(t) = x'(t) \equiv \lim_{h \longrightarrow 0} \frac{x(t + h) - x(t)}{h} = \lim_{\triangle t \longrightarrow 0} \frac{x(t + \triangle t) - x(t)}{\triangle t} \tag{330}$$

From a geometrical point of view, if we plot the graph of the function $x(t)$ then the velocity is the slope of the tangent line to the graph of $x(t)$.

Example 64. Suppose that the position of a car is given by $x(t) = t^2 - 3t + 1$ where x is measured in miles and t in minutes. At what time is the velocity of the car 3 miles per minute?

We use formula 330 to find $v(t)$.

$$v(t) = \lim_{h \longrightarrow 0} \frac{x(t+h) - x(t)}{h} = \lim_{h \longrightarrow 0} \frac{(t+h)^2 - 3(t+h) + 1 - \left(t^2 - 3t + 1\right)}{h} \tag{331}$$

We simplify the numerator using the binomial formula

$$\begin{aligned} & (t+h)^2 - 3(t+h) + 1 - \left(t^2 - 3t + 1\right) \\ = \;& t^2 + 2th + h^2 - 3t - 3h + 1 - t^2 + 3t - 1 \\ = \;& h(2t + h - 3) \end{aligned} \tag{332}$$

Therefore the velocity is

$$v(t) = \lim_{h \longrightarrow 0} \frac{h(2t + h - 3)}{h} = \lim_{h \longrightarrow 0} (2t + h - 3) = 2t - 3 \tag{333}$$

We want to find the value of t such that $v(t) = 3$ which gives the equation $2t - 3 = 3$ or $2t = 6$ so at $t = 3$ minutes the velocity is 3 miles per minute.

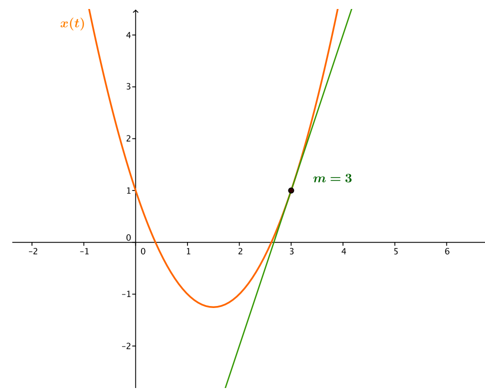


Figure 96: Graph of $x(t)$ and the tangent line at $t = 3$

Another physical problem related to the origins of calculus was the following: during the Renaissance the lenses proved of great importance for navigation and land explorations. Therefore, it became very important to understand the interaction between lenses and light, in particular how is light reflected when it strikes an object like a lens. By the laws of optics (law of reflection) if we assume that light travels like a ray (that is in straight lines) then the interaction between the lens and the light will be determined in part by the angle that the light makes when it strikes the lens. The question is then how to measure this angle. To specify an angle we need to specify two rays, one ray can certainly be the ray that represents the light, the other ray turned out to be the tangent line to the curve that represents the lens (see next figure). Since we need the derivative to find the tangent line we can see how calculus started to become important for the Scientific Revolution.

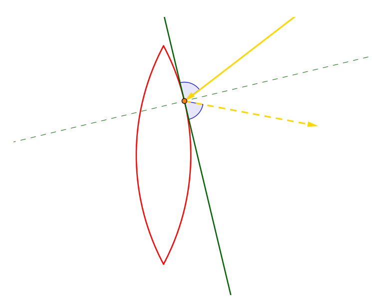


Figure 97: Reflection Law

Because the derivative has found many applications, the notation can change depending on the context on which is used. If $y = f(x)$ is a function, we have seen that the derivative can be written as $y'(x)$ or $f'(x)$. Now, going back to the definition

$$f'(x) = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h} \tag{334}$$

we can write it in a more suggestive notation, which comes from the velocity example. First of all, since $y = f(x)$, instead of writing $f(x)$ or $f(x+h)$ we can write $y(x)$ and $y(x+h)$. In this way the numerator becomes $y(x+h) - y(x)$ which we can write as $\triangle y$. Similarly, h can be written as $(x+h) - (x)$ which we can write as $\triangle x$ so we end up with

$$y'(x) = \lim_{h \longrightarrow 0} \frac{\triangle y}{\triangle x} \tag{335}$$

which captures the idea that the slope of the tangent line is the limit of the slopes of the secant lines. The Leibniz notation emphasizes this idea that the derivative is a "limit of quotients" so instead of writing $y'(x)$ we write $\frac{dy}{dx}$,

$$\frac{dy}{dx} = y'(x) = f'(x) \tag{336}$$

Where dy and dx are considered as individual entities[36], which represent a *very* small change, usually called *infinitesimals*. We will see that in many ways the derivative behaves as a ratio so Leibniz's notation handles very well those circumstances, however, it should not be taken completely literal since the derivative is *not* defined as the ratio of two infinitesimal quantities.

[36] In particular, dy and dx represent each a single symbol so we can't do something like canceling the d's, for example, $\frac{dy}{dx} \neq \frac{y}{x}$

One place in which the Leibniz notation is useful is to identify the *units* of the derivative. In most applications the quantities of interest have units: for example, position is measured in miles, meters, cost is measured in dollars, time in seconds, hours and so on. Therefore, if our function $y = f(x)$ has some units it is useful to know in advance the units of $\frac{dy}{dx} = y'(x) = f'(x)$. From the Leibniz notation we should expect that the units of $\frac{dy}{dx}$ should be the units of y divided by the units of x, this is correct and can be checked using the definition of the derivative as a limit of a quotient of those quantities.

If $y = f(x)$ is a function then the units of the derivative are

$$\left[\text{units of derivative } y'(x) = \frac{dy}{dx}\right] = \frac{[\text{units of } y]}{[\text{units of } x]} \tag{337}$$

For example, 337 says that the units of $v(t) = x'(t) = \frac{dx}{dt}$ are the units of position divided by time which we know are the units for velocity.

Linear Approximation of Functions and the Derivative as a Rescaling Factor

We mentioned previously that the tangent line gives some sense of the direction in which the curve is "moving", at least if we just look at nearby points to the point of contact. Another way to say this is that if we look carefully at the tangent lines and the graph of a function, after we "zoom in" it is difficult to distinguish between the graph of the function and the tangent line.

To be more precise, suppose $(x, f(x))$ is a point on the graph of $y = f(x)$ and we move a distance $\triangle x$ from x, so we consider the point $(x + \triangle x, f(x + \triangle x))$ on the graph of the function.

We know that the slope of the secant line going through $(x, f(x))$ and $(x + \triangle x, f(x + \triangle x))$ is

$$\frac{f(x + \triangle x) - f(x)}{x + \triangle x - x} = \frac{f(x + \triangle x) - f(x)}{\triangle x} \tag{338}$$

If we call $\triangle y = f(x + \triangle x) - f(x)$ then as long as $\triangle x$ is small, the slope of the secant line should be approximately equal to the slope of the tangent line, which is $y'(x)$, so we should have

$$\frac{\triangle y}{\triangle x} \simeq y'(x) \tag{339}$$

where the symbol $\simeq$ is used instead of an equality to denote the fact that the relationship between the two quantities is approximate. Another way of writing 339 is $\triangle y \simeq y'(x)\triangle x$ and if we switch back to the previous notation we obtain

$$f(x + \triangle x) \simeq f'(x)\triangle x + f(x) \tag{340}$$

The point of 340 is that it tells us how to estimate the value of the function at any point if we know the value of the derivative $f'(x)$ and the function $f(x)$ at one point x, and the distance $\triangle x$ from any other point to this specific point x.

Approximation of Functions:
Suppose $y = f(x)$ is a function and we know the value of $f'(x)$ and $f(x)$ at a specific value x. Then, as long as $\triangle x$ is "small", the value of $f(x + \triangle x)$ is approximately equal to

$$f(x + \triangle x) \simeq f'(x) \triangle x + f(x) \tag{341}$$

A more "numerical" way to look at this relationship is as it is indicated in the following two figures. In the first one we consider the function $f(x) = 3 + 2x$. Instead of plotting the function in the usual way, we represent the input as a horizontal line and the output as another horizontal line and the arrows connect the value of an input x with its output $f(x)$. Near an input x, we consider a small region of width $\triangle x$ around it and we follow its behavior under f. We can see that the interval $\triangle x$ expands into a new interval $\triangle y$ by some factor. In fact, we would like to determine what is this factor. Since

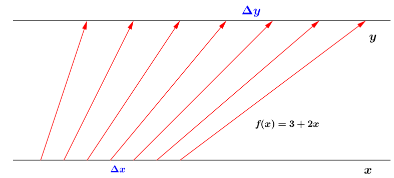


Figure 98: Derivative as an "expansion" factor

$$\triangle y = f(x + \triangle x) - f(x) = 3 + 2(x + \triangle x) - (3 + 2x) = 2\triangle x \tag{342}$$

we can see that $\triangle y$ is twice as big as $\triangle x$. Given that $f'(x) = 2$ we observe in this case that the derivative is precisely the expansion factor needed to transform $\triangle x$ into $\triangle y$. This is basically what we would expect from 341, but now we are not interpreting the equation as the approximation of the function by a straight line, rather, we are considering the derivative as an "instantaneous" expansion factor.

The second example is more complicated because the function $f(x) = x^2$ is quadratic instead of linear. In this case

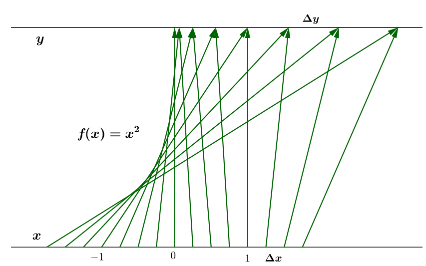


Figure 99: Derivative as an "expansion" factor

$$\triangle y = f(x + \triangle x) - f(x) = (x + \triangle x)^2 - x^2 = 2x\triangle x + (\triangle x)^2 \tag{343}$$

If we ignore the quadratic term then we obtain $\triangle y \simeq 2x\triangle x$, which is again basically what 341 says. However, in this case we can see that the derivative has the following effect:

Interpretation of the Derivative as a Rescaling Factor: Suppose that we represent $f(x)$ by drawing two lines: one with the input values and the other with the output values. Then the derivative can be interpreted as an "instantaneous" rescaling factor: let a be a point of the domain of $f(x)$ and consider $f'(a)$.

- If $|f'(a)| > 1$ then near a the points move away from each other, that is, the distance between them expand
- If $0 \leq |f'(a)| < 1$ then near a the points move near each other, that is, the distance between them contract. The extreme case is when $f'(a) = 0$ since the points try to contract into a point.
- If $|f'(a)| = 1$ then near a the points tend to preserve their relative distances, at least if we ignore higher order terms (h.o.t) of the form $(\triangle x)^2$, $(\triangle x)^3$, etc.
- If $f'(a) < 0$ then near a there is a reversal of the relative ordering of points near a
- If $f'(a) > 0$ then near a the relative ordering of points near a is preserved

Properties of the Derivative

Now that we have a better idea of what the derivative is, we can try to find some of its properties. For example, suppose that we have two functions $f(x)$ and $g(x)$ and we know their derivatives $f'(x)$, $g'(x)$. The sum $h(x) = f(x) + g(x)$ is a new function (for example, if $f(x) = x^2$ and $g(x) = 3 + x$ then $h(x) = x^2 + x + 3$) and we would like to find a relationship between $h'(x)$ and $f'(x)$, $g'(x)$. To find $h'(x)$ we calculate the following limit:

$$\begin{aligned} h'(x) &= \lim_{\triangle x \longrightarrow 0} \tfrac{h(x+\triangle x)-h(x)}{\triangle x} \\ &= \lim_{\triangle x \longrightarrow 0} \tfrac{f(x+\triangle x)-f(x)+g(x+\triangle x)-g(x)}{\triangle x} \\ &= \lim_{\triangle x \longrightarrow 0} \tfrac{f(x+\triangle x)-f(x)}{\triangle x} + \lim_{\triangle x \longrightarrow 0} \tfrac{g(x+\triangle x)-g(x)}{\triangle x} \\ &= f'(x) + g'(x) \end{aligned} \tag{344}$$

Therefore we have found that

$$(f(x) + g(x))' = f'(x) + g'(x) \tag{345}$$

A more interesting example is how to find the derivative of the product $k(x) = f(x)g(x)$. This time we will use the linear approximation 341

$$\begin{cases} f\,(x + \triangle x) \simeq f'(x)\triangle x + f(x) \\ g\,(x + \triangle x) \simeq g'(x)\triangle x + g(x) \\ k\,(x + \triangle x) \simeq k'(x)\triangle x + k(x) \end{cases} \tag{346}$$

and the third equation can be written as

$$f\,(x + \triangle x)\, g\,(x + \triangle x) \simeq (f(x)g(x))'\,\triangle x + f(x)g(x) \tag{347}$$

Now we use the first two equations of 346 to obtain

$$(f'(x)\triangle x + f(x))\,(g'(x)\triangle x + g(x)) \simeq (f(x)g(x))'\,\triangle x + f(x)g(x) \tag{348}$$

If we expand the product in the left hand side of 348 we have

$$f'(x)g'(x)\,(\triangle x)^2 + f'(x)g(x)\triangle x + f(x)g'(x)\triangle x + f(x)g(x) \tag{349}$$

comparing with 348 we can cancel the term $f(x)g(x)$ so we end up with

$$f'(x)g'(x)\left(\triangle x\right)^2 + f'(x)g(x)\triangle x + f(x)g'(x)\triangle x \simeq \left(f(x)g(x)\right)' \triangle x \tag{350}$$

We are almost done with the calculation, the only thing remaining is to drop off the term $f'(x)g'(x)\left(\triangle x\right)^2$. The idea behind this is that the approximation 341 works well only if $\triangle x$ is very small, and when $\triangle x$ is small $\left(\triangle x\right)^2$ is even smaller. For example, if $\triangle x = 0.01$ then $\left(\triangle x\right)^2 \simeq 0.0001$ so this term contributes almost nothing to our approximation. In this way we end up with $f'(x)g(x)\triangle x + f(x)g'(x)\triangle x \simeq \left(f(x)g(x)\right)' \triangle x$ which suggests that

$$\left(f(x)g(x)\right)' = f'(x)g(x) + f(x)g'(x) \tag{351}$$

This can be proved with the limit definition but if one tries to do it one will need to use a continuity argument. To be more precise, when we calculate

$$\begin{aligned}
k'(x) &= \lim_{\triangle x\longrightarrow 0} \frac{k(x+\triangle x)-k(x)}{\triangle x} \\
&= \lim_{\triangle x\longrightarrow 0} \frac{f(x+\triangle x)g(x+\triangle x)-f(x)g(x)}{\triangle x} \\
&= \lim_{\triangle x\longrightarrow 0} \frac{f(x+\triangle x)g(x+\triangle x)-f(x+\triangle x)g(x)+f(x+\triangle x)g(x)-f(x)g(x)}{\triangle x} \\
&= \lim_{\triangle x\longrightarrow 0} \frac{f(x+\triangle x)(g(x+\triangle x)-g(x))}{\triangle x} + \lim_{\triangle x\longrightarrow 0} \frac{(f(x+\triangle x)-f(x))g(x)}{\triangle x} \\
&= \lim_{\triangle x\longrightarrow 0} f\left(x+\triangle x\right) \lim_{\triangle x\longrightarrow 0} \frac{g(x+\triangle x)-g(x)}{\triangle x} + \lim_{\triangle x\longrightarrow 0} g\left(x\right) \lim_{\triangle x\longrightarrow 0} \frac{f(x+\triangle x)-f(x)}{\triangle x} \\
&= f(x)g'(x) + g(x)f'(x)
\end{aligned} \tag{352}$$

we have to assume that $f(x)$ is continuous in order to change $\lim_{\triangle x\longrightarrow 0} f\left(x+\triangle x\right)$ by $f(x)$. Fortunately this is always the case: functions with derivatives must be continuous. This also says that if a function fails to be continuous at some point of its domain then it can't have a derivative at that point.

Differentiable Functions are Continuous:
Let $f(x)$ be a function and $a \in \mathbb{R}$ be an element of the domain of $f(x)$. Suppose that $f'(a)$ exists, that is, the function is differentiable at the point a. Then $f(a)$ is continuous at the point a, that is

$$f(x) \text{ differentiable at } a \longrightarrow f(x) \text{ continuous at } a \tag{353}$$

The contrapositive of this statement is

$$f(x) \text{ discontinuous at } a \longrightarrow f(x) \text{ not differentiable at } a \tag{354}$$

Proof. To see why this is true since $f'(a)$ exists we have that the limit

$$f'(a) = \lim_{h \longrightarrow 0} \frac{f(a+h) - f(a)}{h} \tag{355}$$

exists. In particular, this implies that

$$\lim_{h \longrightarrow 0} (f(a+h) - f(a)) = \lim_{h \longrightarrow 0} \left[\frac{f(a+h)-f(a)}{h} \right] [h] =$$
$$= \left(\lim_{h \longrightarrow 0} \frac{f(a+h)-f(a)}{h} \right) (\lim_{h \longrightarrow 0} h) = (f'(a)) (\lim_{h \longrightarrow 0} h) = 0 \tag{356}$$

so $\lim_{h \longrightarrow 0} (f(a+h) - f(a)) = 0$ and by the properties of the limit this is the same as

$$\lim_{h \longrightarrow 0} f(a+h) = f(a) \tag{357}$$

If we call $x = a + h$ then we see that $\lim_{h \longrightarrow 0} f(a+h) = f(a)$ is the same as $\lim_{x \longrightarrow a} f(x) = f(a)$ and this is precisely the same as saying that $f(x)$ is continuous at a. □

Therefore, if a function is not continuous at a point we know that it won't be differentiable at that point so we have just found a way to "fabricate" lots of examples of functions which are not differentiable.

It is important to notice that the previous property does not say that every continuous function is differentiable, it says that every differentiable function is continuous (as an analogy, every living creature is a plant is clearly very distinct from every plant is a living creature).

For example, consider the function $f(x) = |x|$, the absolute value of x. Recall that

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases} \tag{358}$$

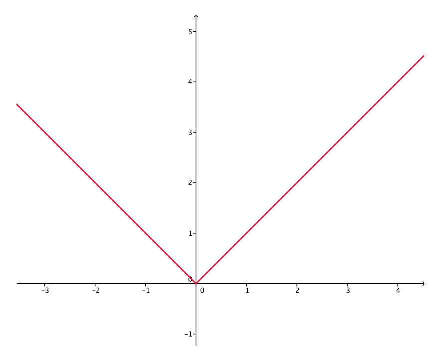


Figure 100: Graph of $y = f(x) = |x|$

Since the derivative gives the slope of the tangent line at a point, we can see that when $x = 0$ there are two conflicting candidates for the tangent line, namely, the two rays that make up the graph of $|x|$. To be more precise, if we claim that $f'(0)$ does not exist we have to say why does

$$f'(0) = \lim_{h \longrightarrow 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \longrightarrow 0} \frac{|h|}{h} \tag{359}$$

does not exist. To see that the previous limit does not exist we simply compute the one-sided limits:

$$\begin{aligned} \lim_{h \longrightarrow 0^+} \frac{|h|}{h} &= \lim_{h \longrightarrow 0^+} \frac{h}{h} = 1 \\ \lim_{h \longrightarrow 0^-} \frac{|h|}{h} &= \lim_{h \longrightarrow 0^-} \frac{-h}{h} = -1 \end{aligned} \tag{360}$$

and because the one-sided limits are different the limit does not exist.

Hence the graph of a function that is differentiable at every point has no peaks or straight-edges. For practical purposes, the most useful formulas for finding derivatives are the following

Properties of the Derivative:

- The derivative of a constant function is 0: suppose $f(x) = c$ is a constant function, for example, $y = f(x) = 2$, $y = f(x) = -3$, $y = f(x) = 0$. Then
$$\frac{dy}{dx} = \frac{dc}{dx} = (c)' = 0 \tag{361}$$

- Power Rule: The derivative of $y = f(x) = x^n$ where n is any real number is nx^{n-1}.
$$\frac{dy}{dx} = \frac{dx^n}{dx} = (x^n)' = nx^{n-1} \tag{362}$$
For example, $\left(x^3\right)' = 3x^2$, $\frac{dx^5}{dx} = 5x^4$, $(\sqrt{x})' = \left(x^{\frac{1}{2}}\right)' = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}$

- If c is a constant and $f(x)$ a differentiable function then
$$\frac{d\,(cf)}{dx} = (cf(x))' = c\,(f(x))' = c\frac{df}{dx} \tag{363}$$
for example, $\frac{d\left(3x^2\right)}{dx} = \left(3x^2\right)' = 3\left(x^2\right)' = 6x$.

- The derivative of a sum is the sum of the derivatives: if $f(x)$ and $g(x)$ are two differentiable functions then
$$\frac{d(f+g)}{dx} = (f(x)+g(x))' = f'(x) + g'(x) = \frac{df}{dx} + \frac{dg}{dx} \tag{364}$$
For example, $\frac{d(x+x^2)}{dx} = \left(x+x^2\right)' = (x)' + \left(x^2\right)' = 1 + 2x$

- The product or Leibniz rule: If $f(x)$, $g(x)$ are two differentiable functions then
$$\frac{d(fg)}{dx} = (f(x)g(x))' = f'(x)g(x) + f(x)g'(x) = \frac{df}{dx}g + f\frac{dg}{dx} \tag{365}$$
$$\textbf{derivative product} = (\textbf{ first factor})' \, (\textbf{second factor}) + (\textbf{first factor}) \, (\textbf{second factor})' \tag{366}$$

- The quotient rule: If $f(x)$, $g(x)$ are two differentiable functions and $g(x) \neq 0$ then
$$\frac{d\left(\frac{f}{g}\right)}{dx} = \left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2} = \frac{\frac{df}{dx}g - f\frac{dg}{dx}}{g^2} \tag{367}$$
$$\textbf{derivative quotient} = \frac{(\textbf{ numerator})' \, (\textbf{denominator}) - (\textbf{numerator}) \, (\textbf{ denominator})'}{\textbf{denominator squared}} \tag{368}$$

Example 65. Find the derivative of $f(r) = \frac{4}{3}\pi r^3$ and $g(u) = \frac{2}{\sqrt{u}}$.

For $f(r)$ we use the property that the constants can be taken "outside" the derivative

$$\frac{d\left(\frac{4}{3}\pi r^3\right)}{dr} = \left(\frac{4}{3}\pi r^3\right)' = \frac{4}{3}\pi\left(r^3\right)' \tag{369}$$

so we only need to find $\frac{dr^3}{dr} = \left(r^3\right)'$. This can be found using the power rule 362 with $n = 3$.

$$\frac{dr^3}{dr} = \left(r^3\right)' = 3r^{3-1} = 3r^2 \tag{370}$$

so

$$f'(r) = \left(\frac{4}{3}\pi r^3\right)' = \frac{4}{3}\pi\left(r^3\right)' = \frac{4}{3}\pi\left(3r^2\right) = 4\pi r^2 \tag{371}$$

For $g(u)$ we follow a similar procedure. First of all we take the constant outside the derivative

$$\left(\frac{2}{\sqrt{u}}\right)' = 2\left(\frac{1}{\sqrt{u}}\right)' = 2\left(u^{-\frac{1}{2}}\right)' \tag{372}$$

To find $\left(u^{-\frac{1}{2}}\right)'$ we use the power rule with $n = -\frac{1}{2}$. Then

$$\left(u^{-\frac{1}{2}}\right)' = -\frac{1}{2}u^{-\frac{1}{2}-1} = -\frac{1}{2}u^{-\frac{3}{2}} = -\frac{1}{2\sqrt{u^3}} \tag{373}$$

In this way

$$g'(u) = \left(\frac{2}{\sqrt{u}}\right)' = 2\left(u^{-\frac{1}{2}}\right)' = 2\left(-\frac{1}{2\sqrt{u^3}}\right) = -\frac{1}{\sqrt{u^3}} \tag{374}$$

Example 66. Find the derivative of $f(x) = \frac{x^3-4x^2+3}{x}$

We have two options. The first one is to apply the quotient rule 367

$$\left(\frac{x^3-4x^2+3}{x}\right)' = \frac{\left(x^3-4x^2+3\right)'(x) - \left(x^3-4x^2+3\right)(x)'}{x^2} \tag{375}$$

To find $\left(x^3-4x^2+3\right)'$ we use the fact that the derivative of the sum is the sum of the derivatives

$$\left(x^3-4x^2+3\right)' = \left(x^3\right)' - 4\left(x^2\right)' + (3)' = 3x^2 - 8x + 0 = 3x^2 - 8x \tag{376}$$

Also, $x' = 1$. In this way

$$\left(\frac{x^3-4x^2+3}{x}\right)' = \frac{\left(3x^2-8x\right)(x) - \left(x^3-4x^2+3\right)(1)}{x^2} = \frac{2x^3-4x^2-3}{x^2} \tag{377}$$

The other way to find the derivative is separating first the numerator and then using the power rules for derivatives

$$\left(\frac{x^3-4x^2+3}{x}\right)' = \left(x^2-4x+\frac{3}{x}\right)' = \left(x^2\right)' - 4(x)' + 3\left(x^{-1}\right)' = 2x - 4 - \frac{3}{x^2} = \frac{2x^3-4x^2-3}{x^2} \tag{378}$$

Example 67. Find the derivative of $f(t) = \frac{4}{t^4} - \frac{3}{t^3} + \frac{2}{t}$

We use the power rule and the rule for sums (we switch to Leibniz's notation to practice it)

$$\frac{d}{dt}\left(\frac{4}{t^4} - \frac{3}{t^3} + \frac{2}{t}\right) = 4\frac{d}{dt}t^{-4} - 3\frac{d}{dt}t^{-3} + 2\frac{d}{dt}t^{-1} = -16t^{-5} + 9t^{-4} - 2t^{-2} = -\frac{16}{t^5} + \frac{9}{t^4} - \frac{2}{t^2} \tag{379}$$

Example 68. Let $f(x) = x^3 - 4x^2$. Find the point(s) on the graph of f where the tangent line is horizontal.

A horizontal tangent line has slope equal to 0. Therefore, we need to find the values of x such that $f'(x) = 0$. Since

$$f'(x) = \left(x^3 - 4x^2\right)' = 3x^2 - 8x \tag{380}$$

we need to solve the equation

$$3x^2 - 8x = 0 \tag{381}$$

which is equivalent to

$$x\,(3x - 8) = 0 \tag{382}$$

and this gives $x = 0$ and $x = \frac{8}{3}$ as can be seen from the graph of $f(x)$.

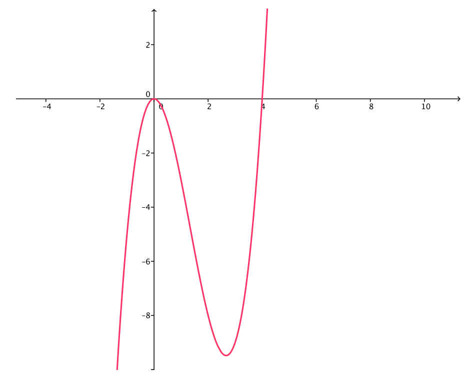


Figure 101: Graph of $f(x) = x^3 - 4x^2$

Example 69. Find the derivative of $f(t) = (t-1)(2t+1)$

We can use the product rule 365

$$[(t-1)(2t+1)]' = (t-1)'\,(2t+1) + (t-1)\,(2t+1)' = (2t+1) + (t-1)(2) = 4t - 1 \tag{383}$$

Example 70. Find the derivative of $f(w) = (w^3 - w^2 + w - 1)(w^2 + 2)$

We use the product rule 365

$$\left[\left(w^3 - w^2 + w - 1\right)\left(w^2+2\right)\right]' = \left(w^3 - w^2 + w - 1\right)'\left(w^2+2\right) + \left(w^3 - w^2 + w - 1\right)\left(w^2+2\right)'$$
$$= \left(3w^2 - 2w + 1\right)\left(w^2+2\right) + \left(w^3 - w^2 + w - 1\right)(2w) = 5w^4 - 4w^3 + 9w^2 - 6w + 2 \tag{384}$$

Example 71. Find the derivative of $f(x) = \left(5x^2+1\right)\left(2\sqrt{x} - 1\right)$

Again we apply the product rule 365

$$\left[\left(5x^2+1\right)\left(2\sqrt{x}-1\right)\right]' = \left(5x^2+1\right)'\left(2\sqrt{x}-1\right) + \left(5x^2+1\right)\left(2\sqrt{x}-1\right)'$$
$$= 10x\left(2\sqrt{x}-1\right) + \left(5x^2+1\right)\left(2\cdot\frac{1}{2\sqrt{x}}\right) = \frac{5x^2+1}{\sqrt{x}} + 10\left(2\sqrt{x}-1\right)x = 25x^{\frac{3}{2}} - 10x + \frac{1}{\sqrt{x}} \tag{385}$$

Example 72. Find the derivative of $f(x) = \frac{x+\sqrt{3x}}{3x-1}$

We use the quotient rule 367

$$\begin{aligned}\left(\frac{x+\sqrt{3x}}{3x-1}\right)' &= \frac{\left(x+\sqrt{3x}\right)'(3x-1)-\left(x+\sqrt{3x}\right)(3x-1)'}{(3x-1)^2} \\ = \frac{\left(1+\frac{\sqrt{3}}{2\sqrt{x}}\right)(3x-1)-\left(x+\sqrt{3x}\right)(3)}{(3x-1)^2} &= -\frac{3\sqrt{3}x+2\sqrt{x}+\sqrt{3}}{2\sqrt{x}(3x-1)^2}\end{aligned} \tag{386}$$

Example 73. Suppose that f, g are functions that are differentiable at $x = 1$ such that $f(1) = 2, f'(1) = -1$, $g(1) = -2$, $g'(1) = 3$. Find the value of $h_1'(1), h_2'(1)$ if $h_1(x) = \frac{xf(x)}{x+g(x)}$ and $h_2(x) = \frac{f(x)g(x)}{f(x)-g(x)}$

First we find $h_1'(x)$ using the quotient rule

$$\begin{aligned}h_1'(x) = \left(\frac{xf(x)}{x+g(x)}\right)' &= \frac{(xf(x))'(x+g(x))-(xf(x))(x+g(x))'}{(x+g(x))^2} \\ &= \frac{(f(x)+xf'(x))(x+g(x))-(xf(x))(1+g'(x))}{(x+g(x))^2}\end{aligned} \tag{387}$$

where in the last step we applied the product rule to $(xf(x))'$. If we take $x = 1$ we find that

$$h_1'(1) = \frac{(f(1)+f'(1))(1+g(1))-(f(1))(1+g'(1))}{(1+g(1))^2} = \frac{(2-1)(1-2)-(2)(1+3)}{(1-2)^2} = -9 \tag{388}$$

Observe that we must differentiate first and evaluate later, otherwise, if we had evaluated first we would have $h_1(1) = -2$ and since the derivative of a constant is zero we would have concluded that the derivative is zero which is incorrect.

Now we find $h_2'(x)$ using the quotient rule

$$\begin{aligned}h_2'(x) = \left(\frac{f(x)g(x)}{f(x)-g(x)}\right)' &= \frac{(f(x)g(x))'(f(x)-g(x))-f(x)g(x)(f(x)-g(x))'}{(f(x)-g(x))^2} \\ &= \frac{(f'(x)g(x)+f(x)g'(x))(f(x)-g(x))-f(x)g(x)(f'(x)-g'(x))}{(f(x)-g(x))^2}\end{aligned} \tag{389}$$

Taking $x = 1$ we find that

$$h_2'(1) = \frac{((-1)(-2)+2(3))(2+2)-2(-2)(-1-3)}{(2+2)^2} = 1 \tag{390}$$

It is also possible to find the derivatives of the trigonometric functions. For example, suppose that we want to find the derivative of

$f(x) = \sin x$. Then

$$\begin{aligned}
\frac{d\sin x}{dx} &= \lim_{\triangle x \longrightarrow 0} \frac{\sin(x+\triangle x)-\sin(x)}{\triangle x} \\
&= \lim_{\triangle x \longrightarrow 0} \frac{\sin(x)\cos(\triangle x)+\cos(x)\sin(\triangle x)-\sin(x)}{\triangle x} \\
&= \lim_{\triangle x \longrightarrow 0} \frac{\sin(x)(\cos(\triangle x)-1)+\cos(x)\sin(\triangle x)}{\triangle x} \\
&= \sin(x)\lim_{\triangle x \longrightarrow 0} \frac{(\cos(\triangle x)-1)}{\triangle x} + \cos(x)\lim_{\triangle x \longrightarrow 0} \frac{\sin(\triangle x)}{\triangle x} \\
&= \sin(x)\cdot 0 + \cos(x)\cdot 1 \\
&= \cos(x)
\end{aligned} \tag{391}$$

Therefore

$$\frac{d}{dx}\sin x = \cos x \tag{392}$$

Using the limit definition we can also find the derivative for cosine and using the quotient rule we can find the remaining trigonometric derivatives

Trigonometric Derivatives: Whenever they are defined, the trigonometric functions are differentiable and their derivatives are

$$\begin{aligned}
(\cos x)' &= \frac{d}{dx}\cos x = -\sin x & (\sec x)' &= \frac{d}{dx}\sec x = \sec x \tan x \\
(\sin x)' &= \frac{d}{dx}\sin x = \cos x & (\csc x)' &= \frac{d}{dx}\csc x = -\csc x \cot x \\
(\tan x)' &= \frac{d}{dx}\tan x = \sec^2 x & (\cot x)' &= \frac{d}{dx}\cot x = -\csc^2 x
\end{aligned} \tag{393}$$

We can also try to find the derivative of e^x: as we will see in a moment the relationship between e^x and the its derivative is unique among the functions. To find $(e^x)'$ we use the definition as a limit

$$(e^x)' = \lim_{h \longrightarrow 0} \frac{e^{x+h}-e^x}{h} = \lim_{h \longrightarrow 0} \frac{e^x e^h - e^x}{h} = e^x \lim_{h \longrightarrow 0} \frac{e^h - 1}{h} \tag{394}$$

Therefore, we need to find

$$\lim_{h \longrightarrow 0} \frac{e^h - 1}{h} \tag{395}$$

Although we won't do the calculation it is possible to show that

$$\lim_{h \longrightarrow 0} \frac{e^h - 1}{h} = 1 \tag{396}$$

so substituting in 394 we have found that the exponential function is its own derivative!

$$\frac{de^x}{dx} = (e^x)' = e^x \tag{397}$$

Once we have seen the implicit differentiation method it will be shown that

$$\frac{d}{dx} \ln x = (\ln x)' = \frac{1}{x} \tag{398}$$

Also, we can find the derivatives of the hyperbolic functions by writing them in terms of the exponential function and using some properties of the derivative like the quotient rule.

Derivative of Hyperbolic Functions:

$$\begin{aligned}
(\cosh t)' &= \tfrac{d}{dt} \cosh t = \sinh t & (\operatorname{sech} t)' &= \tfrac{d}{dt} \operatorname{sech} t = -\operatorname{sech} t \tanh t \\
(\sinh t)' &= \tfrac{d}{dt} \sinh t = \cosh t & (\operatorname{csch} t)' &= \tfrac{d}{dt} \operatorname{csch} t = -\operatorname{csch} t \coth t \\
(\tanh t)' &= \tfrac{d}{dx} \tanh t = (\operatorname{sech} t)^2 & (\coth t)' &= \tfrac{d}{dt} \coth t = -(\operatorname{csch} t)^2
\end{aligned} \tag{399}$$

The Chain Rule

We have seen several algebraic operations that can be done on functions, for example, we can add functions, multiply functions, take quotients of functions, etc. There is a remaining operation that can be performed on functions which is perhaps the most important of all the different operations: the composition of functions.

The composition of functions can be considered as a applying one function after another. For example, consider the following "recipe" to produce a function

1. Start with a number x:

2. Square the number: this gives x^2

3. Add 1 to the result: this gives $x^2 + 1$

4. Take the square root of the result: this gives $\sqrt{x^2+1}$

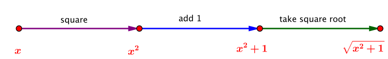


Figure 102: Composition of Functions

Therefore, at the end of this process we have a function which sends x to $\sqrt{x^2+1}$. The important thing is that this function can be considered as the process resulting from using simpler functions in a particular order, and the process itself is known as composing functions.

For example, if we define

$$\begin{gathered} f(x) = x^2 \\ g(x) = x + 1 \\ h(x) = \sqrt{x} \end{gathered} \tag{400}$$

Then we can see that

$$\sqrt{x^2+1} = h\left(x^2+1\right) = h\left(g\left(x^2\right)\right) = h\left(g\left(f(x)\right)\right) \tag{401}$$

so we can say that $\sqrt{x^2+1}$ is the composition of $f(x), g(x), h(x)$. Observe that the composition of functions is read "inside out", meaning that the first function applied is the one inside all the parentheses and the last function performed is the one outside all the parentheses, another way to say this is that the composition is read from right to left, in contrast to the usual way of reading symbols in western culture.

If $f(x)$ is a function and $g(x)$ is another function then the composition of $f(x)$ with $g(x)$ is

$$g\left(f(x)\right) \tag{402}$$

and sometimes it is written as

$$(g \circ f)(x) = g\left(f(x)\right) \tag{403}$$

The notation means: do first f and then g

Using our previous example, we have that

$$\begin{aligned} g \circ f(x) = g(f(x)) = g\left(x^2\right) = x^2 + 1 \\ f \circ g(x) = f\left(g(x)\right) = f\left(x+1\right) = \left(x+1\right)^2 = x^2 + 2x + 1 \end{aligned} \tag{404}$$

We can see from the previous calculation that $g \circ f$ is different from $f \circ g$ so the order of functions when composing them matters.

Example 74. Let $f(x)$ be a function with domain $[10, \infty)$ and $g(x) = 3x - 5$. Find the domain of $(f \circ g)(x)$.

First of all, we can call $h(x) = (f \circ g)(x) = f\left(g(x)\right) = f\left(3x - 5\right)$. Therefore, for $h(x)$ to make sense we need $f(3x-5)$ to make sense since the domain of f is $[10, \infty)$ this implies that $10 \leq 3x - 5$ or $5 \leq x$. Therefore, the domain of $h(x)$ is $[5, \infty)$

The reason why composition is important is that it facilitates the calculation of the derivatives of complicated functions. For example, consider the function $h(x) = \sqrt{2x+1}$. From our rules we have no way to calculate $h'(x)$. In fact we would be forced to apply the definition of the derivative

$$\begin{aligned} h'(x) = \lim_{h \longrightarrow 0} \frac{\sqrt{2x+2h+1} - \sqrt{2x+1}}{h} = \lim_{h \longrightarrow 0} \left(\frac{\left(\sqrt{2x+2h+1}\right)^2 - \left(\sqrt{2x+1}\right)^2}{h\left(\sqrt{2x+2h+1} + \sqrt{2x+1}\right)} \right) \\ = \lim_{h \longrightarrow 0} \frac{2x+2h+1-(2x+1)}{h\left(\sqrt{2x+2h+1}+\sqrt{2x+1}\right)} = \lim_{h \longrightarrow 0} \frac{2}{\left(\sqrt{2x+2h+1}+\sqrt{2x+1}\right)} = \frac{2}{2\sqrt{2x+1}} = \frac{1}{\sqrt{2x+1}} \end{aligned} \tag{405}$$

On the other hand, if we define $f(x) = 2x + 1$ and $g(x) = \sqrt{x}$ then we can see that

$$h(x) = \sqrt{2x+1} = g\left(f(x)\right) \tag{406}$$

so it is natural to ask if there is any relationship between $h'(x)$, $g'(x)$ and $f'(x)$. Applying the rules of the derivatives for $f(x)$ and $g(x)$ we have that

$$\begin{aligned} f'(x) = 2 \\ g'(x) = \frac{1}{2\sqrt{x}} \end{aligned} \tag{407}$$

If we multiply $f'(x)$ and $g'(x)$ we have $g'(x)f'(x) = \frac{1}{\sqrt{x}}$ which is not quite what we need. However, it is almost what we need, in fact, we have that

$$h'(x) = (g \circ f)'(x) = \frac{1}{\sqrt{2x+1}} = [g'(f(x))]\,[f'(x)] \tag{408}$$

The chain rule will say that this is the general situation, and once again the linear approximation 341 helps us see why it works. The linear approximation for $h(x)$ is

$$h'(x + \triangle x) \simeq h'(x)\triangle x + h(x) \tag{409}$$

Using the definition of $h(x)$ this is the same as

$$g\left(f\left(x + \triangle x\right)\right) \simeq \left(g\left(f(x)\right)\right)' \triangle x + g\left(f(x)\right) \tag{410}$$

Now we can approximate $f(x + \triangle x)$ with 341, $f(x + \triangle x) \simeq f'(x)\triangle x + f(x)$. Substituting in 410 we have

$$g\left(f'(x)\triangle x + f(x)\right) \simeq (g(f(x)))'\triangle x + g\left(f(x)\right) \tag{411}$$

Defining $u \equiv f(x)$ [37] we have that $f'(x) \simeq \frac{\triangle u}{\triangle x}$ or $f'(x)\triangle x \simeq \triangle u$ so we can change the left hand side of 411 as

[37] It is going to be clear soon why we use the letter u instead of y

$$g\left(f'(x)\triangle x + f(x)\right) \simeq g\left(\triangle u + u\right) \tag{412}$$

Applying the linear approximation to $g(\triangle u + u)$ we have

$$g\left(\triangle u + u\right) \simeq g'(u)\triangle u + g(u) \simeq g'(f(x))f'(x)\triangle x + g\left(f(x)\right) \tag{413}$$

Substituting in 411 we end up with

$$g'(f(x))f'(x)\triangle x + g\left(f(x)\right) \simeq (g(f(x))' \triangle x + g\left(f(x)\right) \tag{414}$$

which gives an argument as to why

$$(g(f(x)))' = g'\left(f(x)\right) f'(x) \tag{415}$$

Since the chain rule seems more complicated than the previous rule we will see how Leibniz's notation can be useful in this context. Returning to our example we are trying to find $\frac{dh}{dx}$. If think of $2x+1$ as a new variable and call it u then we have

$$h(x) = \sqrt{2x+1} = \sqrt{u} \tag{416}$$

where $u = 2x + 1$. Now, we can calculate the derivative of $\sqrt{u}$ with respect to u,

$$\frac{dh}{du} = \frac{1}{2\sqrt{u}} \tag{417}$$

and if we use a "hybrid" notation and replace u by $2x+1$ we have

$$\frac{dh}{du} = \frac{1}{2\sqrt{u}} = \frac{1}{2\sqrt{2x+1}} \tag{418}$$

on the other hand

$$\frac{du}{dx} = \frac{d\,(2x+1)}{dx} = 2 \tag{419}$$

so 408 can be written as

$$\frac{dh}{dx} = \frac{dh}{du}\frac{du}{dx} \tag{420}$$

so in Leibniz's notation the derivative of a composition follows from the fact that you can "cancel" the numerator and denominator of a product of quotients. Again, this notation is very helpful but can't be taken literally since dh, du, dx are not numbers so they can't be canceled out in that way, but they work as if it were possible to do the cancelation.

This makes more sense if we interpret the derivative as a rescaling factor. Suppose we have the function $h(x) = \frac{1}{3}\left(\frac{x^2-3}{2}\right)^3$ and we want to find $h'(x)$. At some particular value a of the input, we are interpreting $h'(a)$ as the scaling factor between a the length $\triangle x$ of the neighborhood of a $\left(a - \frac{\triangle x}{2}, a + \frac{\triangle x}{2}\right)$ and the length $\triangle y$ of the neighborhood $\left(h(a) - \frac{\triangle y}{2}, h\,(a) + \frac{\triangle y}{2}\right)$, that is, [38]

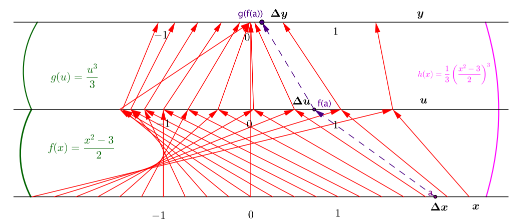


Figure 103: Visual Interpretation of the Chain Rule

[38] It is not necessarily true that if we start with a symmetric neighborhood with respect to a we should get a symmetric neighborhood with respect to $h(a)$ but since this is a heuristic argument we shouldn't be too worried about it

$$\triangle y \simeq h'(a)\triangle x \tag{421}$$

The idea of the chain rule is to interpret $h(x)$ as a composition of two functions. For example, we can take $f(x) = \frac{x^2-3}{2}$ and $g(x) = \frac{1}{3}x^3$ so $h(x) = g\,(f\,(x))$ [39]. The idea is that instead of going in a single journey from the x axis to the y axis, we now introduce an intermediate axis, the u axis, which is the axis where $f(x)$ lands.

[39] There is more than one option, for example, we could have taken $f(x) = x^2 - 3$ and $g(x) = \frac{1}{24}x^3$

We start on the x axis at the value a with neighborhood $\left(a - \frac{\triangle x}{2}, a + \frac{\triangle x}{2}\right)$ of length $\triangle x$. To send a to the u axis we simply apply f to a to obtain $f(a)$ and the neighborhood $\left(a - \frac{\triangle x}{2}, a + \frac{\triangle x}{2}\right)$ is sent to a neighborhood $\left(f(a) - \frac{\triangle u}{2}, f(a) + \frac{\triangle u}{2}\right)$ of length $\triangle u$ which is obtained by our rescaling factor $f'(a)$:

$$\triangle u \simeq f'(a)\triangle x \tag{422}$$

Given that we are at $f(a)$, to get to the y axis we apply g to $f(a)$ to obtain $g\,(f\,(a))$ and the neighborhood $\left(f(a) - \frac{\triangle u}{2}, f(a) + \frac{\triangle u}{2}\right)$ is sent to a neighborhood $\left(g\,(f\,(a)) - \frac{\triangle y}{2}, g\,(f\,(a)) + \frac{\triangle y}{2}\right)$ of length $\triangle y$ which is obtained by our rescaling factor $g'\,(f\,(a))$: [40]

[40] Observe that $g'(f(a))$ means the derivative g' evaluated at the input $f(a)$

$$\triangle y \simeq g'\,(f\,(a))\,\triangle u \tag{423}$$

Now, $g\,(f\,(a)) = h(a)$ so this is the same as the neighborhood $\left(h(a) - \frac{\triangle y}{2}, h\,(a) + \frac{\triangle y}{2}\right)$. If we substitute 422 into 423 we get

$$\triangle y \simeq g'\,(f\,(a))\,f'(a)\triangle x \tag{424}$$

and comparing to 421 we get

$$h'(a)\triangle x \simeq g'\,(f\,(a))\,f'(a)\triangle x \tag{425}$$

so we should have

$$h'(a) \simeq g'(f(a))\, f'(a) \tag{426}$$

The Chain Rule says that we can actually take the last approximation as an equality.

Chain Rule:
Let f, g be two functions such that f is differentiable at x and g is differentiable at $f(x)$. Then the composite function $h(x) \quad = \quad g(f(x))$ is differentiable at x with derivative

$$h'(x) = [g'(f(x))]\,[f'(x)] \tag{427}$$

In Leibniz's notation, if we write $u = f(x)$ then the chain rule becomes

$$\frac{dh}{dx} = \frac{dh}{du}\frac{du}{dx} \tag{428}$$

Proof. We want to calculate

$$\lim_{\triangle x \longrightarrow 0} \frac{h(x + \triangle x) - h(x)}{\triangle x} = \lim_{\triangle x \longrightarrow 0} \frac{\triangle y}{\triangle x} \tag{429}$$

where we have defined

$$\triangle y \equiv h(x + \triangle x) - h(x) = g(f(x + \triangle x)) - g(f(x)) \tag{430}$$

If we write $u = f(x)$ and define the expressions

$$\begin{aligned} \alpha(\triangle x) &\equiv \tfrac{f(x+\triangle x) - f(x)}{\triangle x} - f'(x) \\ \beta(\triangle u) &\equiv \tfrac{g(u+\triangle u) - g(u)}{\triangle u} - g'(u) \end{aligned} \tag{431}$$

The differentiability of f at x and of g at u implies that

$$\begin{aligned} \lim\nolimits_{\triangle x \longrightarrow 0} \alpha(\triangle x) &= 0 \\ \lim\nolimits_{\triangle u \longrightarrow 0} \beta(\triangle u) &= 0 \end{aligned} \tag{432}$$

On the other hand, it is easy to see from the equations 431 that

$$\begin{aligned} f(x + \triangle x) &= f(x) + [f'(x) + \alpha(\triangle x)]\,\triangle x \\ g(u + \triangle u) &= g(u) + [g'(u) + \beta(\triangle u)]\,\triangle u \end{aligned} \tag{433}$$

We use the equation for $f(x + \triangle x)$ in 433 to rewrite $\triangle y$ as follows

$$\begin{aligned} \triangle y &= h(x + \triangle x) - h(x) \\ &= g(f(x + \triangle x)) - g(f(x)) \\ &= g(f(x) + [f'(x) + \alpha(\triangle x)]\,\triangle x) - g(f(x)) \\ &= g(u + \triangle u) - g(u) \end{aligned} \tag{434}$$

where we took $u = f(x)$ and $\triangle u = [f'(x) + \alpha(\triangle x)] \triangle x$. By the second equation of 433 we can write $\triangle y$ as

$$\begin{aligned} \triangle y &= [g'(u) + \beta(\triangle u)] \triangle u \\ &= [g'(u) + \beta(\triangle u)] [f'(x) + \alpha(\triangle x)] \triangle x \end{aligned} \tag{435}$$

Observe that as $\triangle x \longrightarrow 0$ we have $\triangle u \longrightarrow 0$ so $\beta(\triangle u) \longrightarrow 0$. Therefore, the derivative is

$$\begin{aligned} (g \circ f)'(x) &= \lim_{\triangle x \longrightarrow 0} \frac{\triangle y}{\triangle x} \\ &= \lim_{\triangle x \longrightarrow 0} [g'(u) + \beta(\triangle u)] [f'(x) + \alpha(\triangle x)] \\ &= \lim_{\triangle x \longrightarrow 0} [g'(u) + \beta(\triangle u)] \lim_{\triangle x \longrightarrow 0} [f'(x) + \alpha(\triangle x)] \\ &= g'(u) f'(x) \\ &= g'(f(x)) f'(x) \end{aligned} \tag{436}$$

and the Chain Rule has been proved. □

Example 75. Find the derivative of $f(x) = \left(x^2 + 5\right)^5$

In what follows we will use Leibniz's notation for the chain rule. The idea in all the problems involving the chain rule is to start making change of variables until we recognize a differentiation rule. For example, for $f(x) = \left(x^2 + 5\right)^5$ we don't know the rule for the entire expression, however, we do know the rule for a power like u^5, so we make first the change of variables

$$u = x^2 + 5 \tag{437}$$

so $f(x)$ becomes

$$f(x) = u^5 \tag{438}$$

In this way, an adapted version of 428 says that

$$f'(x) = \frac{df}{dx} = \frac{df}{du}\frac{du}{dx} \tag{439}$$

For $\frac{df}{du}$ we consider f as a function of u,

$$f(u) = u^5 \tag{440}$$

so

$$\frac{df}{du} = 5u^4 \tag{441}$$

The other term is $\frac{du}{dx}$ which is easy to differentiate

$$\frac{du}{dx} = \frac{d}{dx}\left(x^2 + 5\right) = 2x \tag{442}$$

If we substitute both derivatives we find that

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx} = \left(5u^4\right)(2x) = 10x\left(x^2 + 5\right)^4 \tag{443}$$

where we substituted in the last step u in terms of x.

Example 76. Find the derivative of $f(t) = \frac{1}{2}\left(2t^2+t\right)^{-3}$

We just need to find $\frac{d}{dt}\left(2t^2+t\right)^{-3}$ since the constant is easy to handle. Again, we use the change of variables $u = 2t^2+t$ and by the chain rule

$$\frac{d}{dt}\left(2t^2+t\right)^{-3} = \frac{d}{du}\left(2t^2+t\right)^{-3}\frac{du}{dt} = \frac{d}{dt}u^{-3}\frac{d}{dt}\left(2t^2+t\right) = \left(-3u^{-4}\right)(4t+1) = -\frac{3(4t+1)}{\left(2t^2+t\right)^4} \tag{444}$$

Therefore

$$\left(\frac{1}{2}\left(2t^2+t\right)^{-3}\right)' = -\frac{3(4t+1)}{2\left(2t^2+t\right)^4} \tag{445}$$

Example 77. Find the derivative of $f(x) = \sqrt[3]{1-x^2}$

We make in this case the change of variable $u = 1-x^2$

$$\frac{d}{dx}\left(1-x^2\right)^{\frac{1}{3}} = \frac{d}{du}u^{\frac{1}{3}}\frac{du}{dx} = \frac{1}{3}u^{-\frac{2}{3}}\frac{d}{dx}\left(1-x^2\right) = \frac{1}{3\sqrt[3]{\left(1-x^2\right)^2}}(-2x) \tag{446}$$

Example 78. Find the derivative of $f(v) = \left(v^{-3}+4v^{-2}\right)^3$

We make the change of variables $u = v^{-3}+4v^{-2}$

$$\frac{d}{dv}\left(v^{-3}+4v^{-2}\right)^3 = \frac{d}{du}u^3\frac{du}{dv} = 3u^2\frac{d}{dv}\left(v^{-3}+4v^{-2}\right) = 3\left(v^{-3}+4v^{-2}\right)^2\left(-3v^{-4}-8v^{-3}\right) \tag{447}$$

Example 79. Find the derivative of $g(s) = \left(s^2+\frac{1}{s}\right)^{\frac{3}{2}}$

We use the change of variable $u = s^2+\frac{1}{s}$. Then

$$\frac{d}{ds}\left(s^2+\frac{1}{s}\right)^{\frac{3}{2}} = \frac{d}{du}u^{\frac{3}{2}}\frac{du}{ds} = \frac{3}{2}u^{\frac{1}{2}}\frac{d}{ds}\left(s^2+s^{-1}\right) = \frac{3}{2}\sqrt{s^2+\frac{1}{s}}\left(2s-s^{-2}\right) \tag{448}$$

Example 80. Suppose $F(x) = g(f(x))$ and $f(2) = 3$, $f'(2) = -3$, $g(3) = 5$ and $g'(3) = 4$. Find $F'(2)$

We use 427

$$F'(x) = \left[g'\left(f(x)\right)\right]f'(x) \tag{449}$$

When $x = 2$ we have

$$F'(2) = \left[g'(f(2))\right]f'(2) = \left[g'(3)\right](-3) = -3(4) = -12 \tag{450}$$

Example 81. Suppose $F(x) = f\left(x^2+1\right)$. Find $F'(1)$ if $f'(2) = 3$

If we call $g(x) = x^2 + 1$ then we have that

$$F(x) = f \circ g(x) \tag{451}$$

By 427 we have

$$F'(x) = [f'(g(x))]\, g'(x) = \left[f'(x^2+1)\right] g'(x) = \left[f'(x^2+1)\right](2x) \tag{452}$$

when $x = 1$ we have

$$F'(1) = [f'(2)]\,(2) = 3\,(2) = 6 \tag{453}$$

Example 82. Find an equation of the tangent line to the graph of the function at the given point. $f(x) = x\sqrt{2x^2+7}$, point $(3, 15)$

The derivative $f'(x)$ will be the slope. Before applying the chain rule, we apply the product rule

$$f'(x) = x'\sqrt{2x^2+7} + x\left(\sqrt{2x^2+7}\right)' = \sqrt{2x^2+7} + x\left(\sqrt{2x^2+7}\right)' \tag{454}$$

To find $\left(\sqrt{2x^2+7}\right)'$ we make the change of variable $u = 2x^2+7$. Then

$$\frac{d}{dx}\left(\sqrt{2x^2+7}\right)' = \frac{d}{du}\sqrt{u}\frac{du}{dx} = \frac{1}{2\sqrt{u}}\frac{d}{dx}\left(2x^2+7\right) = \frac{1}{2\sqrt{2x^2+7}}(4x) = \frac{2x}{\sqrt{2x^2+7}} \tag{455}$$

Therefore

$$f'(x) = \sqrt{2x^2+7} + \frac{2x^2}{\sqrt{2x^2+7}} = \frac{4x^2+7}{\sqrt{2x^2+7}} \tag{456}$$

the slope we need is $f'(3)$

$$f'(3) = \frac{4\,(9)+7}{\sqrt{2\,(9)+7}} = \frac{43}{5} \tag{457}$$

The equation of the tangent line will be $y = \frac{43}{5}x + b$. We use the fact that it must pass through $(3, 15)$ so $15 = \frac{129}{5} + b$ which gives $b = -\frac{54}{5}$. So the equation is

$$y = \frac{43}{5}x - \frac{54}{5} \tag{458}$$

Example 83. Let $f(x)$ and $g(x)$ be differentiable functions with the following information: $f(1) = 3, g(1) = 2$, $f'(1) = 5, g'(1) = -2$, $f'(5) = 4, g'(5) = 1$. Evaluate $k'(1)$ where $k(x) = f\left(g(x) + f(x)\right)$

If we call $h(x) = g(x) + f(x)$ then $k(x) = f\left(h(x)\right)$. By the chain rule 427

$$k'(x) = [f'(h(x))]\, h'(x) = [f'\left(g(x) + f(x)\right)]\left(g'(x) + f'(x)\right) \tag{459}$$

When $x = 1$ we have

$$k'(1) = [f'(g(1) + f(1))]\left(g'(1) + f'(1)\right) = [f'(2+3)]\,(-2+5) = 3f'(5) = 12 \tag{460}$$

Example 84. Find the derivative of $\left(\left(\left(x^3+2\right)^3+2\right)^3+2\right)^3$.

For this problem we are going to require more than one change of variables so we start taking $u_1 = \left(\left(x^3+2\right)^3+2\right)^3+2$. If we call $f(x) = \left(\left(\left(x^3+2\right)^3+2\right)^3+2\right)^3$ then

$$\frac{df}{dx} = \frac{d}{du_1}u_1^3\frac{du_1}{dx} = 3u_1^2\frac{d}{dx}\left(\left(\left(x^3+2\right)^3+2\right)^3+2\right) = 3\left(\left(\left(x^3+2\right)^3+2\right)^3+2\right)^2\frac{d}{dx}\left(\left(x^3+2\right)^3+2\right)^3 \tag{461}$$

To find $\frac{d}{dx}\left(\left(x^3+2\right)^3+2\right)^3$ we make the change of variables $u_2 = \left(x^3+2\right)^3+2$. Then

$$\frac{d}{dx}\left(\left(x^3+2\right)^3+2\right)^3 = \frac{d}{du_2}u_2^3\frac{du_2}{dx} = 3u_2^2\frac{d}{dx}\left(\left(x^3+2\right)^3+2\right) = 3\left(\left(x^3+2\right)^3+2\right)^2\frac{d}{dx}\left(x^3+2\right)^3 \tag{462}$$

Finally, we make the change of variables $u_3 = x^3+2$. Then

$$\frac{d}{dx}\left(x^3+2\right)^3 = \frac{d}{du_3}u_3^3\frac{du_3}{dx} = 3u_3^2\frac{d}{dx}\left(x^3+2\right) = 3\left(x^3+2\right)^2\left(3x^2\right) = 9x^2\left(x^3+2\right)^2 \tag{463}$$

This gives

$$\frac{d}{dx}\left(\left(x^3+2\right)^3+2\right)^3 = 3\left(\left(x^3+2\right)^3+2\right)^2\left[9x^2\left(x^3+2\right)^2\right] = 27x^2\left(x^3+2\right)^2\left(\left(x^3+2\right)^3+2\right)^2 \tag{464}$$

and finally

$$\begin{aligned}\tfrac{df}{dx} &= 3\left(\left(\left(x^3+2\right)^3+2\right)^3+2\right)^2\left[27x^2\left(x^3+2\right)^2\left(\left(x^3+2\right)^3+2\right)^2\right] \\ &= \left[9x\left(x^3+2\right)\left(\left(x^3+2\right)^3+2\right)\left(\left(\left(x^3+2\right)^3+2\right)^3+2\right)\right]^2\end{aligned} \tag{465}$$

Example 85. Find the derivative of $f(x) = x^2\sin\left(9x^2+1\right)$

We use the product rule and the chain rule

$$\begin{aligned}f'(x) &= \left(x^2\right)'\sin\left(9x^2+1\right) + x^2\left(\sin\left(9x^2+1\right)\right)' \\ &= 2x\sin\left(9x^2+1\right) + x^2\left(\cos\left(9x^2+1\right)\right)(18x) \\ &= 2x\sin\left(9x^2+1\right) + 18x^3\cos\left(9x^2+1\right)\end{aligned} \tag{466}$$

Example 86. Find the derivative of $g(x) = \frac{\tan(x)}{\sqrt{3x+1}}$

We use the quotient rule and the chain rule

$$\begin{aligned} g'(x) &= \frac{(\tan x)'\sqrt{3x+1}-(\tan(x))\left(\sqrt{3x+1}\right)'}{\left(\sqrt{3x+1}\right)^2} \\ &= \frac{\left(\sec^2 x\right)\sqrt{3x+1}-\frac{3\tan x}{2\sqrt{3x+1}}}{3x+1} \\ &= \frac{2(3x+1)\left(\sec^2 x\right)-3\tan x}{2(3x+1)^{3/2}} \end{aligned} \tag{467}$$

Example 87. Show that the tangent line to the graph of $e^{\sqrt{x}}$ at any point is not horizontal

By the chain rule the derivative of the exponential with $u = \sqrt{x}$ is

$$\frac{d}{dx}e^{\sqrt{x}} = \frac{d}{du}e^u\frac{du}{dx} = e^u\frac{1}{2\sqrt{x}} = \frac{e^{\sqrt{x}}}{2\sqrt{x}} \tag{468}$$

Since the exponential is never 0 the derivative can't be zero so the slope is never zero and the tangent lines are not horizontal

---

Example 88. Find the derivative of $f(x) = e^{x^3-3x^2}$

We use the chain rule with $u = x^3 - 3x^2$.

$$\frac{df}{dx} = \frac{de^u}{du}\frac{du}{dx} = e^u\left(3x^2-6x\right) = e^{x^3-3x^2}\left(3x^2-6x\right) \tag{469}$$

Example 89. Find the derivative of $f(x) = x^e e^{\sqrt{x}}$

We use the product rule

$$\frac{df}{dx} = \frac{dx^e}{dx}e^{\sqrt{x}} + x^e\frac{d}{dx}e^{\sqrt{x}} = ex^{e-1}e^{\sqrt{x}} + x^e\frac{d}{dx}e^{\sqrt{x}} \tag{470}$$

To find $\frac{d}{dx}e^{\sqrt{x}}$ we use the chain rule with $u = \sqrt{x}$

$$\frac{d}{dx}e^{\sqrt{x}} = \frac{de^u}{du}\frac{du}{dx} = e^u\frac{1}{2\sqrt{x}} = \frac{e^{\sqrt{x}}}{2\sqrt{x}} \tag{471}$$

Therefore

$$\frac{df}{dx} = ex^{e-1}e^{\sqrt{x}} + x^e\frac{e^{\sqrt{x}}}{2\sqrt{x}} \tag{472}$$

Example 90. Find the derivative of $h(x) = \ln\left((8x+4)^3\right)$

We call $u = (8x+4)^3$ and apply the chain rule

$$\frac{dh}{dx} = \frac{d\ln u}{du}\frac{du}{dx} = \frac{1}{u}\frac{du}{dx} = \frac{1}{u}24\,(8x+4)^2 = \frac{24\,(8x+4)^2}{(8x+4)^3} = \frac{24}{8x+4} = \frac{6}{2x+1} \tag{473}$$

Implicit Differentiation and Differentials

Up to this point we have learned to differentiate functions in which y is given as a function of x, that is, $y = f(x)$. Now, in many situations y is not given *explicitly* as a function of x, but it is given *implicitly* as a function of x.

For example, consider the equation of the circle centered at the origin of radius 3, $x^2 + y^2 = 9$. Here y is not given as a function of x explicitly, but there is a relationship between x and y which *in principle* we could use to write y as a function of x. From the equation $x^2 + y^2 = 9$ we have that $y^2 = 9 - x^2$ and if we take square roots on both sides we have $y = \pm\sqrt{9 - x^2}$ where we need to use $\pm$ because y can be positive or negative. Therefore, when we write y as a function of x there are actually two different functions, $y(x) = \sqrt{9 - x^2}$ and $y(x) = -\sqrt{9 - x^2}$ and we can differentiate using the chain rule to obtain $y'(x) = -\frac{x}{\sqrt{9-x^2}}$ and $y(x) = \frac{x}{\sqrt{9-x^2}}$.

However, for most relationships between x and y it is not very practical to write y as a function of x, for example, in the equation $xy - y^2 + x^2 + y^3 = 1$ it is not clear how to write y as a function of x explicitly but in any way if we are only interested in finding $y'(x)$ there is a way of doing this without finding $y(x)$ explicitly and this method is called implicit differentiation.

To see how this works, we will find $y'(x)$ for the circle without using the explicit relationship for y in terms of x. From the equation $x^2 + y^2 = 9$ we know that y is a function of x so if we want we can write $x^2 + (y(x))^2 = 9$. In the method of implicit differentiation to find $y'(x) = \frac{dy}{dx}$ we differentiate *both sides* of the equation with respect to x, that is,

$$\frac{d}{dx}\left(x^2 + (y(x))^2\right) = \frac{d}{dx}9 \tag{474}$$

Observe that differentiating both sides of an equation is no different from multiplying both sides of an equation by a number or squaring both sides of an equation. From 474 we have

$$2x + \frac{d}{dx}(y(x))^2 = 0 \tag{475}$$

to find $\frac{d}{dx}(y(x))^2$ we can make the change of variables $u = y(x)$ and use the chain rule

$$\frac{d}{dx}\left(y(x)\right)^2 = \frac{du^2}{du}\frac{du}{dx} = 2u\frac{du}{dx} = 2y\frac{dy}{dx} = 2yy' \tag{476}$$

Substituting in 475 we have

$$2x + 2yy' = 0 \tag{477}$$

or

$$y' = -\frac{x}{y} \tag{478}$$

This does not seem exactly the same as the previous derivatives we found but if we use that $y = \pm\sqrt{9 - x^2}$ then 478 is the same expression as the one found before. Observe that even tough we did use the explicit form of y as a function of x, we did that *after* differentiating not before differentiating, and that is the big difference with trying to find first y as a function of x. In practice, since finding y as a function of x is difficult, we will be satisfied with writing y' in the form 478.

It would be nice to have a way to solve this problem while treating the variables x and y symmetrically, after all, x and y have the same role in the equation $x^2 + y^2 = 9$. Fortunately, there is such a method, called the method of **differentials**. Trying to give a formal definition of differentials is beyond the scope of this course, however, for now they can be considered as what happens with the linear approximation $\triangle y \simeq f'(x)\triangle x$ if we ignore higher order terms like $(\triangle x)^2$, $(\triangle x)^3$, etc. We will write

$$dy = f'(x)dx \tag{479}$$

to indicate that the linear approximation has taken place and we have ignored the higher order terms. Observe that this seems trivial if we use Leibniz's notation for the derivative $\frac{dy}{dx} = f'(x)$ because it seems as if we just sent dx to the other side. However, this is precisely the reason why Leibniz's notation is so useful: it is a mnemonic device that help us remember the correct relationships between different quantities.

Rules for Differentials: given variables x, y, z, etc. their differentials are written dx, dy, dz, etc. They satisfy the following properties:

- **Linearity**: $d(x+y) = dx + dy$
- The differential of a constant is zero: $dc = 0$
- **Product rule**: let u, v be variables. Then

$$d(uv) = (du)v + u(dv) \tag{480}$$

- **Quotient rule:** let u, v be variables. Then

$$d\left(\frac{u}{v}\right) = \frac{(du)v - u(dv)}{v^2} \tag{481}$$

- **Differential of a function:** suppose that the variable u can be written as a function of another variable, that is, $u = u(x)$. Then

$$du = u'(x)dx \tag{482}$$

For example

$$\begin{aligned} u = x^n &\longrightarrow du = nx^{n-1}dx \\ u = e^x &\longrightarrow du = e^x dx \\ u = \ln(x) &\longrightarrow du = \tfrac{dx}{x} \\ u = \sin x &\longrightarrow du = \cos x dx \\ u = \cos x &\longrightarrow du = -\sin x dx \end{aligned} \tag{483}$$

- **Cancellation property**: if $udx = vdx$ then $u = v$. In particular, if $udx = 0$ then $u = 0$.
- **Differential of equations:** If $u = v$ then $du = dv$

We will now show how the method of differentials can help us solve the problem of the circle. If we start with $x^2 + y^2 = 9$ by properties of differentials we can take differentials on both sides of the equations to get

$$d\left(x^2 + y^2\right) = d(9) \tag{484}$$

Now we use linearity and that differentials of constants are zero to obtain

$$dx^2 + dy^2 = 0 \tag{485}$$

To find dx^2 and dy^2 we use the formulas for differential of functions 483 to obtain

$$2xdx + 2ydy = 0 \tag{486}$$

Canceling 2 on both sides of the equations we have the following equation

$$xdx + ydy = 0 \tag{487}$$

To recover 478 we assume that y is a function of x, that is, $y = y(x)$. Then we can use the formula 482 to write

$$dy = y'(x)dx \tag{488}$$

so we can substitute in equation 487 to obtain

$$xdx + yy'(x)dx = 0 \tag{489}$$

We can factorize dx in the last equation to get

$$(x + yy'(x))\,dx = 0 \tag{490}$$

and the cancellation property of differentials says that

$$x + yy'(x) = 0 \tag{491}$$

so

$$y'(x) = -\frac{x}{y} \tag{492}$$

which is precisely equation 478.

The advantage of using differentials is that we treated x and y in a symmetric way until the very end, where we assumed that y was a function of x. On the other hand, we could have assumed just as easily that x is a function of y, that is, $x = x(y)$. In this case dx can be calculated as

$$dx = x'(y)dy \tag{493}$$

and replacing in equation 487 we get

$$xx'(y)dy + ydy = 0 \tag{494}$$

If we factor dy we obtain

$$(xx'(y) + y)\,dy = 0 \tag{495}$$

and by the cancellation property we get

$$xx'(y) + y = 0 \tag{496}$$

so

$$x'(y) = -\frac{y}{x} \tag{497}$$

Example 91. Find an equation of the tangent line to the graph of the function f defined by the equation at the indicated point: $x^2y^3 - y^2 + xy - 1 = 0$ at point $(1, 1)$

We need to find y' first by differentiating implicitly

$$\frac{d}{dx}\left(x^2y^3 - y^2 + xy - 1\right) = \frac{d}{dx}0 \tag{498}$$

we use the product rule

$$2xy^3 + x^2\frac{d}{dx}y^3 - \frac{d}{dx}y^2 + y + x\frac{d}{dx}y = 0 \tag{499}$$

if we use the chain rule we can see that $\frac{d}{dx}y^3 = 3y^2\frac{d}{dx}y$ and $\frac{d}{dx}y^2 = 2y\frac{d}{dx}y$ so we have

$$2xy^3 + 3x^2y^2y' - 2yy' + y + xy' = 0 \tag{500}$$

We group together the terms with y'

$$y'\left(3x^2y^2 - 2y + x\right) = -2xy^3 - y \tag{501}$$

so

$$y' = \frac{-2xy^3 - y}{3x^2y^2 - 2y + x} \tag{502}$$

At the point $(1,1)$ we have $x = 1$ and $y = 1$ so y' becomes

$$y' = \frac{-3}{2} \tag{503}$$

The equation of the tangent line is $y = -\frac{3}{2}x + b$ and since it passes through $(1,1)$ it becomes $y = -\frac{3}{2}x + \frac{5}{2}$.

If we use the method of differentials we take differentials on both sides of the equation $x^2y^3 - y^2 + xy - 1 = 0$ to obtain

$$d\left(x^2y^3 - y^2 + xy - 1\right) = d(0) \tag{504}$$

Now we use linearity and the differential of a constant to get

$$d\left(x^2y^3\right) - d(y^2) + d(xy) = 0 \tag{505}$$

Now the product rule and differential of a function rule gives

$$d\left(x^2\right)y^3 + x^2d\left(y^3\right) - 2ydy + (dx)y + x(dy) = 0 \tag{506}$$

or

$$2xy^3dx + 3x^2y^2dy - 2ydy + ydx + xdy = 0 \tag{507}$$

For simplicity we group together all the terms dx together and all the terms with dy together

$$\left(2xy^3 + y\right)dx + \left(3x^2y^2 - 2y + x\right)dy = 0 \tag{508}$$

If we want to write y as a function of x, $y = y(x)$, then $dy = y'(x)dx$ and we obtain

$$\left(2xy^3 + y\right)dx + \left(3x^2y^2 - 2y + x\right)y'(x)dx = 0 \tag{509}$$

and using the cancellation property of the differential it is clear that we get 502.

On the other hand, if we want to write x as a function of y, $x = x(y)$ then $dx = x'(y)dy$ and we obtain

$$\left(2xy^3 + y\right)x'(y)dy + \left(3x^2y^2 - 2y + x\right)dy = 0 \tag{510}$$

We can use the cancellation property to get

$$x'(y) = -\frac{\left(3x^2y^2 - 2y + x\right)}{2xy^3 + y} \tag{511}$$

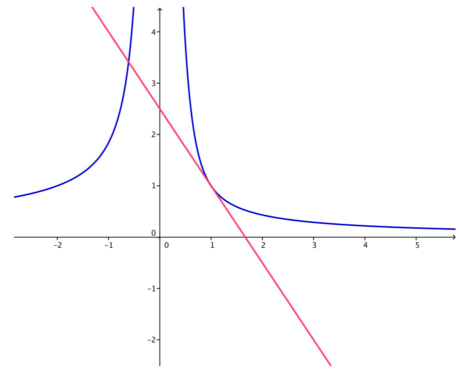


Figure 104: Curve $x^2y^3 - y^2 + xy - 1 = 0$ and tangent line at $(1,1)$

Example 92. Find an equation of the tangent line to the graph of the function f defined by the equation at the indicated point. $(x-y-1)^3 = x$ at point $(1,-1)$

Again, we find y' by differentiating implicitly.

$$\frac{d}{dx}(x-y-1)^3 = \frac{d}{dx}x \tag{512}$$

we make the change of variable $u = x-y-1$ so

$$\frac{d}{dx}(x-y-1)^3 = \frac{d}{du}u^3\frac{du}{dx} = 3u^2\frac{d}{dx}(x-y-1) = 3(x-y-1)^2(1-y') \tag{513}$$

Therefore,

$$3(x-y-1)^2(1-y') = \frac{d}{dx}x = 1 \tag{514}$$

which gives

$$1-y' = \frac{1}{3(x-y-1)^2} \tag{515}$$

or

$$y' = 1-\frac{1}{3(x-y-1)^2} \tag{516}$$

At point $(1,-1)$ we have that

$$y' = 1-\frac{1}{3(1+1-1)^2} = 1-\frac{1}{3} = \frac{2}{3} \tag{517}$$

The equation of the tangent line is $y = \frac{2}{3}x+b$ and since it passes through $(1,-1)$ it must be $y = \frac{2}{3}x-\frac{5}{3}$.

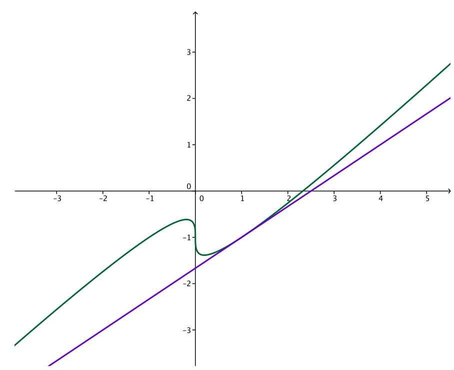


Figure 105: Curve $(x-y-1)^3 = x$ and tangent line at $(1,-1)$

To use differentials in the equation $(x-y-1)^3 = x$ we introduce a variable $u = x-y-1$ so that the equation becomes

$$u^3 = x \tag{518}$$

Taking differentials on both sides

$$3u^2du = dx \tag{519}$$

but since $u = x-y-1$ we get $du = dx-dy$ or

$$3(x-y-1)^2(dx-dy) = dx \tag{520}$$

We group the dx, dy terms together

$$\left(3(x-y-1)^2-1\right)dx - 3(x-y-1)^2dy = 0 \tag{521}$$

using the relationships $dy = y'(x)dx$ or $dx = x'(y)dy$ we can find whichever derivative we want.

Example 93. Find $\frac{dy}{dx}$ by implicit differentiation: $\sqrt{x+y} = x$.

First of all, with $u = x + y$ we have

$$\frac{d}{dx}\sqrt{x+y} = \frac{d}{du}\sqrt{u}\frac{du}{dx} = \frac{1}{2\sqrt{u}}\frac{d}{dx}(x+y) = \frac{1}{2\sqrt{x+y}}(1+y') \quad (522)$$

so if we differentiate $\sqrt{x+y} = x$ on both sides we have

$$\frac{1}{2\sqrt{x+y}}(1+y') = 1 \quad (523)$$

or

$$y' = 2\sqrt{x+y} - 1 \quad (524)$$

This can also be solved with differentials by using $u = x + y$ to get the equation $\sqrt{u} = x$ and

$$\frac{1}{2\sqrt{u}}du = dx \quad (525)$$

substituting back u we get

$$\frac{1}{2\sqrt{x+y}}(dx + dy) = dx \quad (526)$$

using either $dx = x'(y)dy$ or $dy = y'(x)dx$ we find $x'(y)$ or $y'(x)$.

Example 94. Find $\frac{dy}{dx}$ by implicit differentiation: $xy^{\frac{3}{2}} = x^2 + y^2$

If we differentiate both sides, use the product rule and the chain rule we have

$$y^{\frac{3}{2}} + x\frac{d}{dx}y^{\frac{3}{2}} = 2x + 2yy' \quad (527)$$

Also, by the chain rule it can be seen that $\frac{d}{dx}y^{\frac{3}{2}} = \frac{3}{2}y^{\frac{1}{2}}y'$ so

$$y^{\frac{3}{2}} + \frac{3}{2}xy^{\frac{1}{2}}y' = 2x + 2yy' \quad (528)$$

which gives

$$y' = \frac{2x - y^{\frac{3}{2}}}{\frac{3}{2}xy^{\frac{1}{2}} - 2y} \quad (529)$$

Again, we can also solve this problem by taking differentials

$$y^{\frac{3}{2}}dx + \frac{3}{2}xy^{\frac{1}{2}}dy = 2xdx + 2ydy \quad (530)$$

and then use either $dx = x'(y)dy$ or $dy = y'(x)dx$ to find $x'(y)$ or $y'(x)$.

Derivatives of Inverse Functions

Exponentials and Logarithms

We mentioned before that the derivative of $\ln x$ can be found using implicit differentiation. To see how this is done call

$$y = \ln x \tag{531}$$

This equation is the same as

$$e^y = x \tag{532}$$

and if we take differentials on both sides we obtain

$$e^y dy = dx \tag{533}$$

But $dy = (\ln x)' dx$ so the last equation is the same as

$$e^y (\ln x)' dx = dx \tag{534}$$

and the cancellation property says that

$$e^y (\ln x)' = 1 \tag{535}$$

or

$$(\ln x)' = \frac{1}{e^y} = \frac{1}{x} \tag{536}$$

which is what we wanted to show. A similar procedure can be used to find the derivatives of the inverse trigonometric functions as we will know show.

Derivatives of Inverse Trigonometric Functions

Suppose we want to find the derivative of $y = \arccos x$. By the definition of the inverse function this equation is the same as

$$\cos y = x \tag{537}$$

and if we take differentials on both sides we obtain

$$-\sin y dy = dx \tag{538}$$

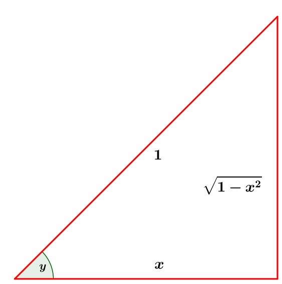


Figure 106: Derivative of $y = \arccos x$

On the other hand, we have that $dy = (\arccos x)'\, dx$ so we can substitute in the previous equation to obtain

$$-\sin y\, (\arccos x)'\, dx = dx \tag{539}$$

By the cancellation property we obtain

$$(\arccos x)' = -\frac{1}{\sin y} = -\frac{1}{\sqrt{1-x^2}} \tag{540}$$

In a similar way we can find [41]

[41] The formulas for the derivatives of $\operatorname{arccsc} x$ and $\operatorname{arcsec} x$ depend on the domain which is used to define them up to a sign so they won't be included in the list

Derivatives of Inverse Trigonometric Functions:

$$\begin{aligned} \frac{d}{dx}\arcsin x &= \frac{1}{\sqrt{1-x^2}} & \frac{d}{dx}\operatorname{arccsc} x &= -\frac{1}{x\sqrt{x^2-1}} \\ \frac{d}{dx}\arccos x &= -\frac{1}{\sqrt{1-x^2}} & \frac{d}{dx}\operatorname{arcsec} x &= \frac{1}{x\sqrt{x^2-1}} \\ \frac{d}{dx}\operatorname{arctan} x &= \frac{1}{1+x^2} & \frac{d}{dx}\operatorname{arccot} x &= -\frac{1}{1+x^2} \end{aligned} \tag{541}$$

General Formula for Derivative of the Inverse

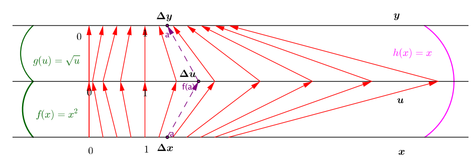


Figure 107: Derivative of the function and its inverse

Suppose we have a function $u = f(x)$ that can be inverted in some neighborhood of a point a. [42] We call the inverse function $f^{-1}(x)$, although it should be emphasized that f^{-1} is not the same as $\frac{1}{f}$. For simplicity call $g = f^{-1}$. If we use the interpretation of the derivative as a rescaling factor it should be pretty clear from the picture that

[42] An inverse function may not exist for the entire domain, like in the case of the trigonometric functions

$$g'(f(a)) = \frac{1}{f'(a)} \tag{542}$$

To see why this should be the case we give an argument using the chain rule before giving the proof. The defining property of the derivative is

$$g(f(x)) = x \tag{543}$$

If we differentiate both sides with respect to x we obtain

$$g'(f(x))\, f'(x) = 1 \tag{544}$$

so

$$g'(f(x)) = \frac{1}{f'(x)} \tag{545}$$

which is 542.

Derivative of the Inverse: Suppose that f is invertible in a neighborhood of a with inverse $g = f^{-1}$. Suppose that f is differentiable at a, with derivative $f'(a) \neq 0$ and suppose that g is continuous at $f(a)$. Then g is differentiable at $f(a)$ with derivative

$$\left(f^{-1}\right)'(f(a)) = \frac{1}{f'(a)} \tag{546}$$

Proof. By the limit definition

$$\begin{aligned} g'(f(a)) &= \lim_{\triangle u \longrightarrow 0} \tfrac{g(f(a)+\triangle u)-g(f(a))}{\triangle u} \\ &= \lim_{\triangle u \longrightarrow 0} \tfrac{g(f(a)+\triangle u)-a}{\triangle u} \end{aligned} \tag{547}$$

Now, call

$$\triangle x = g(f(a)+\triangle u) - a \tag{548}$$

Because g is continuous at $f(a)$ we have that as $\triangle u \longrightarrow 0$, $\triangle x \longrightarrow 0$. Moreover, the last equation is the same as

$$\triangle x + a = f^{-1}(f(a)+\triangle u) \tag{549}$$

and applying f to both sides and the cancellation property $f\left(f^{-1}u\right) = u$ we obtain for $u = f(a)+\triangle u$ that

$$f(\triangle x + a) = f(a) + \triangle u \tag{550}$$

or

$$\triangle u = f(a+\triangle x) - f(a) \tag{551}$$

In terms of $\triangle x$ the limit 547 can be written as

$$\begin{aligned} g'(f(a)) &= \lim_{\triangle u \longrightarrow 0} \tfrac{g(f(a)+\triangle u)-a}{\triangle u} \\ &= \lim_{\triangle x \longrightarrow 0} \tfrac{\triangle x}{f(a+\triangle x)-f(a)} \\ &= \lim_{\triangle x \longrightarrow 0} \tfrac{1}{\frac{f(a+\triangle x)-f(a)}{\triangle x}} \\ &= \tfrac{1}{f'(a)} \end{aligned} \tag{552}$$

□

As an application of the previous theorem, if we take $f(a) = \cos x$ then formula 546 says that

$$(\arccos)'(\cos a) = \tfrac{1}{-\sin a} \tag{553}$$

If we let $x = \cos a$ then $\sin a = \sqrt{1-\sin^2 a} = \sqrt{1-x^2}$ so we obtain

$$(\arccos)' x = -\tfrac{1}{\sqrt{1-x^2}} \tag{554}$$

which is the formula we found before.

Higher Order Derivatives

If we start with a function $y = f(x)$ the derivative $y' = f'(x)$ can be considered as a new function. For example, if $y = x^3$ then $y' = 3x^2$ and we may want to differentiate y' to obtain $\left(3x^2\right)' = 6x$. Since $6x$ is the result of differentiating x^3 two times we say that $6x$ is the second derivative of x^3 and we write $6x = \left(x^3\right)''$.

- If $y = f(x)$ is a function its first derivative is $y' = f'(x)$. If we consider y' as a function and differentiate it again we call the result $y'' = f''(x)$ and say that y'' is the second derivative of y. Similarly, if y'' is differentiable then we write y''' for the derivative of y'' and say that y''' is the third derivative of y.
- For convenience, if we continue this process instead of writing y'''' we write $y^{(4)}$, so $y^{(n)}$ denotes differentiating y n times.
- In Leibniz's notation, we denote the higher derivatives by

$$\frac{d^2y}{dx^2} \qquad \frac{d^3y}{dx^3} \qquad \frac{d^ny}{dx^n} \tag{555}$$

etc.

We will see later that the higher derivatives have a geometric interpretation, in particular the second derivative, but for now we will do some examples of finding higher derivatives to show that there is nothing new to learn about how to calculate them.

Example 95. Find the second derivative of $g(t) = t^2\left(3t+1\right)^4$

First we need to find $g'(t)$. We use the product rule and the chain rule

$$g'(t) = 2t\left(3t+1\right)^4 + t^2\frac{d}{dt}\left(3t+1\right)^4 = 2t\left(3t+1\right)^4 + 12t^2\left(3t+1\right)^3$$
$$= \left(3t+1\right)^3\left(2t\left(3t+1\right)+12t^2\right) = \left(3t+1\right)^3\left(18t^2+2t\right) \tag{556}$$

where we factorized $\left(3t+1\right)^3$ because it facilitates differentiating again.

To find $g''(t)$ we use the product rule and the chain rule

$$\begin{aligned} g''(t) &= \left(\tfrac{d}{dt}(3t+1)^3\right)\left(18t^2+2t\right)+(3t+1)^3(36t+2) \\ &= 9(3t+1)^2\left(18t^2+2t\right)+(3t+1)^3(36t+2) \end{aligned} \tag{557}$$

Example 96. Find the third derivative of $g(s)=\sqrt{3s-2}$

We apply the chain rule to find the first derivative

$$\frac{dg}{ds}=\frac{3}{2\sqrt{3s-2}}=\frac{3}{2}(3s-2)^{-\frac{1}{2}} \tag{558}$$

Now we differentiate again with the chain rule

$$\frac{d^2g}{ds^2}=-\frac{9}{4}(3s-2)^{-\frac{3}{2}} \tag{559}$$

We differentiate one more time with the chain rule

$$\frac{d^3g}{ds^3}=\frac{81}{8}(3s-2)^{-\frac{5}{2}} \tag{560}$$

Example 97. The distance s (in feet) covered by a car after t sec is given by $s=-t^3+8t^2+20t$, $0\le t\le 6$. Find a general expression for the car's acceleration at any time t, $0\le t\le 6$. Show that the car is decelerating after $2\frac{2}{3}$ sec.

Since the velocity is $v=\frac{ds}{dt}$ the acceleration is $a=\frac{dv}{dt}=\frac{d^2s}{dt^2}$ so we have to differentiate the position twice with respect to time.

$$v=\frac{ds}{dt}=-3t^2+16t+20 \tag{561}$$

$$a=\frac{dv}{dt}=\frac{d^2s}{dt^2}=-6t+16 \tag{562}$$

The car is decelerating when $a<0$, that is when $-6t+16<0$ or $\frac{16}{6}<t$, so it is decelerating after $2\frac{2}{3}$ seconds.

Example 98. Find the second derivative $\frac{d^2y}{dx^2}$ of each of the functions defined implicitly by the equation: $y^2-xy=8$

First we differentiate with respect to x using the chain rule and the product rule

$$2y\frac{dy}{dx}-y-x\frac{dy}{dx}=0 \tag{563}$$

this gives

$$y'=\frac{y}{2y-x} \tag{564}$$

To find y'' we differentiate y' using the quotient rule

$$y''=\frac{y'(2y-x)-y(2y'-1)}{(2y-x)^2}=\frac{\left(\frac{y}{2y-x}\right)(2y-x)-y\left(\frac{2y}{2y-x}-1\right)}{(2y-x)^2}=\frac{2y(y-x)}{(2y-x)^3} \tag{565}$$

Example 99. Find the second derivative $\frac{d^2y}{dx^2}$ of each of the functions defined implicitly by the equation: $x^{1/3} + y^{1/3} = 1$

By the chain rule and the power rule

$$\frac{1}{3}x^{-\frac{2}{3}} + \frac{1}{3}y^{-\frac{2}{3}}y' = 0 \tag{566}$$

so

$$y' = -\frac{x^{-\frac{2}{3}}}{y^{-\frac{2}{3}}} = -\frac{y^{\frac{2}{3}}}{x^{\frac{2}{3}}} \tag{567}$$

We use the quotient rule to find y''

$$\begin{aligned} y'' &= -\frac{\left(y^{\frac{2}{3}}\right)'\left(x^{\frac{2}{3}}\right) - y^{\frac{2}{3}}\left(\frac{2}{3}x^{-\frac{1}{3}}\right)}{x^{\frac{4}{3}}} = -\frac{2}{3}\frac{y^{-\frac{1}{3}}y'x^{\frac{2}{3}} - y^{\frac{2}{3}}x^{-\frac{1}{3}}}{x^{\frac{4}{3}}} \\ &= -\frac{2}{3}\frac{-y^{-\frac{1}{3}}\frac{y^{\frac{2}{3}}}{x^{\frac{2}{3}}}x^{\frac{2}{3}} - y^{\frac{2}{3}}x^{-\frac{1}{3}}}{x^{\frac{4}{3}}} = \frac{2}{3}\frac{y^{\frac{1}{3}} + y^{\frac{2}{3}}x^{-\frac{1}{3}}}{x^{\frac{4}{3}}} \end{aligned} \tag{568}$$

Example 100. Find the second derivative of $f(t) = t^2e^{-2t}$

The first derivative is

$$f'(t) = \left(t^2\right)' e^{-2t} + t^2\left(e^{-2t}\right)' = 2te^{-2t} + t^2\left(e^{-2t}\right)(-2) = \left(2t - 2t^2\right)e^{-2t} \tag{569}$$

The second derivative is

$$\begin{aligned} f''(t) &= \left(2t - 2t^2\right)' e^{-2t} + \left(2t - 2t^2\right)\left(e^{-2t}\right)' \\ &= (2 - 4t)e^{-2t} + \left(2t - 2t^2\right)\left(e^{-2t}\right)(-2) = \left(2 - 8t + 4t^2\right)e^{-2t} \end{aligned} \tag{570}$$

Exponential Models and Logarithmic Differentiation

Logarithmic Differentiation

In the method of logarithmic differentiation we find a derivative by taking the (natural) logarithm on both sides of the equation and then differentiating.

Example 101. Find the derivative of $y = 2^x$.

To find the derivative we take natural log on both sides of the equation

$$\ln y = \ln\left(2^x\right) = x \ln 2 \tag{571}$$

Now we differentiate with respect to x and use the chain rule

$$\frac{d \ln y}{dy}\frac{dy}{dx} = \ln 2 \tag{572}$$

and using the formula for the derivative of the logarithm we have

$$\frac{1}{y}y' = \ln 2 \tag{573}$$

so

$$\left(2^x\right)' = 2^x \ln 2 \tag{574}$$

Example 102. Let $f(x) = \left(1+x^2\right)^{\ln(1-x)}$. a) What is the domain of $f(x)$? b) Use logarithmic differentiation to find $f'(x)$

a) The only problem that $f(x)$ can have is when $\ln(1-x)$ becomes undefined. To avoid this from happening we need $1 - x > 0$ so $1 > x$. Therefore, the domain for $f(x)$ is $(-\infty, 1)$

b) Call $y = \left(1+x^2\right)^{\ln(1-x)}$. Taking natural log on both sides

$$\ln y = [\ln(1-x)]\left[\ln(1+x^2)\right] \tag{575}$$

differentiating with respect to x, using the product rule and chain rule

$$\frac{y'}{y} = \left(\frac{d}{dx}\ln(1-x)\right)\ln(1+x^2) + \ln(1-x)\frac{d}{dx}\ln(1+x^2) = -\frac{\ln(1+x^2)}{1-x} + \ln(1-x)\frac{2x}{1+x^2} \tag{576}$$

Therefore

$$y' = y\left[-\frac{\ln(1+x^2)}{1-x} + \ln(1-x)\frac{2x}{1+x^2}\right] = \left(1+x^2\right)^{\ln(1-x)}\left[-\frac{\ln(1+x^2)}{1-x} + \ln(1-x)\frac{2x}{1+x^2}\right] \tag{577}$$

Example 103. Let $f(x) = x^{2x+7}$. Find the equation of the line tangent to the graph of f at $x = 1$. If it is possible to write the slope as an integer, do so.

We need to find $f'(x)$. To do so we use the logarithmic derivative. If $y = x^{2x+7}$ then

$$\ln y = (2x+7)\ln x \tag{578}$$

Differentiating implicitly with respect to x

$$\frac{d}{dx}\ln y = 2\ln x + \frac{2x+7}{x} \tag{579}$$

by the chain rule

$$\frac{d}{dx}\ln y = \frac{d\ln y}{dy}\frac{dy}{dx} = \frac{1}{y}y' \tag{580}$$

so we have

$$\frac{1}{y}y' = 2\ln x + \frac{2x+7}{x} \tag{581}$$

when $x = 1$ we have that $y = 1^{2+7} = 1$ so the value of $y'(1)$ is

$$y' = 2\ln 1 + \frac{2+7}{1} = 9 \tag{582}$$

Therefore, the slope is 9 and the equation for the tangent line is $y = 9x + b$. Since it must go through $(1,1)$ $b = -8$ so the equation for the tangent line is $y = 9x - 8$.

Exponential Models

The exponential function appears in many different contexts, which may be somewhat surprising. From a mathematical point of view, one reason for the usefulness of the exponential function is that it is its own derivative and in many mathematical models we want the rate of change of a quantity to be proportional to the quantity itself which leads to exponential functions.

The first example is the compound interest formula which we discussed when we defined the number e. We start with an initial amount of money P (the so called principal) at a nominal interest rate per year r (for example, if the interest rate is 8% per year then the nominal rate is $r = 0.08$). If the interest is going to be distributed in m conversion periods per year (for example, $m = 12$ means that the interest will be divided monthly) and t is the term (that is, the number of years that have passed) the total money A accumulated at the end of t years is given by

Compounded Interest Formula:

$$A = P\left(1 + \frac{r}{m}\right)^{mt} \tag{583}$$

Example 104. Find the accumulated amount A if the principal P is invested at an interest rate of r per year for t years if $P = 12000$, $r = 0.08$, $t = 10$, compounded quarterly

We use the formula 583 with $m = 4$

$$A = 12000\left(1 + \frac{0.08}{4}\right)^{4\cdot 10} \simeq 26496.48 \tag{584}$$

Since the interest rate is being divided into m periods, we end up making more money than we would have if the interest rate were applied at the end of the year. Therefore, sometimes we use the effective rate of interest instead of the nominal rate of interest. The effective rate of interest r_{eff} is the interest that we would need to make the same money in case the bank applies the interest annually (that is, with $m = 1$). Therefore, the effective interest rate satisfies

$$P\left(1 + r_{eff}\right)^t = P\left(1 + \frac{r}{m}\right)^{mt} \tag{585}$$

canceling the P and taking natural log on both sides we have

$$t\ln\left(1 + r_{eff}\right) = mt\ln\left(1 + \frac{r}{m}\right) \tag{586}$$

canceling the t's on both sides we have

$$\ln\left(1 + r_{eff}\right) = m\ln\left(1 + \frac{r}{m}\right) = \ln\left(1 + \frac{r}{m}\right)^m \tag{587}$$

and equating arguments we have

$$1 + r_{eff} = \left(1 + \frac{r}{m}\right)^m \tag{588}$$

which implies that the effective rate is

Effective rate interest:

$$r_{eff} = \left(1 + \frac{r}{m}\right)^m - 1 \tag{589}$$

Example 105. Find the effective rate corresponding to the given nominal rate a) 10%/year compounded semiannually, b) 9%/year compounded quarterly

a) We use 589 with $r = 0.1$, $m = 2$

$$r_{eff} = \left(1 + \frac{0.1}{2}\right)^2 - 1 \simeq 0.1025 \tag{590}$$

b) We use 589 with $r = 0.09$, $m = 4$

$$r_{eff} = \left(1 + \frac{0.09}{4}\right)^4 - 1 \simeq 0.093 \tag{591}$$

Now, if we let $m \longrightarrow \infty$ then 583 becomes the compounded continuous interest rate

Continuous Compound Interest Formula:

$$A = Pe^{rt} \tag{592}$$

Example 106. Ten years ago, Joe invested \$20000 into an account with continuously compounding interest. The account is now worth \$22000. a) What is the interest rate on the account? b) Ten years ago, Joe's brother Raul also invested \$20000 into an account with interest compounded monthly. By a miraculous coincidence, Raul's account is now also worth \$22000. At what interest rate did Raul invest his money?

a) We use formula 592 with $P = 20000$, $A = 22000$ and $t = 10$

$$22000 = 20000e^{10r} \tag{593}$$

taking natural log on both sides of the equation we have

$$\ln\left(\frac{11}{10}\right) = 10r \tag{594}$$

so

$$r = \frac{1}{10}\ln\left(\frac{11}{10}\right) \simeq 0.0095301 \simeq 0.95301\% \tag{595}$$

b) We use 583 with $P = 20000$, $A = 22000$, $t = 10$ and $m = 12$

$$22000 = 20000\left(1 + \frac{r}{12}\right)^{120} \tag{596}$$

taking natural log on both sides of the equation we have

$$\ln\left(\frac{11}{10}\right) = 120\ln\left(1 + \frac{r}{12}\right) \tag{597}$$

this is the same as

$$\ln\left(1 + \frac{r}{12}\right) = \frac{\ln\left(\frac{11}{10}\right)}{120} \tag{598}$$

and "exponentiating" both sides we have

$$1 + \frac{r}{12} = e^{\frac{\ln\left(\frac{11}{10}\right)}{120}} \tag{599}$$

so

$$r = 12\left(e^{\frac{\ln\left(\frac{11}{10}\right)}{120}} - 1\right) \simeq 0.0095348 \simeq 0.95348\% \tag{600}$$

Example 107. How long will it take for an investment to triple if it earns interest at a nominal interest rate of 7% per year, compounded monthly?

We use 583 with $A = 3P$, $r = 0.07$, $m = 12$

$$3P = P\left(1 + \frac{0.07}{12}\right)^{12t} \tag{601}$$

taking natural log on both sides we have

$$\ln 3 = 12t \ln\left(1 + \frac{0.07}{12}\right) \tag{602}$$

so

$$t = \frac{\ln 3}{12\ln\left(1 + \frac{0.07}{12}\right)} \simeq 15.74 \tag{603}$$

Many simple models involve solving equations of the form

$$\frac{dQ}{dt} = kQ \tag{604}$$

where k is a constant and $Q(t)$ is the quantity that we are studying. For example, $Q(t)$ might be the size of a population, the temperature of a room, the decay of a radioactive material, learning curves and the money earned at a continuous interest rate. Basically, 604 that the rate of change of the quantity Q at time t is determined by the amount $Q(t)$ at that instant. The functions Q that satisfy 604 are of the form

$$Q(t) = Q_0 e^{kt} \tag{605}$$

where Q_0 is the quantity at time 0.

The following examples illustrate how these models behave.

Example 108. A cup of water is heated to brew a cup of white tea. When the water in the cup reaches 168 degrees, a tea bag is added to the cup, and the cup is placed on a table in a 68 degree room. The temperature of the tea is modeled by the equation $T(t) = Ae^{-kt} + 68$, where T represents temperature, t represents time in minutes since the tea was added to the cup and k, A are positive constants. a) Five minutes after the tea bag is added to the cup, the tea is 148 degrees. Determine A and k. b) How fast is the temperature of the tea changing five minutes after the tea bag is added to the cup?

a) Because t is measured after the tea was added to the cup, $t = 0$ represents the moment when the temperature is 168 degrees, so $T(0) = 168$. This gives

$$168 = A + 68 \tag{606}$$

or

$$A = 100 \tag{607}$$

Also, we have $T(5) = 148$ which is the same as

$$148 = Ae^{-5k} + 68 = 100e^{-5k} + 68 \tag{608}$$

which gives after taking natural log

$$\ln\left(\frac{4}{5}\right) = -5k \tag{609}$$

so

$$k \simeq 0.044 \tag{610}$$

Example 109. The growth rate of the bacterium Escherichia coli, a common bacterium found in the human intestine, is proportional to its size. Under ideal laboratory conditions, when this bacterium is grown in a nutrient broth medium, the number of cells in a culture doubles approximately every 20 min. a) If the initial cell population is 100, determine the growth of the number of cells of this bacterium as a function of time t (in minutes). b) How long will it take for a colony of 100 cells to increase to a population of 1 million? c) If the initial cell population were 1000, how would this alter our model?

a) We use the solution 605

$$Q(t) = Q_0 e^{kt} \tag{611}$$

the initial population is 100 so $Q(0) = Q_0 = 100$ and so we have

$$Q(t) = 100e^{kt} \tag{612}$$

We have that the population has doubled during the first 20 min, so $Q(20) = 200$,

$$200 = 100e^{20k} \tag{613}$$

taking natural log this gives

$$k = \frac{1}{20}\ln 2 \simeq 0.034 \tag{614}$$

therefore $Q(t) = 100e^{0.034t}$

b) We want to find t such that

$$1000000 = 100e^{0.034t} \tag{615}$$

so taking natural log

$$t = \frac{\ln 10000}{0.034} \simeq 270 \tag{616}$$

c) If the initial population were 1000, the value of k would be the same and the time needed to reach 1 million would be around 203 minutes.

The Derivative as a Related Rate

Up to this point we have seen the geometrical interpretation of the derivative and some of its uses in physics. In fact, these are particular interpretations of the derivative and become special cases of the general interpretation, which is that the derivative gives the rate of change of one quantity with respect to another.

When we think about change we are used to think about change with respect to time (or position). However, there are many kinds of change and they don't have to be with respect to time (or position). For example, the price of electricity can depend (among other factors) on the price of oil. In the best case scenario, if we are able to hold all the other variables constant (the economist's *ceteris paribus*) then we might be able to write the price of electricity e as a function of the price of oil o and we could let $e(o)$ to represent this function.

For example, $o = 1$ might represent that the price for a gallon of oil is \$1 and $e(1) = 0.2$ says that the price for electricity is \$0.2 when the oil is \$1. If the price of oil is changed a little bit, from o to $o + h$, then the price of electricity changes from $e(o)$ to $e(o + h)$ and the average change is

$$\frac{e(o+h)-e(o)}{o+h-o} = \frac{e(o+h)-e(o)}{h} \tag{617}$$

We have seen this story before, as $h \longrightarrow 0$ the average change becomes an instantaneous change so $\frac{de}{do}$ represents the rate of change of the price of electricity with respect to the price of oil.

If a quantity y is a function of a quantity x, $y = f(x)$ then $\frac{dy}{dx}$ represents the instantaneous rate of change of y with respect to x.

Example 110. The volume of a right-circular cylinder of radius r and height h is $V = \pi r^2 h$. Suppose that the radius and height of the cylinder are changing with respect to time t. a) Find the relationship between $\frac{dV}{dt}$, $\frac{dr}{dt}$ and $\frac{dh}{dt}$. b) At a certain instant of time, the radius and height of the cylinder are 2 and 6 inches and are increasing at the rate of 0.1 and 0.3 inches/second respectively. How fast is the volume of the cylinder increasing?

a) We have that r and h are functions of time so we might write $V(t) = \pi (r(t))^2 h(t)$. By the product rule and the chain rule

$$\frac{dV}{dt} = \pi \left[\left(\frac{d}{dt} (r(t))^2 \right) h(t) + (r(t))^2 \frac{dh}{dt} \right] = \pi \left[2r(t) \frac{dr(t)}{dt} h(t) + (r(t))^2 \frac{dh(t)}{dt} \right] \tag{618}$$

b) For some value of t we have $r(t) = 2$ in, $h(t) = 6$ in, $\frac{dr}{dt} = 0.1$ in/sec and $\frac{dh}{dt} = 0.3$ in/sec. Then

$$\frac{dV}{dt} = \pi \left[2 \cdot 2 \cdot 0.1 \cdot 6 \frac{\text{in}^3}{\text{sec}} + 2^2 \cdot 0.3 \frac{\text{in}^3}{\text{sec}} \right] = 3.6\pi \frac{\text{in}^3}{\text{sec}} \tag{619}$$

Example 111. Suppose the quantity demanded weekly of the Super Titan radial tires is related to its unit price by the equation $p + x^2 = 144$ where p is measured in dollars and x is measured in units of a thousand. How fast is the quantity demanded changing when $x = 9$, $p = 63$ and the price/tire is increasing at the rate of \$2/week?

From $p + x^2 = 144$ differentiating with respect to t we have

$$\frac{dp}{dt} + 2x \frac{dx}{dt} = 0 \tag{620}$$

We have $\frac{dp}{dt} = 2$, $x = 9$ so

$$2 + 18 \frac{dx}{dt} = 0 \tag{621}$$

so

$$\frac{dx}{dt} = -\frac{1}{9} \tag{622}$$

The negative sign implies that the quantity being demanded is decreasing.

Example 112. In calm waters, oil spilling from the ruptured hull of a grounded tanker spreads in all directions. If the area polluted is a circle and its radius is increasing at a rate of 2ft/sec, determine how fast the area is increasing when the radius of the circle is 40 ft.

The area of a circle is $A = \pi r^2$. Differentiating with respect to time

$$\frac{dA}{dt} = 2\pi r \frac{dr}{dt} \tag{623}$$

we have $r = 40$ ft and $\frac{dr}{dt} = 2 \frac{\text{ft}}{\text{sec}}$ so

$$\frac{dA}{dt} = 160\pi \frac{\text{ft}}{\text{sec}^2} \tag{624}$$

Example 113. A lamppost 10 feet tall stands on a walkway that is perpendicular to a wall. The distance from the post to the wall is 15 feet. A 6 foot man moves on the walkway toward the wall at the rate if 5 feet per second. When he is 5 feet from the wall, how fast is the shadow of his head moving up the wall?

In the figure the lamppost is the red segment, the person is the green segment, the wall is the orange-brown segment. s represents the distance between the person and the wall at a particular instant and h is the height of the shadow at that same instant. l is the distance between the base of the wall and the intersection with the ground of the line from the lamp to the man's head.

The position of the person x with respect to the origin is

$$x = 15 - s \tag{625}$$

Since he is walking at 5 feet per second we have that

$$\frac{dx}{dt} = 5 \tag{626}$$

We want to find $\frac{dh}{dt}$ when $s = 5$, that is, when $x = 10$. By basic trigonometry, we have three right triangles that share the angle θ so

$$\tan\theta = \frac{10}{15+l} = \frac{6}{s+l} = \frac{h}{l} \tag{627}$$

From 627 we have the system of equations

$$\begin{cases} \frac{10}{15+l} = \frac{6}{s+l} \\ \frac{6}{s+l} = \frac{h}{l} \end{cases} \tag{628}$$

To find h we use the second equation of 628 to write

$$h = \frac{6l}{s+l} \tag{629}$$

We want h to be a function of just one variable so we use the first equation of 628 to write l as a function of s. Observe that the first equation is equivalent to

$$10s + 10l = 90 + 6l \tag{630}$$

or

$$l = \frac{90-10s}{4} = \frac{45-5s}{2} \tag{631}$$

Substituting in 629 we get

$$h = \frac{6\left(\frac{45-5s}{2}\right)}{s + \frac{45-5s}{2}} = \frac{270-30s}{45-3s} = \frac{90-10s}{15-s} \tag{632}$$

Given that $x = 15 - s$ we can write $s = 15 - x$ so h as a function of x is

$$h(x) = \frac{90-10\,(15-x)}{15-(15-x)} = \frac{10x-60}{x} = 10 - \frac{60}{x} \tag{633}$$

Using that $\frac{dx}{dt} = 5$ we have by the chain rule that

$$\frac{dh}{dt} = \frac{dh}{dx}\frac{dx}{dt} = 5\frac{dh}{dx} = 5\left(\frac{60}{x^2}\right) = \frac{300}{x^2} \tag{634}$$

Therefore, when $x = 10$ we have $\frac{dh}{dt} = 3$ so the shadow is moving up the wall at the rate of 3 feet per second.

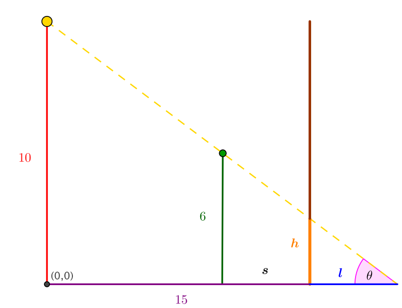


Figure 108: Speed shadow

The First Derivative Test

Recall that a function is increasing on an interval (a, b) when $x_1 < x_2$ implies that $f(x_1) < f(x_2)$, that is, as we move from left to right along the x axis the values of $f(x)$ increase. Likewise, a function is decreasing on an interval (a, b) when $x_1 < x_2$ implies that $f(x_1) > f(x_2)$, that is, as we move from left to right along the x axis the values of $f(x)$ decrease.

For example, as we will see in a moment, the function $f(x) = (x^2 - 1)(x - 2)$ is increasing on the intervals $\left(-\infty, \frac{1}{3}\left(2 - \sqrt{7}\right)\right)$, $\left(\frac{1}{3}\left(2 + \sqrt{7}\right), \infty\right)$ and decreasing on the interval $\left(\frac{1}{3}\left(2 - \sqrt{7}\right), \frac{1}{3}\left(2 + \sqrt{7}\right)\right)$.

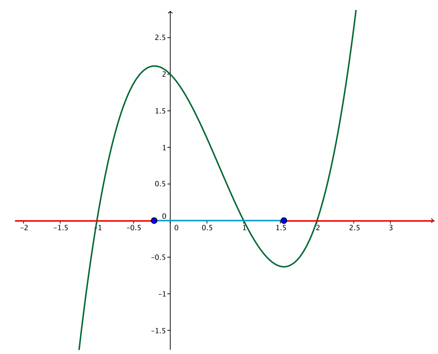


Figure 109: Function $y = (x^2 - 1)(x - 2)$

If we think of the tangent lines to the points on the curve, we can see that on the intervals in which the function is increasing the slopes of the tangent lines are positive (hence the first derivatives are positive) and on the interval in which the function is decreasing the slopes of the tangent lines are negative (hence the first derivatives are negative).

Therefore, it seems that there is a close relationship between being an increasing (decreasing) function on an interval and having a positive (negative) derivative on that interval. Also, from the previous example, the behavior of the function changed at the two blue points indicated in the figure and as we will see with a simple calculation the tangent lines are those two points have zero slope. The following result tells us that the sign of the derivative determines if a function is increasing or decreasing.

Suppose that f is differentiable at a point p:

1. If $f'(p) > 0$ then f is increasing in some neighborhood of p
2. If $f'(p) < 0$ then f is decreasing in some neighborhood of p

Proof. We will show just 1. since 2. is entirely analogous. Recall that the limit definition of the derivative is [43]

$$f'(p) = \lim_{x \longrightarrow p} \frac{f(x) - f(p)}{x - p} \tag{635}$$

which we can write as

$$0 = \lim_{x \longrightarrow p} \left(\frac{f(x) - f(p)}{x - p} - f'(p) \right) \tag{636}$$

[43] This follows from making the change of variables $h = x - p$ or $x = h + p$ in the limit definition $f'(p) = \lim_{h \longrightarrow 0} \frac{f(p+h) - f(p)}{h}$

In terms of the epsilon delta definition this means that for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, p, f) > 0$ such that if $0 < |x-p| < \delta$ then $\left|\frac{f(x)-f(p)}{x-p} - f'(p)\right| < \varepsilon$. This is the same as saying that if x belongs to the neighborhood $(p-\delta, p+\delta)$ then $-\varepsilon < \frac{f(x)-f(p)}{x-p} - f'(p) < \varepsilon$. The last inequality can be written as

$$-\varepsilon + f'(p) < \frac{f(x)-f(p)}{x-p} < \varepsilon + f'(p) \tag{637}$$

If we apply the definition of the limit for $\varepsilon = f'(p)$ then there exists some neighborhood $(p-\delta, p+\delta)$ for which 637 holds, which now becomes

$$0 < \frac{f(x)-f(p)}{x-p} < 2f'(p) \tag{638}$$

If $x > p$ we can multiply both sides of the first inequality by $x-p$ to obtain $0 < f(x) - f(p)$ or $f(p) < f(x)$. If $x < p$ then multiplying both sides of the first inequality by $x-p$ to obtain $0 > f(x) - f(p)$ or $f(p) > f(x)$. But we have just shown that on $(p-\delta, p+\delta)$, $f(x)$ is increasing $\square$

Now, what if $f'(p) = 0$? Well, basically anything can happen. For example, if $p = 0$ and $f(x) = x^3$ then the function is actually increasing on all the real line $\mathbb{R}$ so in particular it is going to be increasing in any neighborhood of p. If $f(x) = -x^3$ then the function is actually decreasing on all the real line $\mathbb{R}$ so in particular it is going to be decreasing in any neighborhood of $p = 0$. If $f(x) = 0$ then the function is always constant so it won't be decreasing nor increasing on any neighborhood around 0.

However, if we look at the linear approximation $\triangle y \simeq f'(x)\triangle x$ then around $x = p$ we get $\triangle y \simeq 0$ so this means that near p the function is almost constant, that is, any changes in its values happen because of higher order terms like $(\triangle x)^2$, not because of the linear term $\triangle x$. Therefore, it is usual to refer to this point as an stationary point.

First Derivative Test, Part 1:
Suppose that $y = f(x)$ is a differentiable function.

- A point x is called a **stationary point** of $f(x)$ if $f'(x) = 0$
- If $f'(x)$ is positive on an interval then $f(x)$ is increasing on that interval
- If $f'(x)$ is negative on an interval then $f(x)$ is decreasing on that interval

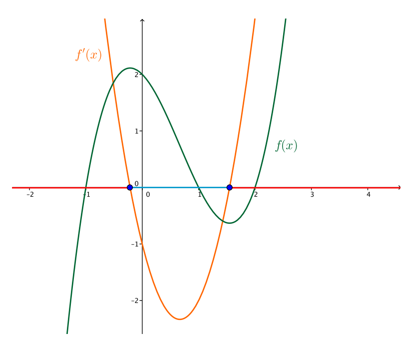


Figure 110: Graph of $f(x)$ and $f'(x)$

Example 114. Determine where $f(x) = \left(x^2-1\right)(x-2)$ is increasing, decreasing, and the stationary points of $f(x)$

We need to analyze the sign of $f'(x)$. By the product rule

$$f'(x) = \left(x^2-1\right)'(x-2) + \left(x^2-1\right)(x-2)' = 2x(x-2) + x^2 - 1 = 3x^2 - 4x - 1 \tag{639}$$

To know when does $f'(x) > 0$ and $f'(x) < 0$ we need to find the roots of $3x^2 - 4x - 1$ to factorize it. Taking $a = 3$, $b = -4$, $c = -1$ we use the formula for the roots of a quadratic polynomial

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{4 \pm \sqrt{16+12}}{6} = \frac{2 \pm \sqrt{7}}{3} \tag{640}$$

so the roots are $r_1 = \frac{2+\sqrt{7}}{3}$ and $r_2 = \frac{2-\sqrt{7}}{3}$. This implies that

$$f'(x) = 3x^2 - 4x - 1 = 3(x-r_1)(x-r_2) = 3\left(x - \frac{2+\sqrt{7}}{3}\right)\left(x - \frac{2-\sqrt{7}}{3}\right) \tag{641}$$

The stationary points of $f(x)$ are the points for which $f'(x) = 0$ so $x = \frac{2+\sqrt{7}}{3}$ and $x = \frac{2-\sqrt{7}}{3}$ are the stationary points. To determine the sign of $f'(x)$ we make a table with the signs of the factors of the derivative:

	$\left(-\infty, \frac{2-\sqrt{7}}{3}\right)$	$\left(\frac{2-\sqrt{7}}{3}, \frac{2+\sqrt{7}}{3}\right)$	$\left(\frac{2+\sqrt{7}}{3}, \infty\right)$
$x - \frac{2+\sqrt{7}}{3}$	$-$	$-$	$+$
$x - \frac{2-\sqrt{7}}{3}$	$-$	$+$	$+$
$f'(x) = 3\left(x - \frac{2+\sqrt{7}}{3}\right)\left(x - \frac{2-\sqrt{7}}{3}\right)$	$+$	$-$	$+$
Behavior $f(x)$	$\nearrow$	$\searrow$	$\nearrow$

Here $\nearrow$ represents that the function is increasing on that interval and $\searrow$ represents that the function is decreasing on that interval. Therefore, $f(x) = \left(x^2-1\right)(x-2)$ is increasing on the intervals $\left(-\infty, \frac{1}{3}\left(2-\sqrt{7}\right)\right)$, $\left(\frac{1}{3}\left(2+\sqrt{7}\right), \infty\right)$ and decreasing on the interval $\left(\frac{1}{3}\left(2-\sqrt{7}\right), \frac{1}{3}\left(2+\sqrt{7}\right)\right)$. This is the same as what we said before.

Example 115. Determine the intervals in which $g(x) = x^3 + 3x^2 + 1$ is increasing and decreasing

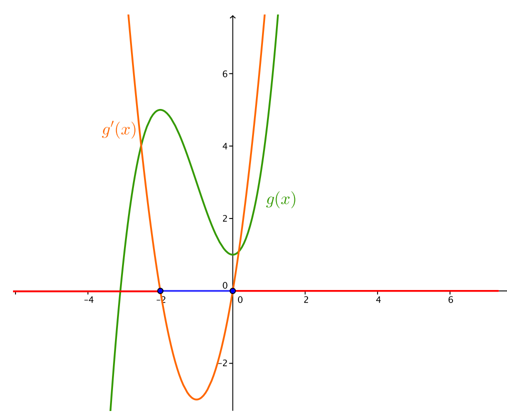


Figure 111: Graph of $g(x)$ and $g'(x)$

We differentiate first to use the first derivative test

$$g'(x) = 3x^2 + 6x = 3x(x+2) \tag{642}$$

To determine the signs of $g'(x)$ it is always useful to find first the stationary points. We can see that they are $x = 0$ and $x = -2$. We make the table of signs

	$(-\infty, -2)$	$(-2, 0)$	$(0, \infty)$
x	$-$	$-$	$+$
$x+2$	$-$	$+$	$+$
$g'(x) = 3x(x+2)$	$+$	$-$	$+$
Behavior $g(x)$	$\nearrow$	$\searrow$	$\nearrow$

By the first derivative test, $g(x)$ is increasing on the intervals $(-\infty, -2)$, $(0, \infty)$ and decreasing on the interval $(-2, 0)$

Example 116. Determine the intervals in which $h(x) = \frac{1}{2x+3}$ is increasing and decreasing

In this case

$$h'(x) = -\frac{2}{(2x+3)^2} \tag{643}$$

and because $(2x+3)^2$ is positive we have that $h'(x)$ is always negative so $h(x)$ is decreasing. The only thing to observe is that because $h(x)$ and $h'(x)$ are not defined at $x = -\frac{3}{2}$ then $h(x)$ is decreasing on the intervals $\left(-\infty, -\frac{3}{2}\right)$ and $\left(-\frac{3}{2}, \infty\right)$ separately, and it is not decreasing if both intervals are considered at the same time.

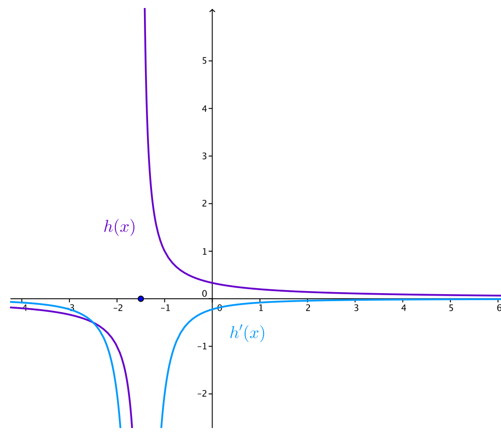


Figure 112: Graph of $h(x)$ and $h'(x)$

Example 117. Show that the cubic function $f(x) = ax^3 + bx^2 + cx + d$ $(a \neq 0)$ has no stationary points if and only if $b^2 - 3ac < 0$.

We have that $f'(x) = 3ax^2 + 2bx + c$. The roots of the polynomial $f'(x)$ are

$$r = \frac{-2b \pm \sqrt{(2b)^2 - 4(3a)(c)}}{6a} = \frac{-b \pm \sqrt{b^2 - 3ac}}{3a} \tag{644}$$

Therefore, if $b^2 - 3ac < 0$ then $\sqrt{b^2 - 3ac}$ is undefined and no roots can exist, therefore, no critical points.

Up to this point we have used the first derivative to determine the intervals on which a function is decreasing and increasing. Moreover, if we look at the graphs of the previous examples, we can see that the behavior of the function changes at the stationary points of the function. Also, for the examples we have seen so far, at the stationary points the function seems to achieve the highest and smallest value, at least with respect to an interval near the stationary point. Those points are called relative maximum and relative minimum for the function.

Suppose $y = f(x)$ is a function.

- A point x is called a relative maximum if the function achieves its highest value at x with respect to the points in some neighborhood of x.
- A point x is called a relative minimum if the function achieves its lowest value at x with respect to the points in some neighborhood of x.
- A point x is a relative extrema if it is a relative maximum or a relative minimum.

The examples so far indicate that the relative extrema occur at the stationary points. When the function is differentiable, this is in general true. If the function is not differentiable everywhere, for example, if the function is $f(x) = |x|$ then $x = 0$ is a relative minimum [44] but there is no derivative at $x = 0$. This means that we should enlarge our concept of stationary point.

[44] Actually an absolute minimum

A point p is a **critical point** if

- p belongs to the domain of f
- $f'(p) = 0$ or $f'(p)$ is undefined

The concept of critical point allows us to state the second part of the first derivative test

First Derivative Test, Part 2: The relative extrema of $f(x)$ are all critical points

Proof. Suppose that p is a relative extrema of $f(x)$. Then either $f'(p)$ does not exist or it does exist. It may not exists like in the example of the absolute value. If it exists it is either positive, negative or zero. But the first two possibilities implies that f is increasing or decreasing in some neighborhood of p, contradicting the fact that p is a relative extrema. Therefore, the only possibility is for $f'(p) = 0$. In any case, this means that p has to be a critical point. □

However, not every critical point is a relative extrema. For example, the function $y = x^3$ has a critical point at $x = 0$ but 0 is neither a relative maximum nor a relative minimum.

Now we would like a criterion to guarantee when our critical point is a relative maximum or relative minimum. This is what the third part of the First Derivative Test addresses. The proof this is part will have to wait until we have the Mean Value Theorem, but the statement will hopefully seem pretty intuitive.

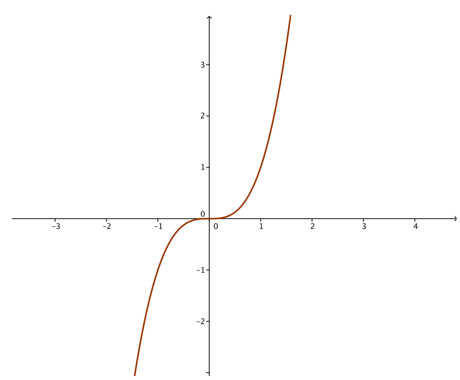


Figure 113: Critical point which is not a relative extrema

First Derivative Test, Part 3:

- Suppose f is differentiable in a deleted neighborhood of a critical point p, and suppose the derivative f' has one sign to the left of p and the opposite sign to the right of p
- If the sign of f' changes from negative to positive then p is a relative minimum.

$$\searrow \quad \mathbf{minimum} \quad \nearrow \tag{645}$$

- If the sign of f' changes from positive to negative then p is a relative maximum

$$\nearrow \quad \mathbf{maximum} \quad \searrow \tag{646}$$

The Second Derivative Test

Besides the First Derivative Test, it is also possible to classify the relative extrema with the second derivative. We said before that if we "zoom in" the graph of a function, the graph looks very similar to its tangent line. However, that statement is actually true when the point we are zooming in is not a stationary point, that is, when the derivative is not zero.

For stationary points, the graph of the function looks more like a parabola (or two parabolas pieced together in the case of 0 for x^3) so we should study how to classify the maxima and minima for parabolas first.

The most general parabola is of the form $y = ax^2 + bx + c$. The derivative is $y' = 2ax + b$ so there is only one stationary point which happens at $x = -\frac{b}{2a}$. Let's call this point

$$x_c = -\frac{b}{2a} \tag{647}$$

We want to determine if x_c is a relative maximum or a relative minimum. For this we need to compare $y(x)$ with $y(x_c)$ where x is close to x_c. Since it is close we can write $x = x_c + \triangle x$ where $\triangle x$ is a small quantity. Therefore

$$\begin{aligned} y(x) - y(x_c) &= y\left(-\tfrac{b}{2a} + \triangle x\right) - y\left(-\tfrac{b}{2a}\right) \\ &= a\left(-\tfrac{b}{2a} + \triangle x\right)^2 + b\left(-\tfrac{b}{2a} + \triangle x\right) + c - a\left(-\tfrac{b}{2a}\right)^2 - b\left(-\tfrac{b}{2a}\right) - c \\ &= a\left(\tfrac{b^2}{4a^2} - \tfrac{b\triangle x}{a} + (\triangle x)^2\right) - \tfrac{b^2}{2a} + b\triangle x - a\left(\tfrac{b^2}{4a^2}\right) + \tfrac{b^2}{2a} \\ &= a(\triangle x)^2 \end{aligned} \tag{648}$$

We can see that the *sign* of the difference $y(x) - y(x_c)$ is determined completely by the *sign* of a. For example, if a is positive then $y(x) - y(x_c) = a(\triangle x)^2 > 0$ so $y(x) > y(x_c)$ which implies that x_c is a relative minimum. Similarly, if a is negative we have $y(x_c) > y(x)$ so x_c is a relative maximum [45]. Now, a parabola for which $a > 0$ was called concave up and a parabola for which $a < 0$ was called concave down so we have found the following

[45] Actually, this calculation shows that x_c is an absolute maximum or absolute minimum since we did not assume anything about the size of $\triangle x$

If $y = ax^2 + bx + c$ is a parabola then:

- The only critical point is $x_c = -\frac{b}{2a}$
- If the parabola is concave up $(a > 0)$ then x_c is a relative minimum
- If the parabola is concave down $(a < 0)$ then x_c is a relative maximum

Now the question becomes: is there a way to characterize a in terms of the derivatives of y? We see that $y' = 2ax + b$ involves both a and b so this can't be our candidate to specify a. However, $y'' = 2a$ so up to a factor of 2 it uniquely characterizes a. Inspired in this fact we define the following

If $y = f(x)$ is a function

- The function is called concave up at the point x if $y''(x) > 0$
- The function is called concave down at the point x if $y''(x) < 0$

Finally, the following results gives us an answer to our question on how to determine when a critical point is a relative maximum or a relative minimum.

Second Derivative Test:
Suppose that x_c is a stationary point of the function $y = f(x)$, that is, $f'(x_c) = 0$. Then

- x_c is a relative minimum if the function is concave up at x_c, that is, $f''(x_c) > 0$,

$$\begin{array}{ccc} + & & + \\ & \cup & \end{array} \tag{649}$$

- x_c is a relative maximum if the function is concave down at x_c, that is, $f''(x_c) < 0$

$$\begin{array}{ccc} - & & - \\ & \cap & \end{array} \tag{650}$$

Proof. We will show the first case since the second one is entirely analogous. Since the second derivative $f''(p)$ exists, looking at the limit definition it is obvious that f' exists on a neighborhood of p. Call $g(x) = f'(x)$. Then $g(x_c) = 0$ and $g'(x_c) > 0$, that is, g is increasing in a neighborhood of x_c. In particular, since g is zero at x_c we must have on that neighborhood that g is negative to the left of x_c and g is positive to the right of x_c. Since $g = f'$ then f' is decreasing to the left of x_c and f' is increasing to the left of x_c. The First Derivative Test then shows that f must have a relative minimum at x_c, which is what we wanted to show. □

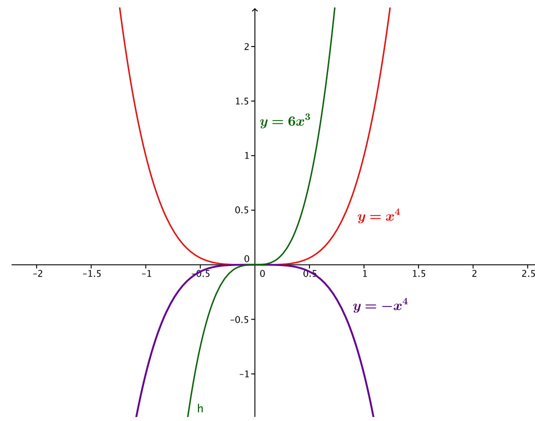


Table 11: Critical points x_c with $f''(x_c) = 0$

Now, what happens if $f''(x_c) = 0$? In that case x_c can be either a relative maximum, a relative minimum or neither. For example, if $y = x^4$ then $x_c = 0$ is a relative minimum; if $y = -x^4$ then $x_c = 0$ is a relative maximum and if $y = x^3$ then $x_c = 0$ is not a relative extrema. However, it is possible, to give some classification of the points satisfying $f''(p) = 0$, regardless of whether they are critical points or not, but in order to do this we need a more geometrical description of the second derivative of a function.

Suppose that f is differentiable at some point p. Then it is easy to check that the equation of the tangent line going through point $(p, f(p))$ is $y = T_p(x) = f'(p)(x-p) + f(p)$. In geometrical terms, being concave up or down translates into having the function above the tangent line or below the tangent line respectively. [46]

[46] Strictly speaking, this characterization is the actual definition of concavity. However, for practical purposes we looks first at the second derivative which is why we started mentioning that criteria.

Let f be differentiable at p and let $T_p(x) = f'(p)(x-p) + f(p)$ be the equation of the tangent line.

- Suppose that $f''(p) > 0$, that is, f is concave up at p. Then f' is increasing in a neighborhood of p and in a neighborhood of p we have $f(x) \geq T(x)$, that is, $y = f(x)$ lies above its tangent at p
- Suppose that $f''(p) < 0$, that is, f is concave down at p. Then f' is decreasing in a neighborhood of p and in a neighborhood of p we have $f(x) \leq T(x)$, that is, $y = f(x)$ lies below its tangent at p

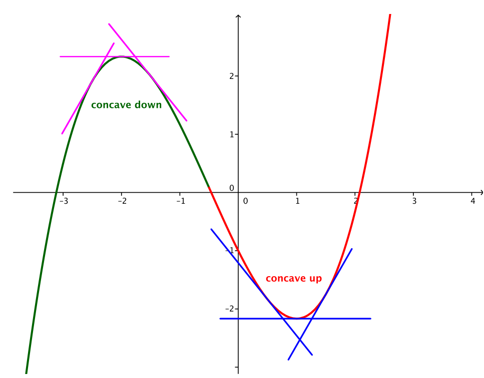


Figure 114: Geometrical Interpretation of Concavity

Now we can describe in better terms, what happens at the points for which $f''(p) = 0$. If our function is for example $f(x) = x^3$ then the concavity may change at $p = 0$. However, if the function is for example $f(x) = x^4$ then there is no concavity change at $p = 0$. The first type of points are called inflection points. [47]

[47] Strictly speaking, the inflection point is the point on the graph of the curve, that is, one must specify both the x coordinate and the y coordinate of the inflection point. However, once we know the x coordinate the y coordinate is simply $f(x)$ so not much is gained.

Example 118. Determine where $g(x) = x^3 - x$ is increasing, decreasing, concave up, concave down. Determine the inflection points, the relative maxima and minima.

First we find the critical points.

$$g'(x) = 3x^2 - 1 = \left(\sqrt{3}x\right)^2 - 1^2 = \left(\sqrt{3}x - 1\right)\left(\sqrt{3}x + 1\right) \qquad (651)$$

so the critical points are $x_1 = \frac{1}{\sqrt{3}}$ and $x_2 = -\frac{1}{\sqrt{3}}$. Now we make the table to determine where $g(x)$ it is increasing and decreasing.

	$\left(-\infty,-\frac{1}{\sqrt{3}}\right)$	$\left(-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$	$\left(\frac{1}{\sqrt{3}},\infty\right)$
$\sqrt{3}x-1$	$-$	$-$	$+$
$\sqrt{3}x+1$	$-$	$+$	$+$
$g'(x)=3x^2-1$	$+$	$-$	$+$
Behavior $g(x)$	$\nearrow$	$\searrow$	$\nearrow$

Therefore, $g(x)$ is increasing on $\left(-\infty,-\frac{1}{\sqrt{3}}\right)$, $\left(\frac{1}{\sqrt{3}},\infty\right)$ and decreasing on $\left(-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$.

To determine the concavity of $g(x)$ we find $g''(x)$.

$$g''(x)=\left(3x^2-1\right)'=6x \tag{652}$$

Since $g''\left(x_1\right)=6\left(\frac{1}{\sqrt{3}}\right)>0$ we must have that $x_1=\frac{1}{\sqrt{3}}$ is a relative minimum. Since $g''\left(x_2\right)=6\left(-\frac{1}{\sqrt{3}}\right)<0$ we must have that $x_2=-\frac{1}{\sqrt{3}}$ is a relative maximum. Moreover, since $g''(x)>0$ for $x>0$ then $g(x)$ is concave up on $(0,\infty)$. Since $g''(x)<0$ for $x<0$ then $g(x)$ is concave down on $(-\infty,0)$.

	$(-\infty,0)$	$(0,\infty)$
$g''(x)=6x$	$-$	$+$
Behavior $g(x)$	concave down	concave up

We can also see from the last table that $x=0$ is an inflection point for $f(x)$.

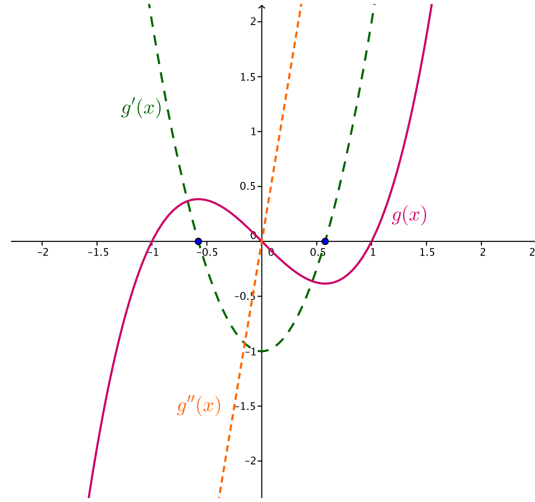


Figure 115: Graph of $g(x)$, $g'(x)$, $g''(x)$

Example 119. Determine where $h(t)=\frac{t^2}{t-1}$ is increasing, decreasing, concave up, concave down. Determine the inflection points, the relative maxima and minima.

We have that

$$h'(t)=\frac{2t\left(t-1\right)-t^2}{\left(t-1\right)^2}=\frac{t^2-2t}{\left(t-1\right)^2}=\frac{t\left(t-2\right)}{\left(t-1\right)^2} \tag{653}$$

Therefore, the critical points are $t_1=0$ and $t_2=2$. Since $h(t)$ is undefined at $t=1$ it is best to separate the interval $(0,2)$ into $(0,1)$ and $(1,2)$ for the table of signs

	$(-\infty,0)$	$(0,1)$	$(1,2)$	$(2,\infty)$
t	$-$	$+$	$+$	$+$
$t-2$	$-$	$-$	$-$	$+$
$(t-1)^2$	$+$	$+$	$+$	$+$
$h'(t)=\frac{t(t-2)}{(t-1)^2}$	$+$	$-$	$-$	$+$
Behavior $h(t)$	$\nearrow$	$\searrow$	$\searrow$	$\nearrow$

Therefore, $h(t)$ is increasing on $(-\infty,0)$, $(2,\infty)$ and decreasing on $(0,1)$, $(1,2)$.

To determine the concavity of $h(t)$ we find $h''(t)$.

$$h''(t)=\left(\frac{t^2-2t}{\left(t-1\right)^2}\right)'=\frac{\left(2t-2\right)\left(t-1\right)^2-\left(t^2-2t\right)\left(2\left(t-1\right)\right)}{\left(t-1\right)^4}=\frac{2}{\left(t-1\right)^3} \tag{654}$$

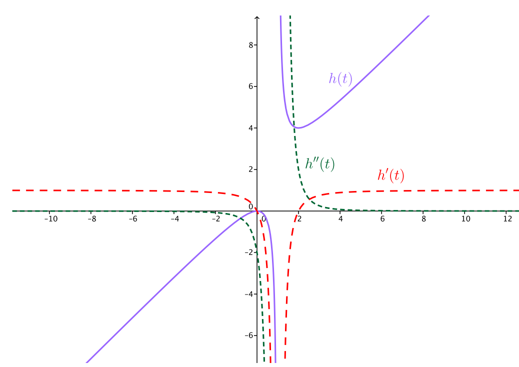


Figure 116: Graph of $h(t)$, $h'(t)$ and $h''(t)$

	$(-\infty,1)$	$(1,\infty)$
$h''(t)=\frac{2}{(t-1)^3}$	$-$	$+$
Behavior $g(x)$	concave down	concave up

If $t > 1$ we have that $h''(t) > 0$ so $h(t)$ is concave up on $(1,\infty)$. Also, because $h''(2) = 2$ we have that $t_2 = 2$ is a relative minimum. If $t < 1$ we have $h''(t) < 0$ so $h(t)$ is concave down on $(-\infty,1)$. Because $h''(0) = -2$ we have that $t_1 = 0$ is a relative maximum.

In this case there are no inflection points.

Example 120. The altitude of a rocket (in feet) t seconds into flight is given by $f(t) = -t^3 + 54t^2 + 480t + 10$ $(t \geq 0)$. a) Find an expression for the rocket's velocity at any time t and an expression for the rocket's acceleration at any time t. b) At what time does the rocket reach the highest point of its flight?

a) The velocity is the derivative of the position (altitude) of the rocket

$$v = \frac{df}{dt} = -3t^2 + 108t + 480 \tag{655}$$

The acceleration is the derivative of the velocity

$$a = \frac{dv}{dt} = -6t + 108 \tag{656}$$

b) The the highest point of the flight is a critical point of the position $f(t)$, so we need to find when does $\frac{df}{dt} = 0$, that is, when is the velocity zero. We need to solve then

$$-3t^2 + 108t + 480 = 0 \tag{657}$$

and the polynomial can be factorized as

$$(-3t + 120)(t + 4) = 0 \tag{658}$$

The possible times (critical points) are $t_1 = 40$ and $t_2 = -4$. Since we only care about positive values of time in this problem the time is $t_1 = 40$. We can see that this value is a relative maximum because $a(40) = f''(40) = -132$.

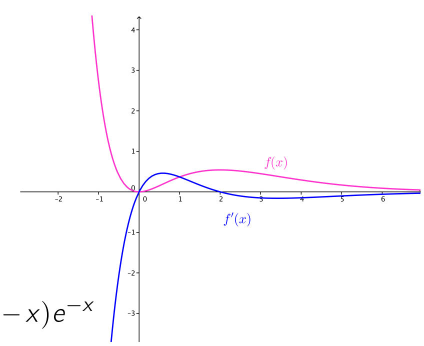


Figure 117: Graph of $y = x^2e^{-x}$ and its derivative

Example 121. Determine the intervals where $f(x) = x^2e^{-x}$ is increasing and where it is decreasing

We need to find the first derivative

$$f'(x) = \left(x^2\right)' e^{-x} + x^2\left(e^{-x}\right)' = 2xe^{-x} + x^2\left(e^{-x}\right)(-1) = e^{-x}\left(2x - x^2\right) = x(2-x)e^{-x} \tag{659}$$

The table of signs is

	$(-\infty,0)$	$(0,2)$	$(2,\infty)$
x	$-$	$+$	$+$
$2-x$	$+$	$+$	$-$
e^{-x}	$+$	$+$	$+$
$f'(x) = x(2-x)e^{-x}$	$-$	$+$	$-$
Behavior $f(x)$	$\searrow$	$\nearrow$	$\searrow$

Example 122. Determine the intervals of concavity of the function $f(x) = \frac{e^x - e^{-x}}{2}$

We have

$$f'(x) = \frac{e^x + e^{-x}}{2} \tag{660}$$

$$f''(x) = \frac{e^x - e^{-x}}{2} = f(x) \tag{661}$$

The concavity is determined by the sign of $f''(x)$. For example, $f''(x)$ is positive when $\frac{e^x - e^{-x}}{2} > 0$, that is, when $e^x > \frac{1}{e^x}$ which is the same as $e^{2x} > 1$. Since $e^0 = 1$ this implies from the graph of e^x that we need $2x > 0$ or $x > 0$. Similarly, $f''(x)$ is negative when $x < 0$.

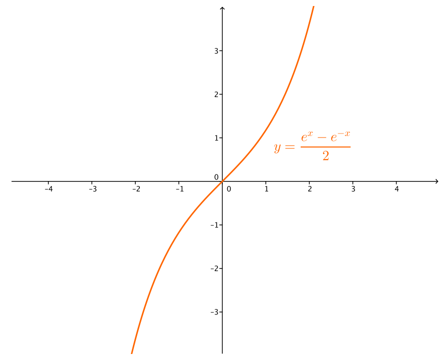


Figure 118: $y = \frac{e^x + e^{-x}}{2}$

Extreme Value Theorem and Optimization Problems

Consider the function $f(x) = x$. If we try to find the critical points of $f(x)$ we calculate $f'(x) = 1$ and since it is never 0 there are no critical points to the function, in fact, it is always increasing since the derivative is positive. However, suppose that we don't want to work with the entire real line $\mathbb{R}$ but instead with restrict the values of the function to a closed interval, for example, we want to study $f(x) = x$ on the interval $[1, 3]$. Then because we restricted the values of x, something new happens: now the function can have a maximum and a minimum!

This should not come as a surprise, because we restricted the values the function can take so it is not "free" to take all the values it wanted to take. In applications this is crucial because sometimes restrictions have to be placed on the values of a variable, for example, there is no point in studying negative values of a population.

Actually, if we restrict our variable to an interval we need to see what happens at the endpoints to get a complete picture of the behavior of the function.

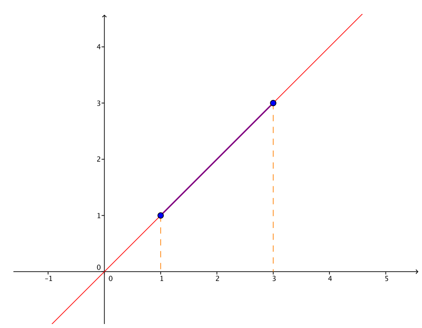


Figure 119: $y = x$ restricted to the interval $[1, 3]$

- **Extreme Value Theorem:** suppose that f is a continuous function on the *closed* interval $[a, b]$. Then f achieves on $[a, b]$ an absolute maximum and an absolute minimum

- **Guide to Optimization Problems:** Suppose that we want to optimize a quantity $y = f(x)$: that is, we want to find the values of x which gives the relative extrema of f or the absolute maxima or absolute minima of f. There are two cases to consider:

1. If the variable x has no constraints that is, it can be any real number or the domain is an *open* interval (or intervals), then the relative extrema are the critical points that satisfy the first or second derivative test.

2. If the variable x satisfy the inequalities $a \leq x \leq b$, then to find the relative extrema we proceed as 1. and to find the absolute maximum and minimum we make a table with all the critical points inside $[a, b]$ together with the endpoints a and b. We evaluate the function at each of the previous points and we compare their values to identify the highest and lowest value, which corresponds to the absolute maximum and absolute minimum respectively.

Example 123. Farmer Frank wants to enclose a square grazing area and a circular corral for his horses. He has 400 meters of fencing, and he will use all of this material to enclose the two areas (although he will settle for just one of the two if it is to his advantage) a) How should Farmer Frank divide the fencing in order to maximize the total area enclosed? b) How would Farmer Frank divide the fencing in order to minimize the total area enclosed?

Let's call l the length of the square and r the radius of the circular corral. Then the area of the square is l^2 and the area of the corral is πr^2. Therefore, the total area enclosed is

$$A = l^2 + \pi r^2 \tag{662}$$

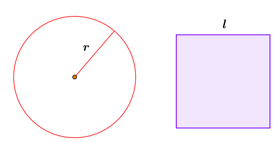


Figure 120: Farmer's Frank Farm

We can't differentiate just yet, because l, r are two different variables so A is a function of two variables which we will not learn how to handle. However, there is a relationship between l and r because we must use 400 meters of fencing. The perimeter of the square is $4l$ and the perimeter of the circle is $2\pi r$ so the condition is

$$4l + 2\pi r = 400 \tag{663}$$

so we can use it to find l as a function of r

$$l = 100 - \frac{\pi r}{2} \tag{664}$$

If we substitute l in the formula for the area 662 we end up with

$$A = \left(100 - \frac{\pi r}{2}\right)^2 + \pi r^2 = 10000 - 100\pi r + \frac{\pi^2 r^2}{4} + \pi r^2 = 10000 - 100\pi r + \left(\frac{\pi^2}{4} + \pi\right) r^2 \quad (665)$$

The critical point of A satisfies

$$\frac{dA}{dr} = -100\pi + 2\left(\frac{\pi^2}{4} + \pi\right) r = 0 \quad (666)$$

so the critical value of the radius is

$$r = \frac{50}{\frac{\pi}{4} + 1} = \frac{200}{\pi + 4} \quad (667)$$

We can see that this value gives a relative minimum because $\frac{d^2A}{dr^2} = 2\left(\frac{\pi^2}{4} + \pi\right) > 0$. For this value of r using 664 the length of the square has to be

$$l = 100 - \frac{\pi}{2}\left(\frac{200}{\pi + 4}\right) = 100\left(1 - \frac{1}{\pi + 4}\right) = 100\left(\frac{\pi + 3}{\pi + 4}\right) \quad (668)$$

Therefore, if $r = \frac{200}{\pi+4}$ and $l = 100\left(\frac{\pi+3}{\pi+4}\right)$ the area fenced is a minimum.

Now, since we found only one critical point it seems that there is only a minimum area enclosed but this does not seem to make a lot of physical sense because that would imply that with only 400 meters of fencing we could make enclose arbitrarily large areas. The problem is that there are physical constraints that need to be taken into account.

For example, it does not make sense to have a negative value for the radius or the length of the square, so we need $l \geq 0$ and $r \geq 0$. From 664 the condition $l \geq 0$ implies

$$100 - \frac{\pi r}{2} \geq 0 \quad (669)$$

so

$$\frac{200}{\pi} \geq r \quad (670)$$

This implies that the value of the radius must satisfy the inequalities $0 \leq r \leq \frac{200}{\pi}$ and therefore we are actually working with a closed interval and we need to apply the previous criteria. Therefore, we need to evaluate $A(r)$ at the endpoints, that is, we need to find $A(0)$ and $A\left(\frac{200}{\pi}\right)$. We have $A(0) = 10000$ and $A\left(\frac{200}{\pi}\right) = \frac{40000}{\pi}$.

We have $A\left(\frac{200}{\pi}\right) > A(0)$ and because there is no other candidate for the maximum this implies that the relative maximum occurs when $r = \frac{200}{\pi}$ and $l = 0$.

Example 124. Suppose you own a plot of land shaped like a right triangle. You have decided to construct a fence from the northern tip to some point on the souther border. A buyer offers you \$3 per square yard for the portion of your plot to the west of the fence and \$2 per square yard for the portion to the east. Fencing costs you \$1 per yard. Where should you build the fence to maximize your profit? (assume that the base has length $\sqrt{2}$ and the height is 1)

Consider the following figure

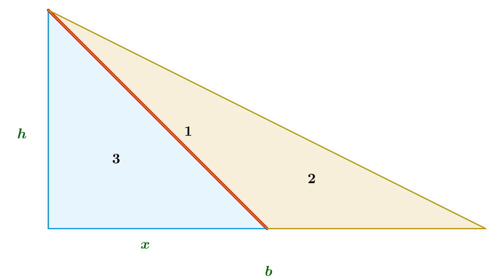


Figure 121: Plot of land-triangle shaped

Here b is the base of the triangle and h is height of the triangle (these are constant for the problem). We call x the position in which the fence touches the base. To find the profit we need first to find the areas of the two regions divided by the fence and the length of the fence

$$\begin{aligned} \text{area left plot } &= \tfrac{bx}{2} & \text{profit left plot} &= \tfrac{3bx}{2} \\ \text{area right plot} &= \tfrac{bh}{2} - \tfrac{bx}{2} & \text{profit right plot} &= 2\left(\tfrac{bh}{2} - \tfrac{bx}{2}\right) \\ \text{length fence} &= \sqrt{h^2 + x^2} & \text{cost fence} &= 1\sqrt{h^2 + x^2} \end{aligned} \tag{671}$$

Therefore, the net profit is

$$P(x) = \frac{3bx}{2} + bh - bx - \sqrt{h^2 + x^2} \tag{672}$$

First we find the critical points

$$P'(x) = \frac{3b}{2} - b - \frac{x}{\sqrt{h^2 + x^2}} = \frac{b}{2} - \frac{x}{\sqrt{h^2 + x^2}} \tag{673}$$

so the critical points satisfy

$$\frac{b}{2} = \frac{x}{\sqrt{h^2 + x^2}} \tag{674}$$

to find the values of x we square both sides of the equations

$$b^2h^2 + b^2x^2 = 4x^2 \tag{675}$$

Therefore

$$x = \frac{bh}{\sqrt{4 - b^2}} \tag{676}$$

We use the fact that $b = \sqrt{2}$ and $h = 1$ so

$$x = \frac{\sqrt{2}}{\sqrt{4 - 2}} = 1 \tag{677}$$

Now we need to see if $x = 1$ corresponds to a relative maximum or minimum. First of all using the values of b and h

$$P'(x) = \frac{\sqrt{2}}{2} - \frac{x}{\sqrt{1 + x^2}} \tag{678}$$

The second derivative is

$$P''(x) = -\frac{\sqrt{1 + x^2} - x\frac{x}{\sqrt{1+x^2}}}{1 + x^2} \tag{679}$$

when $x = 1$ then

$$P''(1) = -\frac{\sqrt{2} - \frac{1}{\sqrt{2}}}{2} = -\frac{\frac{1}{\sqrt{2}}}{2} = -\frac{1}{2\sqrt{2}} \tag{680}$$

so $P''(1) < 0$ which implies that $x = 1$ is a relative maximum. Since $P(1) = \frac{3\sqrt{2}}{2} - \sqrt{2} = \frac{\sqrt{2}}{2}$.

By geometrical reasons we have that $0 \leq x \leq \sqrt{2}$ so we are in the same situation as in the previous problem, that is, we must see what happens at the endpoints. We have that $P(0) = \sqrt{2} - 1$ and $P\left(\sqrt{2}\right) = 3 + \sqrt{2} - 2 - \sqrt{3} = 1 + \sqrt{2} - \sqrt{3}$. Both $P(0)$ and $P\left(\sqrt{2}\right)$ are smaller than $P(1)$ so the fence should be built with $x = 1$.

Example 125. A truck gets $600/x$ mpg when driven at a constant speed of x mph (between 50 and 70 mph). If the price of fuel is \$3/gallon and the driver is paid \$18/hour, at what speed between 50 and 70 mph is it most economical to drive (for the company)?

Suppose that the truck wants to travel a distance l at speed x. Since the speed is constant the total driving time is

$$\text{total time } = \frac{l}{x} \tag{681}$$

in that time the driver will get paid

$$\text{money paid } = 18\frac{l}{x} \tag{682}$$

At the same time, the amount of fuel spent by traveling a distance l

$$\text{gallons spent} = \frac{\text{distance travelled}}{\text{efficiency (miles per galon)}} = \frac{l}{600/x} = \frac{lx}{600} \tag{683}$$

Therefore, the money spent on fuel is

$$\text{money spent fuel } = 3 \cdot \frac{lx}{600} = \frac{lx}{200} \tag{684}$$

In this way, the total money spent by the company is

$$P(x) = \frac{18l}{x} + \frac{lx}{200} = l\left(\frac{18}{x} + \frac{x}{200}\right) \tag{685}$$

The first derivative is

$$P'(x) = l\left(-\frac{18}{x^2} + \frac{1}{200}\right) \tag{686}$$

so the critical point satisfies

$$\frac{1}{200} = \frac{18}{x^2} \longrightarrow x = 60 \tag{687}$$

We can see that

$$P''(x) = l\left(\frac{36}{x^3}\right) \tag{688}$$

so $P''(60) > 0$ and therefore $x = 60$ corresponds to a minimum. Since $50 \leq x \leq 70$ we need to compare $P(50)$ and $P(70)$ but it can be seen that $P(60)$ is indeed the least value so the most economic strategy is that the driver should travel at 60 mph.

Example 126. Find two nonnegative numbers x and y whose sum is 300 and for which x^2y is a maximum.

We want to maximize x^2y. Given that $x + y = 300$ we have $y = 300 - x$ so the function to maximize is

$$f(x) = x^2y = x^2(300 - x) = 300x^2 - x^3 \tag{689}$$

The derivative of the function is

$$f'(x) = 600x - 3x^2 = 3x(200 - x) \tag{690}$$

Therefore the critical values are $x_1 = 0$ and $x_2 = 200$. Also, since x has to be positive and can't be bigger than 300 the restriction for x is $0 \leq x \leq 300$ therefore another candidate is $x_3 = 300$. Comparing $f(0)$, $f(200)$ and $f(300)$ we see that the maximum occurs when $x = 200$ and $y = 100$.

Example 127. Suppose that a beverage manufacturer wants to decide whether or not it should increase or decrease the airtime of the company's tv commerical to raise profits from its products. Suppose that x represents the number of tv hours per month in commercials. Thanks to the company's own research analysis, they know that the profit they make with x hours of commercials is $p(x) = 2\sqrt{x}$ million dollars and the fee for the TV commercial is \$10 000 per minute. If the company is currently buying 4 hours of commercials per month should the company increase or decrease the number of commercial hours?

Since a commercial costs \$10 000 per minute the company spends \$600 000 in one hour or 0.6 million dollars in one hour so in x hours the company has to spend $c(x) = 0.6x$ million dollars. Therefore, the net revenue is $r(x) = 2\sqrt{x} - 0.6x$ million dollars.

To see if the company has to spend more money in commercials or not we have to calculate $r'(4)$. Observe that

$$r'(x) = \frac{1}{\sqrt{x}} - 0.6 \tag{691}$$

Since

$$r'(4) = \frac{1}{2} - 0.6 = -0.1 \tag{692}$$

we see that $r'(4) < 0$ so the company should spend less money in tv commericals. In fact, the critical point is when $r'(x) = 0$ or $\frac{1}{\sqrt{x}} = 0.6$ which gives $x = \frac{25}{9} \simeq 2.77$. This critical point corresponds to a maximum (check using the second derivative) so the company should spend around 2.77 hours in tv commercials.

Example 128. There are two types of bubbles is sodas: relatively small bubbles that become smaller and eventually disappear while other bubbles are relatively large bubbles which eventually rise up to the surface. An explanation on why this happens is that CO_2 in carbonated drinks is supersaturated so it becomes more stable as a gas than when it is dissolved in fluid. Hence, the energy of a bubble decreases in proportion to its volume while on the other hand the surface tension acts on the surface between the bubble and the fluid which increases the energy of the bubble. Hence, a simplified model for the energy of the bubble is $E(r) = -c_1 r^3 + c_2 r^2$ where c_1, c_2 are some positive constants and r is the radius of the bubble. From physical considerations the bubble tries to minimize $E(r)$. Show this implies that there are two different behaviors depending on whether or not a radius is larger or smaller than the radius that maximizes the energy.

First we find the critical points by differentiating the energy

$$E'(r) = -3c_1 r^2 + 2c_2 r = r\left(-3c_1 r + 2c_2\right) \tag{693}$$

So the critical points are $r = 0$ and $r = \frac{2c_2}{3c_1}$. However, we will ignore the first critical point since the bubbles should have some positive radius (that is $r > 0$). Since $E''(r) = -6c_1 r$ we can see that $r_{max} = \frac{2c_2}{3c_1}$ gives a maximum. A qualitative behavior of $E(r)$ is given in the following figure.

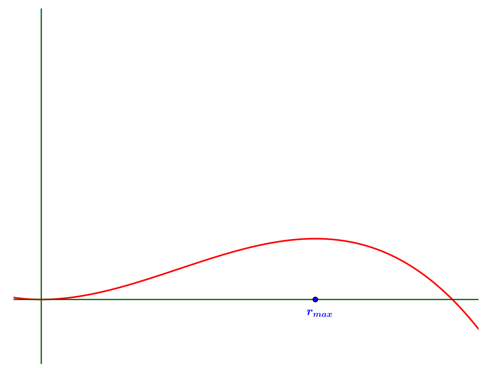


Figure 122: $E(r)$ with $c_1 = c_2 = 1$

From the figure we observe that if $r < r_{max}$ then the radius of the bubble will decrease so that the energy of the bubble can stay away from the maximum. On the other hand, if $r > r_{max}$ then the radius of the bubble will increase so that the energy of the bubble can stay away from the maximum. Therefore there are two different behaviors for the bubbles depending on its radius and this explains our observation.

Example 129. Suppose that we want to save a person who is drowning at the sea. Obviously it would be preferable to minimize the time it takes you to that person but your speed in the sand v_s is larger than your speed in the water v_w. At what point should you enter the water?

Suppose that you are the black dot and start at the origin $(0,0)$. We will cross the water at point (x, y) and the person drowning is located at point (a, b). The actual trajectory are the two green segments and the total time is

$$t_{total} = t_{sand} + t_{water} \tag{694}$$

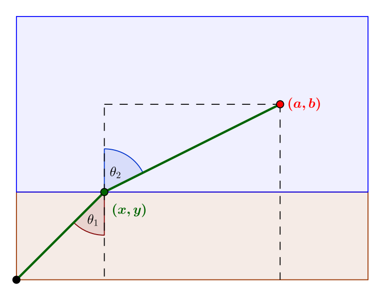


Figure 123: Snell's Law

Since the speed is constant in each medium we have by Pythagoras

$$t_{total} = \frac{\text{distance sand}}{v_s} + \frac{\text{distance water}}{v_w} = \frac{\sqrt{x^2 + y^2}}{v_s} + \frac{\sqrt{(a-x)^2 + (b-y)^2}}{v_w} \tag{695}$$

Now, y is a constant in this problem because it is just the distance between the shoreline and the x axis. Therefore, t_{total} is in fact just a function of x:

$$t_{total}(x) = \frac{\sqrt{x^2+y^2}}{v_s} + \frac{\sqrt{(a-x)^2+(b-y)^2}}{v_w} \tag{696}$$

and we want to find when does

$$\frac{dt_{total}}{dx} = 0 \tag{697}$$

(observe that in this problem t is differentiate with respect to x when usually x is the one differentiated with respect to time!). A simple calculation gives

$$\frac{dt_{total}}{dx} = \frac{x}{v_s\sqrt{x^2+y^2}} - \frac{a-x}{v_w\sqrt{(a-x)^2+(b-y)^2}} \tag{698}$$

The critical point must satisfy

$$\frac{x}{v_s\sqrt{x^2+y^2}} = \frac{a-x}{v_w\sqrt{(a-x)^2+(b-y)^2}} \tag{699}$$

Looking at the geometry of the problem we have

$$\sin\theta_1 = \frac{x}{\sqrt{x^2+y^2}} \qquad \sin\theta_2 = \frac{a-x}{\sqrt{(a-x)^2+(b-y)^2}} \tag{700}$$

so the condition for the critical point 699 can be written as

$$\frac{\sin\theta_1}{v_s} = \frac{\sin\theta_2}{v_w} \tag{701}$$

We can check that if x satisfies 699 then it is in fact a minimum so we have found the point at which we must enter the water: it is the point for which 701 is holds. As a matter of fact 701 is known as Snell's Law because light traveling two different media minimizes its total time in the same way as we do trying to save the drowning person.

Curve Sketching

A classical application of calculus is to determine how to sketch the graph of a function. Granted, the use of computers has somewhat diminished the need to know how to sketch the graph of a function, but it is a good way to test our knowledge of calculus. Before giving the basic algorithm on how to sketch a curve, we introduce some terminology.

- **Vertical Asymptotes:** Suppose that $f(x)$ is a function. Then the vertical line $x = a$ is a vertical asymptote if at least one of the one sided limits $\lim_{x\longrightarrow a^+} f(x)$, $\lim_{x\longrightarrow a^-} f(x)$ is ∞ or $-\infty$
- **Horizontal Asymptotes:** The horizontal line $y = b$ is a horizontal asymptote for the graph of the function if $\lim_{x\longrightarrow\infty} f(x) = b$ or $\lim_{x\longrightarrow-\infty} f(x) = b$.

Guide to Curve Sketching : Suppose $y = f(x)$ is a function. To sketch the graph of $f(x)$

1. Determine the domain of $f(x)$ and find if they are vertical asymptotes (often they occur at the points where f is undefined)
2. Determine if they are horizontal asymptotes or not by studying $\lim_{x\longrightarrow\infty} f(x)$ and $\lim_{x\longrightarrow-\infty} f(x)$
3. Find the points x where $f(x) = 0$ and $f(0)$ if possible.
4. Find $f'(x)$ and use the first derivative test to determine where $f(x)$ is increasing and decreasing.
5. Find the critical points and try to classify them either with the first or second derivative test
6. Use $f''(x)$ to determine the concavity of f
7. Find the inflection points of $f(x)$

Example 130. Sketch the graph of the function $f(x) = 2x^3 + 1$

Following the previous guide:

- Because $f(x)$ is a polynomial the domain of $f(x)$ is $\mathbb{R}$ and there are no vertical asymptotes
- We have $\lim_{x \longrightarrow \infty} \left(2x^3 + 1\right) = \infty$ and $\lim_{x \longrightarrow -\infty} \left(2x^3 + 1\right) = -\infty$ so there are no horizontal asymptotes.
- $f(x) = 0$ when $2x^3 = -1$ or $x = -\frac{1}{\sqrt[3]{2}}$. Also, $f(0) = 1$
- We have $f'(x) = 6x^2$ so the only critical point is $x = 0$. Moreover, because $f'(x)$ is always positive then $f(x)$ is always increasing.
- We have $f''(x) = 12x$ so $x = 0$ is an inflection point.
- If $x < 0$ then $f(x)$ is concave down and if $x > 0$ then $f(x)$ is concave up

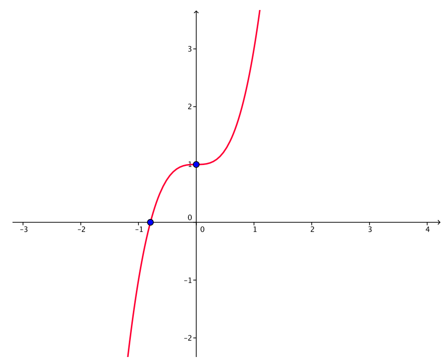


Figure 124: $y = 2x^3 + 1$

Example 131. Sketch the graph of the function $f(x) = \sqrt[3]{x^2}$

Following the previous guide:

- The cubic root of any number always exists so the domain of $f(x)$ is $\mathbb{R}$. Also, there are no vertical asymptotes.
- We have $\lim_{x \longrightarrow \infty} \sqrt[3]{x^2} = \infty$ and $\lim_{x \longrightarrow -\infty} \sqrt[3]{x^2} = \infty$ so there are no horizontal asymptotes.
- $f(x) = 0$ when $x = 0$.
- We have $f'(x) = \frac{2}{3}x^{-\frac{1}{3}} = \frac{2}{3}\frac{1}{\sqrt[3]{x}}$ so there are no critical points and $f'(x)$ is not defined at $x = 0$. We have that $f(x)$ is decreasing when $x < 0$ and $f(x)$ is increasing when $x > 0$
- Since there are no critical points we skip this step
- Because $f''(x) = -\frac{2}{9}x^{-\frac{4}{3}} = -\frac{2}{9}\frac{1}{\sqrt[3]{x^4}} < 0$ the function is concave down.

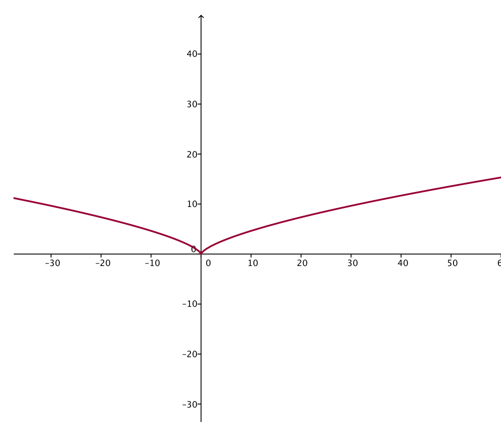


Figure 125: $y = \sqrt[3]{x^2}$

Example 132. Sketch the graph of the function $f(x) = \frac{x^2-9}{x^2-4}$

Following the previous guide:

- Because $f(x) = \frac{(x-3)(x+3)}{(x-2)(x+2)}$ the domain of $f(x)$ is the real line except $x = 2$ and $x = -2$. Observe that

$$\begin{array}{ll} \lim_{x \longrightarrow 2^+} \frac{(x-3)(x+3)}{(x-2)(x+2)} = -\infty & \lim_{x \longrightarrow 2^-} \frac{(x-3)(x+3)}{(x-2)(x+2)} = \infty \\ \lim_{x \longrightarrow -2^+} \frac{(x-3)(x+3)}{(x-2)(x+2)} = \infty & \lim_{x \longrightarrow -2^-} \frac{(x-3)(x+3)}{(x-2)(x+2)} = -\infty \end{array} \tag{702}$$

so $x = 2$ and $x = -2$ are vertical asymptotes.

- To find if there are horizontal asymptotes observe that

$$\begin{aligned}\lim_{x\longrightarrow\infty}\frac{x^2-9}{x^2-4} &= \lim_{x\longrightarrow\infty}\frac{1-\frac{9}{x^2}}{1-\frac{4}{x^2}} = 1\\ \lim_{x\longrightarrow-\infty}\frac{x^2-9}{x^2-4} &= \lim_{x\longrightarrow-\infty}\frac{1-\frac{9}{x^2}}{1-\frac{4}{x^2}} = 1\end{aligned} \tag{703}$$

  so $y=1$ is a horizontal asymptote.

- We have that $f(x)=0$ when $x=3$ or $x=-3$ and $f(0)=\frac{9}{4}$

- We have

$$f'(x) = \frac{(2x)\left(x^2-4\right)-\left(x^2-9\right)(2x)}{\left(x^2-4\right)^2} = \frac{10x}{\left(x^2-4\right)^2} \tag{704}$$

  so $x = 0$ is a critical point. Also, on $(-\infty,-2)$, $(-2,0)$ $f(x)$ is decreasing and on $(0,2)$, $(2,\infty)$ $f(x)$ is increasing.

- The second derivative is

$$f''(x) = \frac{10\left(x^2-4\right)^2-10x(2)(x^2-4)(2x)}{\left(x^2-4\right)^4} = -\frac{10\left(3x^2+4\right)}{\left(x^2-4\right)^3} \tag{705}$$

  given that $f''(0)>0$ the point $x=0$ is a relative minimum.

- To find the concavity of $f(x)$ the sign of $f''(x)$ is determined by the sign of $\left(x^2-4\right)^3$ since $10(3x^2+4)$ is always positive. When $-2<x<2$ we have that $\left(x^2-4\right)^3$ is negative so $f''(x)$ is positive because of the negative sign in front of the fraction. When $x>2$ or $x<-2$ we have that $f''(x)$ is negative so $f(x)$ is concave up on $(-2,2)$ and concave down on the intervals $(-\infty,-2)$ and $(2,\infty)$

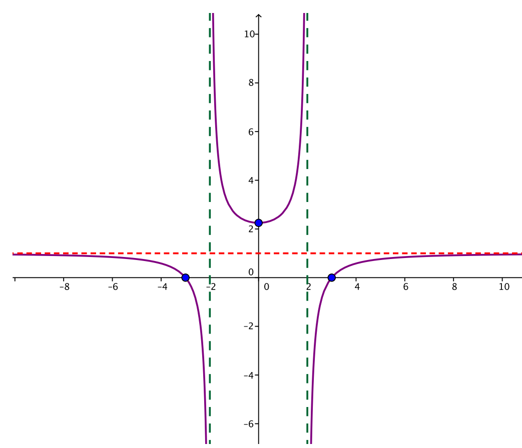


Figure 126: $y=\frac{x^2-9}{x^2-4}$

Mean Value Theorem

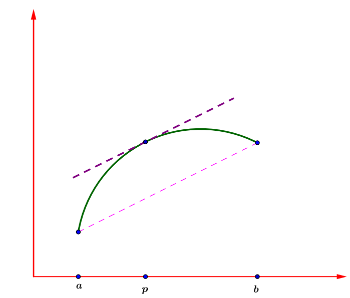


Figure 127: MVT, 1 point

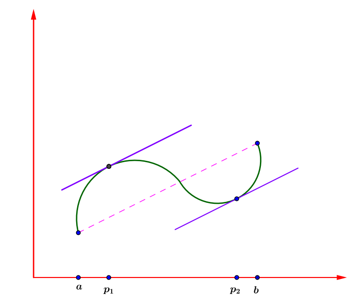


Figure 128: MVT, 2 points

A very powerful theorem about derivatives is the Mean Value Theorem, or MVT. The idea behind it is very intuitive: suppose we have a function defined on $[a, b]$. The slope of the secant line connecting the endpoint $(a, f(a))$, $(b, f(b))$ is

$$m_{[a,b]}f = \frac{f(b) - f(a)}{b - a} \tag{706}$$

The Mean Value Theorem simply states that there must be a point whose tangent line has the same slope as the slope of the secant line connecting the endpoints. As shown in the following figures, there can be more than one point that works, but the Mean Value Theorem guarantees the existence of at least one point, it does not say how many points there are.

Mean Value Theorem: Suppose that $f(x)$ is a continuous function in the closed interval $[a, b]$ and differentiable in the open interval (a, b). Then there is a point $p \in (a, b)$ such that

$$f'(p) = \frac{f(b) - f(a)}{b - a} \tag{707}$$

Proof. The equation of the secant line is

$$y = \frac{f(b) - f(a)}{b - a}(x - a) + f(a) \tag{708}$$

If we define the function on $[a, b]$

$$g(x) \equiv f(x) - \left[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right] \tag{709}$$

then we are measuring the vertical distance between a point on the graph of the function and the secant line. Since g is a continuous function on a closed interval the Extreme Value Theorem says that g achieves an absolute maximum and an absolute minimum. Given that $g(a) = 0 = g(b)$ at least one of these points must be inside (a, b), otherwise the function $g(x)$ would be constant (since the absolute maximum would equal the absolute minimum) and the result would be obvious because $f(x)$ would be equal to the secant line.

Assume without loss of generality that the maximum at p lies in the open interval (a, b). Then p corresponds also to relative maximum so by part 2 of the first derivative test $g'(p) = 0$. But

$$g'(x) = f'(x) - \left(\frac{f(b) - f(a)}{b - a}\right) \tag{710}$$

and evaluating at p we obtain $f'(p) = \frac{f(b)-f(a)}{b-a}$ which is the statement of the mean value theorem. □

There are many consequences to the Mean Value Theorem, which we now present:

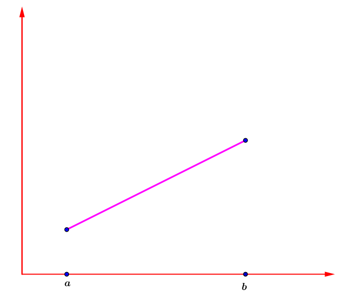


Figure 129: MVT, infinite points

Consequence 1 MTV: Proof the First Derivative Test, Part 3: Suppose f is differentiable in a deleted neighborhood of a critical point p, and suppose the derivative f' has one sign to the left of p and the opposite sign to the right of p

- If the sign of f' changes from negative to positive then p is a relative minimum.
- If the sign of f' changes from positive to negative then p is a relative maximum

Proof. We will show the first case since the second one is entirely analogous. Call the deleted neighborhood of p $(p - \delta, p) \cup (p, p + \delta)$. Take $x \in (p, p + \delta)$. By the Mean Value Theorem applied to the interval (p, x) we obtain the existence of some point c such that

$$f'(c) = \frac{f(x) - f(p)}{x - p} \tag{711}$$

Since $f'(c) > 0$ and $x - p > 0$ this means that $f(x) - f(p) > 0$ so $f(x) > f(p)$. On the other hand, if we take $x \in (p - \delta, p)$ the Mean Value Theorem applied to (x, p) gives the existence of some point c such that

$$f'(c) = \frac{f(p) - f(x)}{p - x} \tag{712}$$

Because $f'(c) < 0$ and $p - x > 0$ we obtain $f(p) - f(x) < 0$ or $f(p) < f(x)$. Regardless of the interval, we see that $f(p) < f(x)$ so f is a relative minimum. □

Consequence 2 **MVT**: If a function f has a vanishing derivative on the interval (a, b) then f must be constant on (a, b)

Proof. Suppose that $x_1 < x_2$ are two points on (a, b). To show that $f(x_1) = f(x_2)$ observe that by the Mean Value Theorem applied to the interval (x_1, x_2) we obtain the existence of some point p such that

$$f'(p) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \tag{713}$$

Since the derivative of f is zero we have $f'(p) = 0$ so $f(x_1) = f(x_2)$ and f must be constant. □

Average Net Change Interpretation of the Mean Value Theorem: Let f be a continuous function on $[a, b]$ with derivative f' on the interval (a, b). The net change of f on $[a, b]$ is by definition

$$\triangle f \equiv f(b) - f(a) \tag{714}$$

and its average net change is

$$\text{average change } f \text{ on } [a, b] = \frac{f(b) - f(a)}{b - a} \tag{715}$$

Then the Mean Value Theorem states that there exists some point p such that the instantaneous change of f at p, that is, $f'(p)$, equals the average change of f on $[a, b]$

$$f'(p) = \frac{f(b) - f(a)}{b - a} \tag{716}$$

In particular, if $x(t)$ represents the position of particle from t_1 to t_2 then the Mean Value Theorem says that there is some time $t_a \in (t_1, t_2)$ for which the instantaneous velocity of the particle at that time is the same as the average velocity of the particle

$$v(t_a) = v_{avg} = \frac{x(t_2) - x(t_1)}{t_2 - t_1} \tag{717}$$

Example 133. A college student, fresh from doing well on her calculus final exam, is driving home for the summer. She enters a freeway at 6pm and her odometer at that moment reads 42452 miles. She exists the freeway at 8pm after driving 150 miles. a) The speed limit on the freeway is 70 mph. Which calculus theorem guarantees that the student broke the speed limit during her trip? Explain. b) Which calculus theorem guarantees that at one point during the student's trip, her odometer read 42500? Explain.

a) Given that the student travelled 150 miles in two hours, here average speed is

$$v_{avg} = \frac{150 \text{ miles}}{2 \text{ hours}} = 75 \text{ miles/hour} \tag{718}$$

By the Mean Value Theorem, there must have been an instant when the student achieved her average speed; so at that instant she broke the speed limit.

b) Since the student drove for 150 miles the odometer reads at 8 pm $42452 + 150 = 42602$ miles. Since the position changes continuously, the intermediate value theorem says that at some instant the odometer read 42500 miles.

Example 134. Show that the equation $2x^3 + 2x^2 + x + 1 = 0$ has only one real root

We have proved before that polynomials of odd degree have at least one real root. Assume that there is more than one root, for example, assume that there are at least two roots $r_1 < r_2$. If we let $p(x) = 2x^3 + 2x^2 + x + 1$ be the polynomial then by the Mean Value Theorem applied to (r_1, r_2) we have the existence of a point c such that

$$p'(c) = \frac{p(r_2) - p(r_1)}{r_2 - r_1} = \frac{0 - 0}{r_2 - r_1} = 0 \tag{719}$$

Now, the derivative of $p(x)$ is

$$p'(x) = 6x^2 + 4x + 2 \tag{720}$$

but the discriminant of $6x^2 + 4x + 2$ is negative, so $p'(x)$ is never zero, which is a contradiction since $p'(c) = 0$. Therefore, there is exactly one real root.

L'Hospital's Theorem

Almost all of the interesting limits can't be evaluated by simple substitution. For example, to find the limit

$$\lim_{x \longrightarrow 0} \frac{\sin x}{x} = 1 \tag{721}$$

we can't just simply substitute x for 0 because we would have to evaluate $\frac{\sin 0}{0} = \frac{0}{0}$ which is clearly impossible. L'Hospital Theorem gives us a quick way to find this limit and many other kinds of limits which are known as indeterminate forms.

L'Hospital's Theorem:
Let f and g be real valued functions and α be any of $a^{\pm}$, $\pm\infty$. Suppose that $g'(x) \neq 0$ on an open interval I that contains a. Suppose that one of the following cases holds:

$$\begin{gathered} \lim_{x \longrightarrow \alpha} f(x) = 0 \qquad \text{AND} \qquad \lim_{x \longrightarrow \alpha} g(x) = 0 \\ \text{OR} \\ \lim_{x \longrightarrow \alpha} f(x) = \pm\infty \qquad \text{AND} \qquad \lim_{x \longrightarrow \alpha} g(x) = \pm\infty \end{gathered} \tag{722}$$

Then provided that

$$\lim_{x \longrightarrow \alpha} \frac{f'(x)}{g'(x)} = L \tag{723}$$

with $L \in \mathbb{R}$ or $L = \pm\infty$ we have

$$\lim_{x \longrightarrow \alpha} \frac{f(x)}{g(x)} = \lim_{x \longrightarrow \alpha} \frac{f'(x)}{g'(x)} \tag{724}$$

Proof. Rather than proving the theorem, we will just give an heuristic argument for the theorem when α is a finite number. If we do the linear approximation around a we can write

$$\begin{aligned} f(x) &\simeq f(a) + f'(a)\,(x-a) \\ g(x) &\simeq g(a) + g'(a)\,(x-a) \end{aligned} \tag{725}$$

Now, assume that we are in the first case: we can take $f(a) = g(a) = 0$ so

$$\lim_{x \longrightarrow a} \frac{f(x)}{g(x)} = \lim_{x \longrightarrow a} \frac{f'(a)(x-a)}{g'(a)(x-a)} = \lim_{x \longrightarrow a} \frac{f'(a)}{g'(a)} = \frac{f'(a)}{g'(a)} = \lim_{x \longrightarrow a} \frac{f'(x)}{g'(x)} \tag{726}$$

For this argument to work rigorously we would need to assume that f', g' are continuous near a and that $g'(a) \neq 0$ but since we are not going to give the full proof of the theorem we won't worry about the details.

$\square$

Example 135. Find $\lim_{x \longrightarrow 0} \frac{\sin x}{x}$ using L'Hospital's rule.

We have by L'Hospital's Theorem

$$\begin{aligned} \lim_{x \longrightarrow 0} \frac{\sin x}{x} &\overset{L'H}{\underset{[\frac{0}{0}]}{=}} \lim_{x \longrightarrow 0} \frac{(\sin x)'}{x'} \\ &= \lim_{x \longrightarrow 0} \cos x \\ &= 1 \end{aligned} \tag{727}$$

which is what we had found before using a geometrical argument.

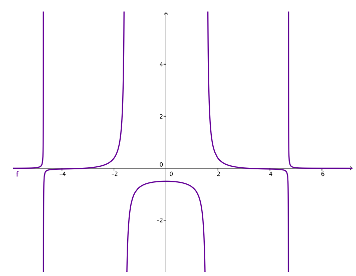


Figure 130: $y = \frac{\sin x - \tan x}{x^3}$

Example 136. Find $\lim_{x \longrightarrow 0} \frac{\sin(x) - \tan(x)}{x^3}$

We have by L'Hospital's Theorem

$$\begin{aligned} \lim_{x \longrightarrow 0} \frac{\sin(x) - \tan(x)}{x^3} &\overset{L'H}{\underset{[\frac{0}{0}]}{=}} \lim_{x \longrightarrow 0} \frac{\cos(x) - \sec^2(x)}{3x^2} \\ &\overset{L'H}{\underset{[\frac{0}{0}]}{=}} \lim_{x \longrightarrow 0} \frac{-\sin x - 2\sec x(\sec x \tan x)}{6x} \\ &= \lim_{x \longrightarrow 0} \frac{-\sin x - 2\sec^2 x \tan x}{6x} \\ &\overset{L'H}{\underset{[\frac{0}{0}]}{=}} \lim_{x \longrightarrow 0} \frac{-\cos x - 2\left(2\sec x \sec x \tan x \tan x + \sec^2 x \sec^2 x\right)}{6} \\ &= \frac{-1 - 2(0+1)}{6} \\ &= -\frac{1}{2} \end{aligned} \tag{728}$$

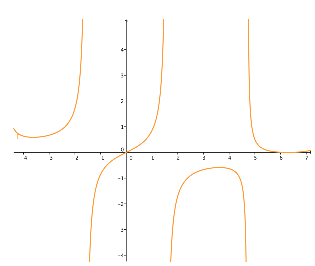


Figure 131: $y = \frac{\sec \theta - 1}{\theta}$

Example 137. Find $\lim_{\theta \longrightarrow 0} \frac{\sec \theta - 1}{\theta}$

We have by L'Hospital's Theorem

$$\begin{aligned} \lim_{\theta \longrightarrow 0} \frac{\sec \theta - 1}{\theta} &\overset{L'H}{\underset{[\frac{0}{0}]}{=}} \lim_{\theta \longrightarrow 0} \sec \theta \tan \theta \\ &= 0 \end{aligned} \tag{729}$$

Example 138. Find $\lim_{x\longrightarrow\infty}\left(\ln(3x+1)-\ln(2x+2)\right)$
By the properties of logarithm and L'Hospital's Theorem

$$\begin{aligned}
\lim_{x\longrightarrow\infty}\left(\ln(3x+1)-\ln(2x+2)\right) &= \lim_{x\longrightarrow\infty}\ln\left(\frac{3x+1}{2x+1}\right)\\
&=^{\text{continuity}} \ln\left(\lim_{x\longrightarrow\infty}\frac{3x+1}{2x+1}\right)\\
&=^{L'H}_{\left[\frac{\infty}{\infty}\right]} \ln\left(\lim_{x\longrightarrow\infty}\frac{3}{2}\right)\\
&= \ln\left(\frac{3}{2}\right)
\end{aligned} \tag{730}$$

Example 139. Find $\lim_{x\longrightarrow 1}\frac{x^{99}-1}{x-1}$

$$\begin{aligned}
\lim_{x\longrightarrow 1}\frac{x^{99}-1}{x-1} &=^{L'H}_{\left[\frac{0}{0}\right]} \lim_{x\longrightarrow 1}99x^{98}\\
&= 99
\end{aligned} \tag{731}$$

As the following examples show, sometimes L'Hospital Rule can't be applied directly. For example, we may have indeterminate forms $0\cdot\infty$ or $\infty-\infty$. In some cases these forms can be solved by turning them into quotients and then applying L'Hospital Rule

Example 140. Find $\lim_{x\longrightarrow 0^+}\sqrt{x}\ln x$
This is an indeterminate form $0\cdot\infty$

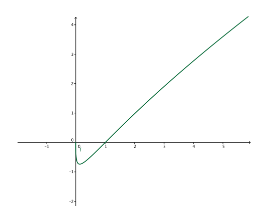


Figure 132: $y=\sqrt{x}\ln x$

$$\begin{aligned}
\lim_{x\longrightarrow 0^+}\sqrt{x}\ln x &=_{0\cdot\infty} \lim_{x\longrightarrow 0^+}\frac{\ln x}{x^{-1/2}}\\
&=^{L'H}_{\left[\frac{\infty}{\infty}\right]} \lim_{x\longrightarrow 0^+}\frac{\frac{1}{x}}{-\frac{x^{-3/2}}{2}}\\
&= \lim_{x\longrightarrow 0^+}-2\sqrt{x}\\
&= 0
\end{aligned} \tag{732}$$

Example 141. Find $\lim_{x\longrightarrow 1}\left(\frac{1}{\ln x}-\frac{1}{x-1}\right)$

This is an indeterminate form $\infty - \infty$

$$\begin{aligned}
\lim_{x\longrightarrow 1}\left(\frac{1}{\ln x}-\frac{1}{x-1}\right) &=_{\infty-\infty} \lim_{x\longrightarrow 1}\frac{x-1-\ln x}{(x-1)\ln x}\\
&=^{L'H}_{\left[\frac{0}{0}\right]} \lim_{x\longrightarrow 1}\frac{1-\frac{1}{x}}{\ln x+\frac{x-1}{x}}\\
&= \lim_{x\longrightarrow 1}\frac{\frac{x-1}{x}}{\frac{x\ln x+x-1}{x}}\\
&= \lim_{x\longrightarrow 1}\frac{x-1}{x\ln x+x-1}\\
&=^{L'H}_{\left[\frac{0}{0}\right]} \lim_{x\longrightarrow 1}\frac{1}{\ln x+1+1}\\
&= \frac{1}{2}
\end{aligned} \tag{733}$$

Other times we run into the indeterminate forms 0^0, ∞^0 and 1^∞. [48]In this case we try to use the property $a^b = e^{b\ln a}$ as the following examples show.

[48] Why is 0^0 considered an indeterminate form? Well, this is because there are two reasonable definitions for 0^0. One option is to define $0^0 = 1$; if we do that then the function $f(x) = 0^x$ (for $x \geq 0$) would not be continuous at 0. The other option is to define $0^0 = 0$; however, this seems to go agains the convention that $b^0 = 1$ for $b > 0$.

Example 142. Find $\lim_{x\longrightarrow\infty} x^{1/x}$

This is an indeterminate form ∞^0

$$\begin{aligned}
\lim_{x\longrightarrow\infty} x^{1/x} &= \lim_{x\longrightarrow\infty} e^{\ln\left(x^{1/x}\right)}\\
&= \lim_{x\longrightarrow\infty} e^{\frac{\ln x}{x}}\\
&=_{\text{continuity}} e^{\lim_{x\longrightarrow\infty}\frac{\ln x}{x}}\\
&=_{L'H} e^{\lim_{x\longrightarrow\infty}\frac{1}{x}}\\
&= e^0\\
&= 1
\end{aligned} \tag{734}$$

Example 143. Find $\lim_{x\longrightarrow 0^+}(\tan 5x)^x$

We have the indeterminate form 0^0

$$
\begin{aligned}
\lim_{x\longrightarrow 0^+} (\tan 5x)^x &= \lim_{x\longrightarrow 0^+} e^{\ln(\tan 5x)^x} \\
&= \lim_{x\longrightarrow 0^+} e^{x\ln(\tan 5x)} \\
&= e^{\lim_{x\longrightarrow 0^+} x\ln(\tan 5x)} \\
&= e^{\lim_{x\longrightarrow 0^+} \frac{\ln(\tan 5x)}{x^{-1}}} \\
&=_{L'H} e^{\lim_{x\longrightarrow 0^+} \frac{\frac{5\sec^2(5x)}{\tan 5x}}{-x^{-2}}} \qquad (735) \\
&= e^{\lim_{x\longrightarrow 0^+} \frac{-5x^2}{\cos(5x)\sin(5x)}} \\
&=_{L'H} e^{\lim_{x\longrightarrow 0^+} \frac{-10x}{-5\sin^2(5x)+5\cos^2(5x)}} \\
&= e^0 \\
&= 1
\end{aligned}
$$

Example 144. Find $\lim_{x\longrightarrow 0^+} (1+\sin 7x)^{\cot(5x)}$
This is an indeterminate form 1^∞:

$$
\begin{aligned}
\lim_{x\longrightarrow 0^+} (1+\sin 7x)^{\cot(5x)} &= \lim_{x\longrightarrow 0^+} e^{\ln\left[(1+\sin 7x)^{\cot(5x)}\right]} \\
&= \lim_{x\longrightarrow 0^+} e^{\cot(5x)\ln(1+\sin 7x)} \\
&=_{\text{continuity}} e^{\lim_{x\longrightarrow 0^+} \frac{\cos(5x)\ln(1+\sin 7x)}{\sin(5x)}} \\
&= e^{\lim_{x\longrightarrow 0^+} \cos(5x) \lim_{x\longrightarrow 0^+} \frac{\ln(1+\sin 7x)}{\sin(5x)}} \\
&= e^{\lim_{x\longrightarrow 0^+} \frac{\ln(1+\sin 7x)}{\sin(5x)}} \\
&=_{L'H} e^{\lim_{x\longrightarrow 0^+} \frac{\frac{7\cos 7x}{1+\sin 7x}}{5\cos(5x)}} \\
&= e^{7/5} \qquad (736)
\end{aligned}
$$

Part V

Integration of Functions

The Integral as Area under the Curve

Up to this point, most of our efforts have been concentrated on calculating derivatives of functions. We saw that from a geometrical point of view, the derivative is the slope of the tangent line and this classical problem turned out to be applicable in many other situations.

In the same spirit, another classical problem turns out to be extremely important and together with the problem of finding tangent lines constitute the two building blocks of calculus. This is the problem of finding areas. For example, we all know that the area of a circle is $A = \pi r^2$, but how can we find this formula? One approach done by the greeks (specially Archimedes) was to approximate the area of the circle by the area of regular polygons inscribed in the circle[49].

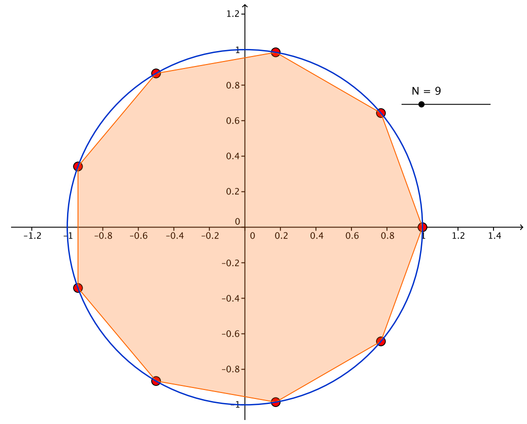


Figure 133: Approximation area of a circle

[49] There is an animation in the following link

If we call P_N the area of the regular polygon with N sides then as a N becomes bigger and bigger we expect that they approximate better the area of the circle so in "the limit" we should have that

$$\pi R^2 = \lim_{N \longrightarrow \infty} P_N \tag{737}$$

(at the time of Archimedes the limit concept was not available but his method of exhaustion is in fact a predecessor for our understanding of limits).

Suppose that we have a function $y = f(x)$. The graph of the function is a curve and from a geometrical point of view we can ask for the value of the area under the curve on an interval $[a, b]$. This area under the curve is called the integral of $f(x)$ on $[a, b]$ and will be denoted $\int_a^b f(x)dx$.[50]

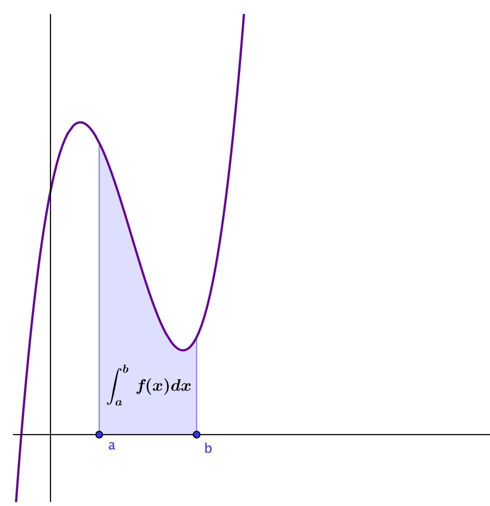


Figure 134: Area under the curve of $y = f(x)$

[50] As we will see soon, geometrically the integral is a "signed" area because if the curve is under the x axis the area is considered negative. Also, the notation will make more sense after we have discussed Riemann sums.

From classical geometry, the formulas we know to compute areas work for triangles, polygons, circles and some other special curves. However, this is not sufficient for many curves that appear in applications, like the one in the previous figure which is the graph of $y = x^3 - 5x^2 + 5x + 5$.

The strategy that we will use is to work in fact with only *one* formula from classical geometry: the formula for the area of a rectangle. Suppose we want to find the area under the curve of $f(x)$ on the interval $[a, b]$. The length of the interval is

$$L = b - a \tag{738}$$

we divide the interval $[a, b]$ into n subintervals of length $\frac{L}{n}$. The length of these subintervals will be called $\triangle x$:

$$\triangle x = \frac{L}{n} = \frac{b-a}{n} \tag{739}$$

These intervals are $[a, a+\triangle x]$, $[a+\triangle x, a+2\triangle x]$, $\cdots$, $[a+(n-2)\triangle x, a+(n-1)\triangle x]$ and $[a+(n-1)\triangle x, a+n\triangle x]$. We will approximate the area under the curve by the area of the rectangles which sit on top of each subinterval and whose heights are the values of the left-endpoints of the subintervals (the following link shows this method).

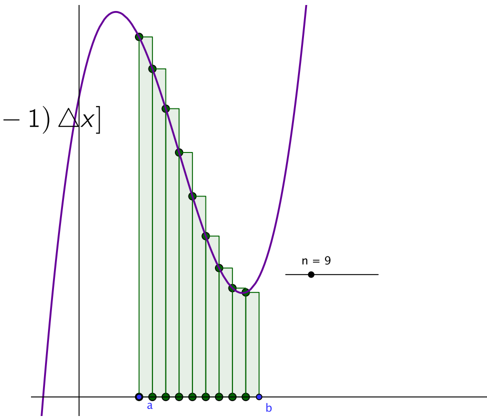


Figure 135: Approximation to the Area

For example, the area of the rectangle on top of $[a, a+\triangle x]$ has height $f(a)$ and base $\triangle x$ so its area is $f(a)\triangle x$. More generally, we have

$$\begin{array}{cc} \text{interval} & \text{area} \\ & \\ [a, a+\triangle x] & f(a)\triangle x \\ [a+\triangle x, a+2\triangle x] & f(a+\triangle x)\triangle x \\ \vdots & \vdots \\ [a+(n-2)\triangle x, a+(n-1)\triangle x] & f(a+(n-2)\triangle x)\triangle x \\ [a+(n-1)\triangle x, a+n\triangle x] & f(a+(n-1)\triangle x)\triangle x \end{array} \tag{740}$$

The total area of all the rectangles is $f(a)\triangle x + f(a+\triangle x)\triangle x + \cdots + f(a+(n-2)\triangle x)\triangle x + f(a+(n-1)\triangle x)\triangle x$. Since $\triangle x$ appears in every term we can factorize it so the total area becomes

$$\text{area rectangles} = (f(a) + f(a+\triangle x) + \cdots + f(a+(n-2)\triangle x) + f(a+(n-1)\triangle x))\triangle x \tag{741}$$

The area under the curve should be the limit of the area in 741 as $n \longrightarrow \infty$. Since $\triangle x = \frac{b-a}{n}$, $n \longrightarrow \infty$ is the same as $\triangle x \longrightarrow 0$. We encountered a similar situation when we defined the derivative, so in the limit $\triangle x$ is replaced by dx and $\int_a^b f(x)dx$ is used to represent the limit of the areas 741.

As an example, we will find $\int_1^2 xdx$, that is, the area under the curve of the line $y = x$ from 1 to 2.

By simple geometry, the area under the curve is equal to the area of the triangle plus the area of the square show in the previous figure so we expect

$$\int_1^2 xdx = 1^2 + \frac{1}{2}1 \cdot 1 = \frac{3}{2} \tag{742}$$

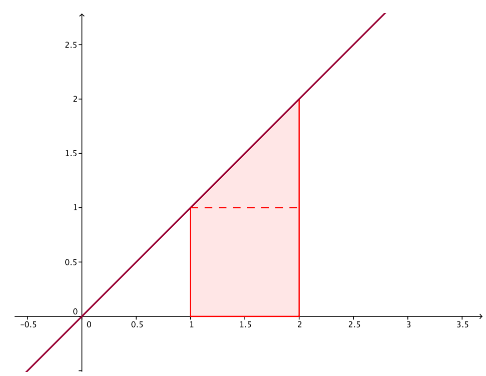


Figure 136: $\int_1^2 xdx$

If we use formula 741 with $a = 1$, $b = 2$ and $\triangle x = \frac{b-a}{n} = \frac{1}{n}$ we have

$$\text{area rectangles} = \left(f(1) + f\left(1+\frac{1}{n}\right) + \cdots + f\left(1+(n-2)\frac{1}{n}\right) + f\left(1+(n-1)\frac{1}{n}\right)\right)\frac{1}{n} \tag{743}$$

Here $f(x) = x$ so the area becomes

$$\text{area rectangles} = \left(\underbrace{1}_{*} + \underbrace{1}_{*} + \underbrace{\frac{1}{n}}_{**} + \cdots + \underbrace{1}_{*} + \underbrace{\frac{n-2}{n}}_{**} + \underbrace{1}_{*} + \underbrace{\frac{n-1}{n}}_{**}\right)\frac{1}{n} \tag{744}$$

We separate the sum in the parentheses into two terms: the terms with $*$ are just 1's and since there are n we have

$$\text{sum} * = n \tag{745}$$

The sum in $**$ all have $\frac{1}{n}$ as a common term so can factorize it to get

$$\text{sum } ** = \frac{1}{n}\left(1+2+\cdots+(n-2)+(n-1)\right) = \frac{1}{n}\left[\frac{(n-1)n}{2}\right] \tag{746}$$

where in the last step we used Gauss formula for the sum of the first $n-1$ terms. Therefore, 744 becomes

$$\text{area rectangles} = \left(n+\frac{1}{n}\left[\frac{(n-1)n}{2}\right]\right)\frac{1}{n} = \left(n+\frac{n-1}{2}\right)\frac{1}{n} = \frac{3n-1}{2n} = \frac{3}{2}-\frac{1}{2n} \tag{747}$$

Given that

$$\lim_{n\longrightarrow\infty}\left(\frac{3}{2}-\frac{1}{2n}\right) = \frac{3}{2} \tag{748}$$

we can see that the approximation of the area by rectangles gives the same result as the one derived by classical geometry.

This calculation shows that even for something as simple as a straight line, the calculation of integrals as a limit of areas can be a long procedure. Therefore, we need a different method for finding integrals and fortunately there is such procedure, given by the Fundamental Theorem of Calculus.

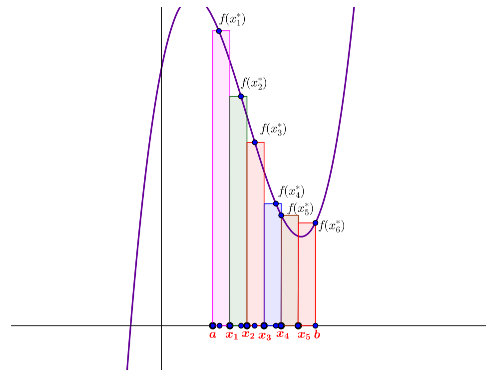


Figure 137: Definition Riemann Sum

Before stating the Fundamental Theorem of Calculus, it is useful to give a precise definition of the definite integral since we will need it to prove the Mean Value Theorem for Integrals, which is necessary for the Fundamental Theorem. Suppose we start with a function f defined on $[a,b]$. To define $\int_a^b f(x)dx$ we divide the interval $[a,b]$ into n intervals with endpoints

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b \tag{749}$$

In general, we don't require all the intervals to have the same length. Rather, we call their lengths

$$\triangle x_i \equiv x_i - x_{i-1} \qquad (i = 1,2,\cdots,n) \tag{750}$$

On each interval we pick an arbitrary point x_i^* and consider the sum

$$S = \sum_{i=1}^{n} f\left(x_i^*\right)\triangle x_i = f\left(x_1^*\right)\triangle x_1 + f\left(x_2^*\right)\triangle x_2 + \cdots + f\left(x_n^*\right)\triangle x_n \tag{751}$$

If we let $l = \max\{\triangle x_1,\cdots,\triangle x_n\}$ then we want to say

$$\int_a^b f(x)dx \equiv \lim_{l\longrightarrow 0}\sum_{i=1}^{n} f\left(x_i^*\right)\triangle x_i \tag{752}$$

Average Value and Mean Value Theorem

Suppose that we want to find the average value of the temperature during the day. If can call our function $T(t)$, then t takes values in the interval $[0, 24]$. For the sake of generality, we will assume that T takes values on some interval $[a, b]$.

The problem we have with trying to find the average value of the temperature is that the usual formula for average involves a finite number of data: for example, if the grades on the exams of 5 students are $80, 60, 90, 85, 85$ then their average grade would be

$$\text{average grade} = \frac{80 + 60 + 90 + 85 + 85}{5} \tag{753}$$

However, temperature is considered a continuous function, so in principle there is an infinite number of data points: one value of the temperature T for each time $t \in [a, b]$. Therefore, we need to approach in another way. What we do is to divide the time interval $[a, b]$ into subintervals

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b \tag{754}$$

In each subinterval we will *assume* that the temperature is constant (or almost constant), so that we can work with a finite number of temperature measurements. That is, we choosing times $t_1^*, \cdots, t_n^*$, one for each subinterval, and we declare the temperature on each subinterval to be equal to $T\left(t_i^*\right)$. Clearly if the temperature fluctuates a lot on a subinterval this would be a bad approximation, however, our intention is making the duration of these intervals smaller so that it becomes better.

Now we have turned the problem into a problem similar to the grades of the students: we have n measurements $T(t_1^*), \cdots, T\left(t_n^*\right)$ and so the average temperature becomes

$$\text{approximation average } T = \frac{T\left(t_1^*\right) + \cdots + T\left(t_n^*\right)}{n} \tag{755}$$

Right now this formula has no apparent relationship with the concept of definite integral we introduced before, so our objective is to find a way to make this relationship evident. Suppose that each subinterval has the same duration $\triangle t$, that is, we take

$$\triangle t = \frac{b - a}{n} \tag{756}$$

Then write n in terms of $\triangle t$ as

$$n = \frac{b - a}{\triangle t} \tag{757}$$

and if we substitute this expression on the formula for the average temperature we obtain

$$\text{approximation average } T = \frac{\left[T\left(t_1^*\right)\triangle t + \cdots + T\left(t_n^*\right)\triangle t\right]}{b-a} \tag{758}$$

Now, suppose that we plot T against t, that is, plot the values of T on the y axis and the values of t on the x axis. Then, with respect to this visual interpretation, each term $T\left(t_i^*\right)\triangle t$ corresponds precisely to the area of a small rectangle we used to define the definite integral!

Given that we expect the approximation to become better with smaller values of $\triangle t$ we make the following definition:

Suppose that $T(t)$ is a function defined on $[a, b]$. Then its **average value**, denoted $\langle T\rangle$, is by definition

$$\langle T\rangle \equiv \frac{1}{b-a}\int_a^b T(t)dt \tag{759}$$

Now, the Mean Value Theorem for Integrals says that if T is continuous, there is some instant of time t_a for which $T(t_a)$ is equal to the average value of the temperature.

Mean Value Theorem for Integral: Suppose that $T(t)$ is a continuous function on $[a, b]$. Then there at least one value $t_{ave} \in [a, b]$ for which

$$T(t_{ave}) = \langle T\rangle = \frac{1}{b-a}\int_a^b T(t)dt \tag{760}$$

Proof. We will use the limit definition of the definite integral. Suppose we divide the interval $[a, b]$ into subintervals whose endpoints are

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b \tag{761}$$

Then the integral is approximated by

$$\sum_{i=1}^{n} T\left(t_i^*\right)\triangle t_i \tag{762}$$

Because T is continuous on $[a, b]$ by the extreme value theorem it achieves an absolute maximum T_{max} and an absolute minimum T_{min}. Therefore, no matter what value we chose for t_i^* we have

$$T_{min} \leq T\left(t_i^*\right) \leq T_{max} \tag{763}$$

and the value for the sum 762 can be bounded by

$$T_{min}(b-a) = \sum_{i=1}^{n} T_{min}\triangle t_i \leq \sum_{i=1}^{n} T\left(t_i^*\right)\triangle t_i \leq \sum_{i=1}^{n} T_{max}\triangle t_i = T_{max}(b-a) \tag{764}$$

Dividing by $b - a$ we obtain

$$T_{min} \leq \frac{1}{b-a} \sum_{i=1}^{n} T\left(t_i^*\right) \triangle t_i \leq T_{max} \tag{765}$$

Taking the limit in which the length of the intervals goes to 0 we obtain

$$T_{min} \leq \frac{1}{b-a} \int_a^b T(t) dt \leq T_{max} \tag{766}$$

Now, $\frac{1}{b-a} \int_a^b T(t) dt$ is simply a number between the smallest and largest value achieved by T and by the intermediate value theorem there must exist a value t_{ave} for which

$$T\left(t_a\right) = \frac{1}{b-a} \int_a^b T(t) dt \tag{767}$$

and the result follows. $\square$

The Fundamental Theorem of Calculus

Suppose $y = f(x)$ is a function. We know that $\int_a^b f(x)dx$ is the area under the curve so it is just a number. Our objective is to find a function related in some way to the integral $\int_a^b f(x)dx$. A natural option is to define

$$A(t) = \int_a^t f(x)dx \tag{768}$$

We imagine, that at time $t = a$ we are at point a and then start walking from a to b. Then $A(t)$ gives the area under the curve covered in time t [51]. Clearly we have $A(a) = 0$ and $A(b) = \int_a^b f(x)dx$.

[51] For the animation click here

We want to find $\frac{dA}{dt}$, that is, the rate at which new area is being covered. By the definition of the derivative

$$A'(t) = \lim_{\triangle t \longrightarrow 0} \frac{A(t + \triangle t) - A(t)}{\triangle t} = \lim_{\triangle t \longrightarrow 0} \frac{\int_a^{t+\triangle t} f(x)dx - \int_a^t f(x)dx}{\triangle t} \tag{769}$$

Looking at the next figure, it is clear that $\int_a^{t+\triangle t} f(x)dx - \int_a^t f(x)dx = \int_t^{t+\triangle t} f(x)dx$

Therefore,

$$A'(t) = \lim_{\triangle t \longrightarrow 0} \frac{\int_t^{t+\triangle t} f(x)dx}{\triangle t} \tag{770}$$

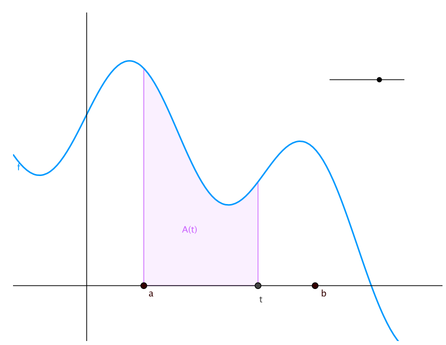


Figure 138: Definition $A(t)$

Now, if $\triangle t$ is *very* small, then $\int_t^{a+\triangle t} f(x)dx$ can be approximated by the area of the rectangle on top of the interval $[t, t + \triangle t]$ with height $f(t)$. This rectangle has area $f(t)\triangle t$ and given that in the limit we expect this approximation to be exact we have

$$A'(t) = \lim_{\triangle t \longrightarrow 0} \frac{f(t)\triangle t}{\triangle t} = f(t) \tag{771}$$

To make this argument more rigorous, observe that if f is continuous then by the Mean Value Theorem for Integrals there exists a time $t_{ave} \in [t, t + \triangle t]$ such that

$$f(t_{ave}) = \frac{\int_t^{t+\triangle t} f(x)dx}{\triangle t} \tag{772}$$

so

$$A'(t) = \lim_{\triangle t \longrightarrow 0} f(t_{ave}) = f(t) \tag{773}$$

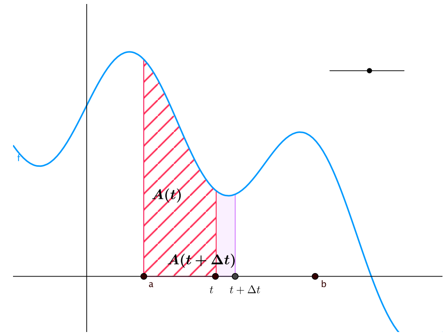


Figure 139: $A(t+\triangle t)-A(t)$

If we use Leibniz notation and the definition of $A'(t)$ then 771 becomes

$$\frac{d}{dt}\left[\int_a^t f(x)dx\right] = f(t) \tag{774}$$

Also, if we use 771 to write $f(x) = A'(x)$ and substitute it in 768 we get

$$A(t) = \int_a^t \frac{dA}{dx}dx \tag{775}$$

The last two equations are so important that they are called the Fundamental Theorem of Calculus , since it says that one is the process is the reverse of the other. [52]

[52] To see why they are the inverses of each other suppose that we start with a function $f(x)$. Call P the process of multiplying by 3, so $Pf(x) = 3f(x)$ and call Q the process of dividing by 3, so $Qf(x) = \frac{f(x)}{3}$. If we apply first P the way to cancel the effect of P and recover $f(x)$ is to apply Q so $QPf(x) = f(x)$. Likewise, if we apply first Q the way to cancel the effect of Q and recover $f(x)$ is to apply P so $PQf(x) = f(x)$.
Similarly, if we start with a function $f(x)$ and integrate it then 774 says that we must differentiate to recover $f(x)$. Likewise, if we start with a function $f(x)$ and differentiate it then 775 says that we must integrate to recover $f(x)$. It is in this sense that integration and differentiation are inverse processes.

Fundamental Theorem of Calculus:

- Suppose that $y = f(x)$ is a continuous function defined on an interval $[a, b]$. Define the "area function"

$$A(t) = \int_a^t f(x)dx \tag{776}$$

  Then

$$\frac{dA(t)}{dt} = f(t) \tag{777}$$

- Therefore, $A(t)$ is a function whose derivative is $f(t)$ and it is called an **antiderivative** of $f(t)$

We will show that the Fundamental Theorem of Calculus really deserves its name because it solves the problem of how to find integrals. First of all, we just saw $A(t)$ must be a function whose derivative is $f(t)$. A function with this property is called an antiderivative.

If $f(t)$ is a function then $F(t)$ is an antiderivative (or indefinite integral) of $f(t)$ if

$$F'(t) = f(t) \tag{778}$$

For example, $\frac{t^2}{2}$, $\frac{t^2}{2}+3$ and $\frac{t^2}{2}-2$ are all antiderivatives of $f(t) = t$.

Although we won't show this now, the only way two different functions can have the same derivative always is if they differ by some constant. [53] In particular, since 777 says that $A(t)$ is an antiderivative of $f(t)$ then any other antiderivative $F(t)$ of $f(t)$ must differ from $A(t)$ by a constant. That is, we must have

$$A(t) = F(t) + c \tag{779}$$

to find c we use the fact that $A(a) = 0$ so

$$0 = F(a) + c \tag{780}$$

[53] Basically, if $F(t)$ and $A(t)$ have the same derivative, that is, $F'(t) = A'(t)$ then $(F(t) - A(t))' = 0$ so $F(t) - A(t)$ is a function with vanishing derivative and we saw as a consequence of the mean value theorem that $F(t) - A(t)$ must be a constant function.

giving

$$A(t) = F(t) - F(a) \tag{781}$$

Substituting 781 in 776 and taking $t = b$ we have found the procedure to calculate integrals

Algorithm for finding integrals:

- Suppose that $y = f(x)$ is a continuous function and we want to find

$$\int_a^b f(x)dx \tag{782}$$

To do this

1. Find a antiderivative $F(x)$ of $f(x)$, that is, a function that satisfies

$$F'(x) = f(x) \tag{783}$$

2. Then

$$\int_a^b f(x)dx = F(b) - F(a) \tag{784}$$

We will use 784 to find

$$\int_1^2 xdx \tag{785}$$

We already know that the answer is $\frac{3}{2}$ but we will see how to obtain this result with the Fundamental Theorem of Calculus. Comparing with 784 we take $a = 1$, $b = 2$, $f(x) = x$. We want to find $F(x)$ such that

$$F'(x) = f(x) = x \tag{786}$$

The question now becomes, which function $F(x)$ has a derivative x? For example, $\frac{x^2}{2}$ is a function whose derivative is x, so we can take $F(x) = \frac{x^2}{2}$ and 784 gives

$$\int_1^2 xdx = F(2) - F(1) = \frac{2^2}{2} - \frac{1^2}{2} = 2 - \frac{1}{2} = \frac{3}{2} \tag{787}$$

which is the same result we found with the definition of the limit! However, it should be clear that this procedure is much more easier!

Now that we know how to compute integrals we can expand the geometrical interpretation of the integral. Suppose we want to find

$$\int_{-3}^3 xdx \tag{788}$$

We still have that $F(x) = \frac{x^2}{2}$ is an antiderivative for x so by 784 we have

$$\int_{-3}^3 xdx = F(3) - F(-3) = \frac{(3)^2}{2} - \frac{(-3)^2}{2} = 0 \tag{789}$$

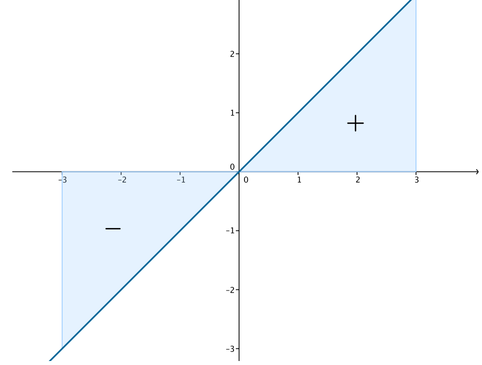


Figure 140: $\int_{-3}^{3} x dx = 0$

so the area under the curve is 0 in this case.

If we look at the graph of the function $y = x$ we can see that the "problem" lies in the fact that the area above the x axis is equal to the area under the x axis so the integral considered the area under the x axis as negative. When we discuss applications of the integrals, we will see that this property of the integral turns out to be of great advantage.

Geometrical Meaning of the Integral: $\int_a^b f(x)dx$ is the **signed or net area** under the curve of $y = f(x)$. If the curve is above the x axis the area is considered positive and if the curve is below the x axis the area is considered negative.

Finding Antiderivatives

We just saw that the problem of finding integrals is reduced to the problem of finding antiderivatives. In analogy to 776, we write $\int f(x)dx$ to denote the antiderivatives of $f(x)$

- If $f(x)$ is a function, all possible antiderivatives of $f(x)$ are represented by the symbol $\int f(x)dx$.
- For example, $\int x dx = \frac{x^2}{2} + c$ means that all functions whose derivative is x are of the form $\frac{x^2}{2} + c$ where c is an arbitrary constant

Given that we know the derivatives of many functions, it is possible to verify the following table

Table Indefinite Integrals:

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + c \qquad n \neq -1 \tag{790}$$

$$\int \frac{dx}{x} = \ln|x| + c \tag{791}$$

$$\int e^x dx = e^x + c \tag{792}$$

$$\begin{array}{cc} \int \sin x dx = -\cos x + c & \int \cos x dx = \sin x + c \\ \int \csc x \cot x dx = -\csc x + c & \int \csc^2 x dx = -\cot x + c \\ \int \sec^2 x dx = \tan x + c & \int \sec x \tan x dx = \sec x + c \\ \int \frac{1}{x^2+1} dx = \arctan x + c & \int \frac{1}{\sqrt{1-x^2}} = \arcsin x + c \end{array} \tag{793}$$

$$\int \sinh x dx = \cosh x + c \qquad \int \cosh x dx = \sinh x + c \tag{794}$$

The following properties of the integral follow from the properties of the derivative:

Properties of the Integral:

- The integral of a sum is the sum of the integral
$$\int (f(x)+g(x))\,dx = \int f(x)dx + \int g(x)dx \tag{795}$$
- Constants can be taken outside the integral
$$\int kf(x)dx = k\int f(x)dx \tag{796}$$
- Integral of a constant:
$$\int kdx = kx + c \tag{797}$$

Example 145. Find the indefinite integral $\int 3x^{-2/3}dx$

By the properties of the integral we can take the constant outside the integral so
$$\int 3x^{-2/3}dx = 3\int x^{-2/3}dx \tag{798}$$
Now we use the rule $\int x^n dx = \frac{x^{n+1}}{n+1} + C$ where C is an arbitrary constant. Taking $n = -2/3$ we have
$$\int x^{-2/3}dx = \frac{x^{-\frac{2}{3}+1}}{-\frac{2}{3}+1} + C = \frac{x^{\frac{1}{3}}}{\frac{1}{3}} + C = 3\sqrt[3]{x} + C \tag{799}$$
Therefore
$$\int 3x^{-2/3}dx = 3\int x^{-2/3}dx = 3\left(3\sqrt[3]{x} + C\right) = 9\sqrt[3]{x} + 3C \tag{800}$$
Now, $3C$ is a new constant so we can call it $C_1 = 3C$ if we want to.

Example 146. Find the indefinite integral $\int \left(x^2 + x + x^{-3}\right)dx$

We use the fact that the integral of a sum is the sum of the integrals
$$\int \left(x^2 + x + x^{-3}\right)dx = \int x^2 dx + \int x dx + \int x^{-3}dx \tag{801}$$
and the rule $\int x^n dx = \frac{x^{n+1}}{n+1} + C$ with $n = 2, 1, -3$ respectively.
$$\begin{gathered} \int x^2 dx = \frac{x^3}{3} + C_1 \\ \int x dx = \frac{x^2}{2} + C_2 \\ \int x^{-3}dx = \frac{x^{-2}}{-2} + C_3 \end{gathered} \tag{802}$$

where C_1, C_2, C_3 are constants. Therefore

$$\int \left(x^2 + x + x^{-3}\right) dx = \frac{x^3}{3} + \frac{x^2}{2} - \frac{1}{2x^2} + C_1 + C_2 + C_3 \tag{803}$$

and if we want we can call $C = C_1 + C_2 + C_3$ to represent the new constant.

Example 147. Find the indefinite integral $\int \frac{u^3+2u^2-u}{3u} du$

Factorizing u in the numerator

$$\frac{u^3 + 2u^2 - u}{3u} = \frac{u\left(u^2 + 2u - 1\right)}{3u} = \frac{1}{3}\left(u^2 + 2u - 1\right) \tag{804}$$

Therefore

$$\int \frac{u^3 + 2u^2 - u}{3u} du = \frac{1}{3}\int \left(u^2 + 2u - 1\right) du = \frac{1}{3}\left(\int u^2 du + 2\int u du - \int 1 du\right) \tag{805}$$

Now

$$\begin{aligned} \textstyle\int u^2 du &= \tfrac{u^3}{3} + C_1 \\ \textstyle\int u du &= \tfrac{u^2}{2} + C_2 \\ \textstyle\int 1 du &= u + C_3 \end{aligned} \tag{806}$$

Substituting in 805 we get

$$\frac{1}{3}\left(\frac{u^3}{3} + C_1 + u^2 + 2C_2 - u - C_3\right) = \frac{1}{3}\left(\frac{u^3}{3} + u^2 - u\right) + \frac{1}{3}\left(C_1 + 2C_2 - C_3\right) \tag{807}$$

we can call $C = \frac{1}{3}\left(C_1 + 2C_2 - C_3\right)$ as a name for the total constant.

There is another rule for finding antiderivatives which follows from the chain rule for derivatives. From the chain rule we know that

$$(f \circ g)'(x) = f'(g(x))g'(x) \tag{808}$$

This implies that

$$\int f'\left(g(x)\right) g'(x) dx = f\left(g(x)\right) + c \tag{809}$$

If we use the chain of variables

$$u = g(x) \tag{810}$$

then

$$du = \frac{du}{dx} dx = g'(x) dx \tag{811}$$

and substituting in 809 we end up with

$$\int f'(u) du = f(u) + c \tag{812}$$

which looks like a tautology under this notation.

Method of Substitution:

- To integrate

$$\int f'(g(x))\, g'(x) dx \tag{813}$$

we make the change of variables

$$u = g(x) \tag{814}$$

- We can write

$$du = g'(x) dx \tag{815}$$

so the integral becomes

$$\int f'(g(x))\, g'(x) dx = \int f'(u) du \tag{816}$$

Example 148. Find the indefinite integral of $\int \frac{4x}{(2x^2+3)^3} dx$

If we use the change of variables

$$u = 2x^2 + 3 \tag{817}$$

then

$$du = \frac{du}{dx} dx = 4x dx \tag{818}$$

We can use 818 to write

$$x dx = \frac{du}{4} \tag{819}$$

so

$$\int \frac{4x}{(2x^2+3)^3} dx = \int \frac{4}{u^3}\frac{du}{4} = \int u^{-3} du = \frac{u^{-2}}{-2} + C = -\frac{1}{2}\frac{1}{(2x^2+3)^2} + C \tag{820}$$

Example 149. Find the indefinite integral of $\int \frac{e^x}{1+e^x} dx$

We make the change of variable

$$u = 1 + e^x \tag{821}$$

then

$$du = \frac{du}{dx} dx = e^x dx \tag{822}$$

so

$$\int \frac{e^x}{1+e^x} dx = \int \frac{du}{u} = \ln u + C = \ln(1 + e^x) + C \tag{823}$$

Example 150. Find the indefinite integral of $\int \frac{\sqrt{\ln x}}{x} dx$
We make the change of variables

$$u = \ln x \tag{824}$$

then

$$du = \frac{du}{dx} dx = \frac{1}{x} dx \tag{825}$$

Therefore

$$\int \frac{\sqrt{\ln x}}{x} dx = \int u^{\frac{1}{2}} du = \frac{u^{\frac{3}{2}}}{\frac{3}{2}} + C = \frac{2}{3}\sqrt{\ln^3 x} + C \tag{826}$$

Example 151. Find the indefinite integral of $\int \left(xe^{-x^2} + \frac{e^x}{e^x+3} \right) dx$
First we separate the integral

$$\int \left(xe^{-x^2} + \frac{e^x}{e^x+3} \right) dx = \underbrace{\int xe^{-x^2} dx}_{(1)} + \underbrace{\int \frac{e^x}{e^x+3} dx}_{(2)} \tag{827}$$

To solve (1) we use the change of variables

$$u = -x^2 \tag{828}$$

Then

$$du = \frac{du}{dx} dx = -2x dx \tag{829}$$

which we write as

$$x dx = -\frac{du}{2} \tag{830}$$

Therefore (1) becomes

$$\int xe^{-x^2} dx = \int e^u \left(-\frac{du}{2} \right) = -\frac{1}{2} \int e^u du = -\frac{1}{2} \left(e^u + C_1 \right) = -\frac{1}{2} e^{-x^2} - \frac{C_1}{2} \tag{831}$$

To solve (2) we use the change of variables

$$v = e^x + 3 \tag{832}$$

then

$$dv = \frac{dv}{dx} dx = e^x dx \tag{833}$$

The integral (2) becomes

$$\int \frac{e^x}{e^x+3} dx = \int \frac{dv}{v} = \ln v + C_2 = \ln\left(e^x + 3\right) + C_2 \tag{834}$$

In this way

$$\int \left(xe^{-x^2} + \frac{e^x}{e^x+3} \right) dx = -\frac{1}{2} e^{-x^2} - \frac{C_1}{2} + \ln\left(e^x + 3\right) + C_2 \tag{835}$$

if we want we can call $C = -\frac{C_1}{2} + C_2$ as the total constant.

Example 152. Find the indefinite integral of $\int \frac{1-\sqrt{x}}{1+\sqrt{x}}dx$

We make the change of variable

$$u = 1+\sqrt{x} \tag{836}$$

then

$$du = \frac{du}{dx}dx = \frac{1}{2\sqrt{x}}dx \tag{837}$$

We can use the previous equation and the fact that $\sqrt{x} = u-1$ to write

$$dx = 2\sqrt{x}du = 2\left(u-1\right)du \tag{838}$$

In this way

$$\int \frac{1-\sqrt{x}}{1+\sqrt{x}}dx = \int \frac{1-(u-1)}{u}2(u-1)du = 2\int \frac{(2-u)(u-1)}{u}du \tag{839}$$

To integrate the last integral we expand the numerator and separate it

$$\frac{(2-u)(u-1)}{u} = \frac{3u-2-u^2}{u} = 3-\frac{2}{u}-u \tag{840}$$

Therefore

$$2\int \frac{(2-u)(u-1)}{u}du = 2\left(3\int du - 2\int \frac{du}{u} - \int udu\right) = 6u-4\ln u - u^2+C \tag{841}$$

Now we replace u in terms of x to obtain

$$\int \frac{1-\sqrt{x}}{1+\sqrt{x}}dx = 6\left(1+\sqrt{x}\right)-4\ln\left(1+\sqrt{x}\right)-\left(1+\sqrt{x}\right)^2+C \tag{842}$$

Example 153. Find the area of the region under the graph of the function $f(x) = 1-\sqrt[3]{x}$ on the interval $[-8,-1]$

The area is

$$A = \int_{-8}^{-1}\left(1-x^{\frac{1}{3}}\right)dx = \left(x-\frac{x^{\frac{4}{3}}}{\frac{4}{3}}\right)\Bigg|_{x=-8}^{x=-1} = \left(-1-\frac{3}{4}\right)-\left(-8-\frac{3}{4}(-8)^{\frac{4}{3}}\right) = \tag{843}$$

Observe that there is no need to include the constant of integration since it cancels out when it is evaluated at the upper limit minus the lower limit.

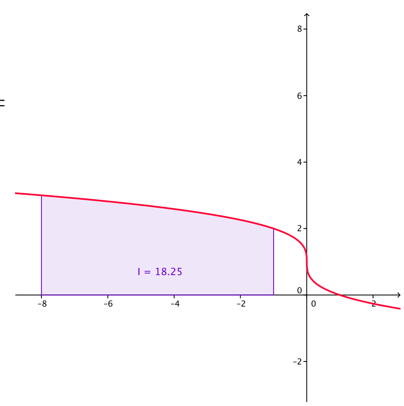


Figure 141: Area under the curve $y = 1-\sqrt[3]{x}$

Example 154. Evaluate $\int_0^1 \frac{dx}{\sqrt{2x+1}}$

We make the change of variables

$$u = 2x+1 \tag{844}$$

then

$$du = 2dx \tag{845}$$

so

$$dx = \frac{du}{2} \tag{846}$$

Since we make a change of variables we change the limits of integration:

$$\begin{aligned} &\text{if } x = 0 \text{ then } u = 1 \\ &\text{if } x = 1 \text{ then } u = 3 \end{aligned} \tag{847}$$

In this way

$$\int_0^1 \frac{dx}{\sqrt{2x+1}} = \int_1^3 \frac{du}{2\sqrt{u}} = \sqrt{u}\Big|_{u=1}^{u=3} = \sqrt{3} - \sqrt{1} = \sqrt{3} - 1 \tag{848}$$

Therefore, in a definite integral, if we make a change of variables there is no need to return to the original variable, instead we change the limits of integration.

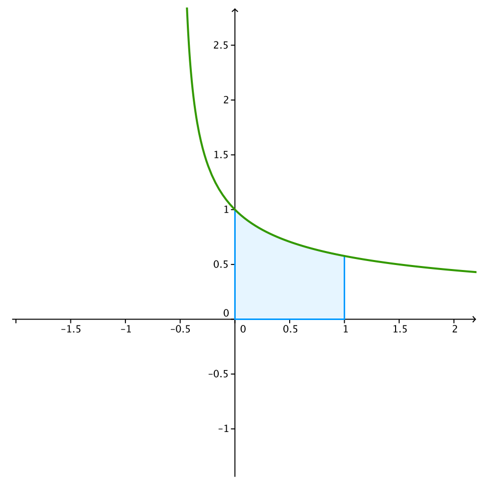


Figure 142: Area under the curve $y = \frac{1}{\sqrt{2x+1}}$

Example 155. Given that $\int_1^3 f(x)dx = 7$ and $\int_3^5 f(x)dx = 2$ find $\int_1^5 f(x)dx$

From the geometrical interpretation we can see that

$$\int_1^5 f(x)dx = \int_1^3 f(x)dx + \int_3^5 f(x)dx = 7 + 2 = 9 \tag{849}$$

When we discussed the exponential models we saw that they consisted in functions $Q(t)$ that satisfy the equation

$$Q'(t) = kQ(t) \tag{850}$$

we mentioned that $Q(t)$ must be of the form $Q_0 e^{kt}$ but we gave no explanation on how we could figure this out other than verification by "brute force". However, with the help of integrals it is possible to justify this. Equation 850 is equivalent to

$$\frac{1}{Q}\frac{dQ}{dt} = k \tag{851}$$

If we "multiply" both sides by dt this is the same as

$$\frac{dQ}{Q} = kdt \tag{852}$$

and integrating both sides of the equation gives

$$\int \frac{dQ}{Q} = \int kdt \tag{853}$$

Now,

$$\begin{aligned} &\int \tfrac{dQ}{Q} = \ln Q + c_1 \\ &\int kdt = kt + c_2 \end{aligned} \tag{854}$$

Therefore, 853 is the same as

$$\ln Q = kt + c_2 - c_1 \tag{855}$$

if we exponentiate both sides of the equation we end up with

$$Q(t) = e^{kt+c_2-c_1} = e^{c_2-c_1}e^{kt} \tag{856}$$

since $e^{c_2-c_1}$ is just a constant we can call it Q_0 so the function $Q(t)$ is

$$Q(t) = Q_0 e^{kt} \tag{857}$$

.In practice, to find the constant Q_0 we need to specify a desired value for Q at some time. These problems are called initial value problems as we will show in the next examples.

Example 156. Find $f(x)$ by solving the initial value problem $f'(x) = 1 + \frac{1}{x^2}$, $f(1) = 2$

If we write $f'(x) = \frac{df}{dx}$ then we need to solve $\frac{df}{dx} = 1 + \frac{1}{x^2}$ and to do so we can "multiply" by dx on both sides to obtain

$$df = \left(1 + \frac{1}{x^2}\right) dx \tag{858}$$

Integrating both sides gives

$$\int df = \int \left(1 + x^{-2}\right) dx \tag{859}$$

and by the basic rules of integration this is the same as

$$f = x - x^{-1} + C \tag{860}$$

where C represents all the constants of integration that show up. To find the value of C we use the initial condition $f(1) = 2$ which gives

$$2 = 1 - 1^{-1} + C \tag{861}$$

so

$$C = 2 \tag{862}$$

and the solution is

$$f(x) = x - \frac{1}{x} + 2 \tag{863}$$

Observe that we can always verify that the solution works by differentiating $\left(x - \frac{1}{x} + 2\right)'$ and observing that it gives $1 + \frac{1}{x^2}$. In the same way, we can verify the initial condition.

Example 157. Find $f(x)$ by solving the initial value problem $f'(x) = 1 + e^x + \frac{1}{x}$, $f(1) = 3 + e$

Again, $\frac{df}{dx} = 1 + e^x + \frac{1}{x}$ is equivalent to

$$df = \left(1 + e^x + \frac{1}{x}\right) dx \tag{864}$$

integrating both sides we have

$$\int df = \int \left(1 + e^x + \frac{1}{x}\right) dx \tag{865}$$

so we end up with

$$f = x + e^x + \ln x + C \tag{866}$$

Using $f(1) = 3 + e$ gives

$$3 + e = 1 + e + \ln 1 + C \tag{867}$$

so

$$C = 2 \tag{868}$$

and

$$f(x) = x + e^x + \ln x + 2 \tag{869}$$

Applications of the Integral

Suppose that we know the velocity of an object (for example, a car) as a function time, that is, we have a function $v(t)$. Given the velocity we may want to find the change in position of the car from time a to time b. We know that the if $v(t)$ is constant then

$$\text{change in position} = (\text{velocity}) \cdot (\text{time ellapsed}) \tag{870}$$

If $v(t)$ is not constant, we can divide the interval of time into small subintervals of time $\triangle t = \frac{b-a}{n}$. If the subintervals are small enough then we may assume that $v(t)$ is approximately constant so the change in position $\triangle x$ on each subinterval is

$$\triangle x \simeq v(t)\triangle t \tag{871}$$

the total change in position will be approximately equal to

$$x \simeq (v(a) + v(a + \triangle t) + \cdots + v(a + (n-2)\triangle t) + v(a + (n-1)\triangle t))\,\triangle t \tag{872}$$

we can see that this formula is basically the same as the formula for the definite integral, in fact, if we plot v as a function of t then the area under the curve represents the distance travelled

- If $v(t)$ represents the velocity of an object the change in position of the object from time a to time b is given by

$$\begin{gathered} x(b) - x(a) = \int_a^b v(t)dt \\ x(t) = \int v(t)dt \end{gathered} \tag{873}$$

- If $a(t)$ represents the acceleration of an object the change in velocity of the object from time a to time b is given by

$$\begin{gathered} v(b) - v(a) = \int_a^b a(t)dt \\ v(t) = \int a(t)dt \end{gathered} \tag{874}$$

Example 158. You are driving a magical Mini Cooper in a dream. The velocity of the car is given by $v(t) = \frac{2}{t^2} + 2e^t$ where t is measured in minutes after you started dreaming and $v(t)$ is measured in miles per minute. If you are two miles from a magical lamppost 1 minute after entering the dream and speeding directly away from it, how far are you from the lamppost t minutes after you started dreaming?

Call $x(t)$ the position measured from the lamppost. We have that $x(1) = 2$ and we want to find $x(t)$. Using 873 we have

$$x(t) - x(1) = \int_1^t 2\left(t^{-2} + e^t\right) dt = 2\left[-t^{-1} + e^t\right]\Big|_1^t = 2\left[-t^{-1} + e^t - (-1+e)\right] = 2\left(-\frac{1}{t} + e^t + 1 - e\right) \tag{875}$$

Therefore

$$x(t) = x(1) + 2\left(-\frac{1}{t} + e^t + 1 - e\right) = 2\left(-\frac{1}{t} + e^t + 2 - e\right) \tag{876}$$

Example 159. While you are studying for your calculus final your friends kidnap you to take you out for ice cream. They lock you into the trunk of a car to keep you from escaping, but fortunately you have an accelerometer and a watch with you. Your acceleration t hours after they start driving is $a(t) = 180 - 360t\ \frac{\text{miles}}{\text{hour}^2}$. The car stops after 1 hour. Assuming the car has been driving in a straight line, and the car is stationary when you start counting time, how far you have traveled?

Call $v(t)$ the velocity of the car after your friends kidnapped you. Because $a = \frac{dv}{dt}$ we can see this problem as an initial value problem. We have

$$v(t) = \int a dt = \int (180 - 360t)\, dt = 180t - 180t^2 + c_1 \tag{877}$$

We want $v(1) = 0$ so the value of the constant must be $c_1 = 0$.

$$v(t) = 180t - 180t^2 \tag{878}$$

To find the position traveled, we have that $v = \frac{dx}{dt}$ so

$$x(t) = \int v dt = \int \left(180t - 180t^2\right) dt = 90t^2 - 60t^3 + c_2 \tag{879}$$

to find c_2 we use that $x(0) = 0$ which gives $c_2 = 0$. Therefore, the distance travelled is

$$x(1) = 30 \text{ miles} \tag{880}$$

We have seen that geometrically the integral is the area under the curve of the function $y = f(x)$ and can be defined as the limit the sum of the areas of small rectangles. So in a sense we can say that

$$\text{area} = \int_a^b f(x)dx \simeq \text{sum of rectangles of area } f(x)\triangle x \tag{881}$$

If we interpret the function as the velocity $v(t)$ then the physical interpretation of the integral is

$$\text{change in position} = \int_a^b v(t)dt \simeq \text{sum of distance } v(t)\triangle t \text{ travelled} \tag{882}$$

If we interpret the function as the acceleration $a(t)$ then we have

$$\text{change in velocity} = \int_a^b a(t)dt \simeq \text{sum of changes of velocity } a(t)\triangle x \text{ experienced} \tag{883}$$

We can see that in all interpretations the integral is some form of "continuous" sum of infinitesimal contributions of the function times the change in the independent variable so if $q(u)$ is our quantity of interest we have

$$\int_a^b q(u)du \simeq \text{sum of contributions of magnitude } q(u)\triangle u \tag{884}$$

From this we can see that

$$\text{units of } \int_a^b q(u)du = [\text{units of } q][\text{units of } u] \tag{885}$$

Example 160. Find the volume of a sphere as $\int A(z)dz$ where $A(z)$ is a cross section of the sphere at height z

We can think of the volume of the sphere being made as the volume of "thick" circles.

Suppose that the sphere has radius R and we take a "thick" circle at a distance z from the center of the sphere and thickness $\triangle z$ (here $-R \leq z \leq R$). The radius of the circle depends on the distance z so we write it as $r(z)$. The volume of this circle is

$$\pi\left(r(z)\right)^2 \triangle z = \pi\left(R^2 - z^2\right)\triangle z \tag{886}$$

where we used Pythagora's theorem in the last equation. Therefore the volume is

$$V = \int_{-R}^{R} \pi\left(R^2 - z^2\right) dz = \frac{4}{3}\pi R^3 \tag{887}$$

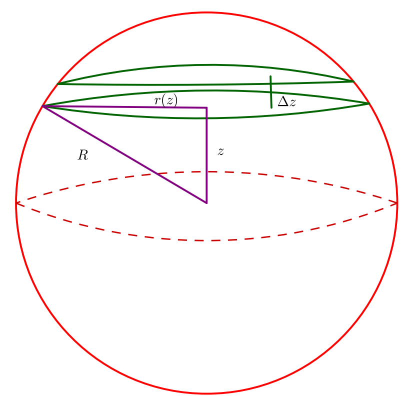


Figure 143: Volume of the sphere

Part VI

Other Integration Techniques

Integration by Parts

It may be clear by now, but finding integrals is much more difficult than finding derivatives. In fact, most of the rules we have used to find integrals have been obtained from the corresponding rules for derivatives. For example, the rule for the integral of a sum

$$\int f(x) + g(x)dx = \int f(x)dx + \int g(x)dx \tag{888}$$

is obtained from the rule

$$\frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx} + \frac{dg}{dx} \tag{889}$$

Similarly, the substitution rule

$$\int f'(g(x))g'(x)dx = f(g(x)) + C \tag{890}$$

is a consequence of the chain rule

$$\frac{d}{dx}f(g(x)) = f'(g(x))g'(x) \tag{891}$$

However, so far we have not used any rule for integrals that mirrors the product rule for derivatives. It should be clear first of all that

$$\int f(x)g(x)dx \neq \left(\int f(x)dx\right)\left(\int g(x)dx\right) \tag{892}$$

This can be seen if we take $f(x) = x$ and $g(x) = x^2$:

$$\begin{gathered}\int f(x)g(x)dx = \int x^3 dx = \tfrac{x^4}{4} + C \\ \left(\int f(x)dx\right)\left(\int g(x)dx\right) = \left(\int x dx\right)\left(\int x^2 dx\right) = \left(\tfrac{x^2}{2} + C_1\right)\left(\tfrac{x^3}{3} + C_2\right)\end{gathered} \tag{893}$$

and in fact it mirrors the corresponding behavior for derivatives

$$\frac{d}{dx}(f(x)g(x)) \neq \frac{df}{dx}\frac{dg}{dx} \tag{894}$$

Therefore, we should start with the correct product rule for derivatives

$$\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}g(x) + f(x)\frac{dg}{dx} \tag{895}$$

If we integrate both sides of the equation we obtain

$$f(x)g(x) + C = \int f'(x)g(x)dx + \int f(x)g'(x)dx \qquad (896)$$

Since the antiderivative notation $\int$ has an implicit constant we typically don't write the constant on the left hand side [54] and so the **integration by parts rule** becomes

[54] However, as we will mention after the examples this means one needs to be careful when using the antiderivative notation to avoid running into paradoxes

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx \qquad (897)$$

Typically, we write $u = f(x)$ and $dv = g'(x)dx$ so that the previous rule becomes

$$\int udv = uv - \int vdu \qquad (898)$$

To remember the rule, one can make an "integration by parts box"

u	v
du	dv

so that the right hand side of 898 can be obtained from the box by drawing the number "7" as the next examples show[55].

[55] Another mnemonic is $uv - \int vdu$ $\iff$ "ultra-violet voodoo"

Example 161. Find $\int \ln x dx$

Use integration by parts with

$u = \ln x$	$v = x$
$du = \frac{1}{x}dx$	$dv = dx$

$$\int \ln x dx = x \ln x - \int \frac{x}{x} dx = x \ln x - x + C \qquad (899)$$

Example 162. Find $\int \frac{\ln x}{\sqrt{x}} dx$

Use integration by parts with

$u = \ln x$	$v = 2\sqrt{x}$
$du = \frac{1}{x}dx$	$dv = \frac{1}{\sqrt{x}}dx$

$$\int \frac{\ln x}{\sqrt{x}} dx = 2\sqrt{x} \ln x - \int \frac{2\sqrt{x}}{x} dx = 2\sqrt{x} \ln x - 4\sqrt{x} + C \qquad (900)$$

Example 163. Evaluate the integral $\int_0^1 \frac{r^3}{\sqrt{16+r^2}} dr$

First we find the indefinite integral $\int \frac{r^3}{\sqrt{16+r^2}} dr$. Observe that

$$\frac{r^3}{\sqrt{16+r^2}} = r^2 \left(\frac{r}{\sqrt{16+r^2}} \right) \qquad (901)$$

and so we use integration by parts using the substitutions

$u = r^2$	$v = \sqrt{16+r^2}$
$du = 2rdr$	$dv = \frac{r}{\sqrt{16+r^2}}dr$

The integration by parts formula gives

$$\begin{aligned} \int \frac{r^3}{\sqrt{16+r^2}}dr &= r^2\sqrt{16+r^2} - \int 2r\sqrt{16+r^2}dr \\ &= r^2\sqrt{16+r^2} - \tfrac{2}{3}(r^2+16)^{3/2} + C \end{aligned} \tag{902}$$

Therefore

$$\begin{aligned} \int_0^1 \frac{r^3}{\sqrt{16+r^2}}dr &= \left(r^2\sqrt{16+r^2} - \tfrac{2}{3}(r^2+16)^{3/2} + C\right)\Big|_{r=0}^1 \\ &= \left(\sqrt{17} - \tfrac{2}{3}(17)^{3/2} + \tfrac{2}{3}(16)^{3/2}\right) \end{aligned} \tag{903}$$

Example 164. Find $\int e^{2x}\sin x dx$

We use integration by parts with

$u = e^{2x}$	$v = -\cos x$
$du = 2e^{2x}dx$	$dv = \sin x dx$

$$\int e^{2x}\sin x dx = -e^{2x}\cos x + \int 2e^{2x}\cos x dx \tag{904}$$

To find $\int 2e^{2x}\cos x dx$ we use integration by parts again, now with

$u = 2e^{2x}$	$v = \sin x$
$du = 4e^{2x}dx$	$dv = \cos x dx$

$$\int 2e^{2x}\cos x dx = 2e^{2x}\sin x - \int 4e^{2x}\sin x dx \tag{905}$$

and substituting this formula in 904 we find that

$$\int e^{2x}\sin x dx = -e^{2x}\cos x + 2e^{2x}\sin x - 4\int e^{2x}\sin x dx \tag{906}$$

Combining the colored terms we conclude that

$$\int e^{2x}\sin x dx = \frac{1}{5}\left(-e^{2x}\cos x + 2e^{2x}\sin x\right) + C \tag{907}$$

Example 165. Find $\int \sec^3\theta d\theta$ using that $\int \sec\theta d\theta = \ln|\sec\theta + \tan\theta| + C$

Notice that $\sec^3\theta = \sec\theta\sec^2\theta$ and so we can use integration by parts in the following way

$u = \sec\theta$	$v = \tan\theta$
$du = \sec\theta\tan\theta d\theta$	$dv = \sec^2\theta d\theta$

$$\int \sec^3\theta d\theta = \sec\theta\tan\theta - \int \sec\theta\tan^2\theta d\theta \tag{908}$$

To find $\int \sec\theta \tan^2\theta d\theta$, recall the identity $\tan^2\theta = \sec^2\theta - 1$ (which is obtained from $\sin^2\theta + \cos^2\theta = 1$ dividing by $\cos^2\theta$) so that

$$\int \sec\theta \tan^2\theta d\theta = \int \sec\theta(\sec^2\theta - 1)d\theta = \int \sec^3\theta - \int \sec\theta d\theta \quad (909)$$

Substituting in equation 908 we obtain

$$\int \sec^3\theta d\theta = \sec\theta\tan\theta - \int \sec^3\theta + \int \sec\theta d\theta \quad (910)$$

Using the formula at the beginning of the problem for $\int \sec\theta d\theta$ we conclude that

$$\int \sec^3\theta = \frac{1}{2}\left(\sec\theta\tan\theta + \ln|\sec\theta + \tan\theta|\right) + C \quad (911)$$

Example 166. Find $\int x^7(x^4+1)^{2/3}dx$

Notice that the integrand can be written as $x^7(x^4+1)^{2/3} = x^4\left(x^3(x^4+1)^{2/3}\right)$. We use integration by parts with

$u = x^4$	$v = \frac{3}{20}(x^4+1)^{5/3}$
$du = 4x^3dx$	$dv = x^3(x^4+1)^{2/3}dx$

$$\begin{aligned}\int x^7(x^4+1)^{2/3}dx &= \tfrac{3}{20}x^4(x^4+1)^{5/3} - \int \tfrac{3}{5}x^3(x^4+1)^{5/3} \\ &= \tfrac{3}{20}x^4(x^4+1)^{5/3} - \tfrac{9}{160}(x^4+1)^{8/3} + C\end{aligned} \quad (912)$$

We end up this section by pointing out the subtleties of the antiderivative $\int$ notation. For example, we know that

$$\int \frac{dx}{x} = \ln|x| + C \quad (913)$$

Suppose that someone wanted to find the left hand side using integration by parts with

$u = \frac{1}{x}$	$v = x$
$du = -\frac{1}{x^2}$	$dv = dx$

We would obtain that

$$\int \frac{dx}{x} = 1 + \int \frac{dx}{x} \quad (914)$$

which seems to imply that $0 = 1$! However, this paradoxical conclusion was obtained by being careless about the meaning of the antiderivative. The only way in which we could obtain that $0 = 1$ from the previous equation is if we subtract $\int \frac{dx}{x}$ from both sides of equation 914 . However, in that case we obtain by the antiderivative rules that

$$\int 0dx = 1 \quad (915)$$

which is fine because $\int 0dx$ consists of all functions which have derivative zero, so instead of $\int 0dx = 0$, we have in fact that $\int 0dx = C$, where C is any constant, not necessarily 0. Therefore, this integration by parts paradox is solve by realizing that the correct **cancellation property for antiderivatives** is

$$\int f(x)dx = g(x) + \int f(x)dx \implies C = g(x) \tag{916}$$

Tabular Integration

A more interactive way to use integration by parts is the method of **tabular integration .** It is easier to understand it through some examples, but the basic idea is to write three columns as the following diagram shows:

(alt) +/-	(diff). u	(int). dv	
+	$u_1 \searrow$	v_0	the top row contains the original integrand
-	$u_2 \searrow$	v_1	is $-\int u_2 v_1$ manageable? If so, stop. If not, continue
$\vdots$			
$(-1)^{n-1}$	$u_n \searrow$	v_{n-1}	
$(-1)^n$	$u_{n+1} \longrightarrow$	v_n	the bottom row represents $(-1)^n \int v_n u_{n+1}$

- In the first column one alternates between $+$ and $-$ signs, starting always with a $+$ sign.
- In the second column one differentiates a row and writes the result in the row below.
- In the third column one integrates a row and writes the result in the row below.
- The top row always contains the original integrand.
- If the product on the second row (with the sign of the first column) $-\int u_2 v_1$ is manageable, that is, if we know how to integrate it, then the process stops and the answer is

$$\int u_1 v_0 = u_1 v_1 - \int u_2 v_1 \tag{917}$$

  If it is not manageable, we iterate the process and apply the algorithm again.

Example 167. Find $\int \ln x dx$ using tabular integration.

(alt) +/-	(diff). u	(int). dv	
+	$\ln x$	1	the top row contains the original integrand
-	$\frac{1}{x}$	x	the product $-\int \frac{1}{x} \cdot x dx = -\int dx = -(x + C)$ is manageable

Since we only needed two rows the answer is

$$\int \ln x dx = +(x \ln x) - (x + C) \tag{918}$$

which is what we found before.

Example 168. Find $\int (x^2 - 3x) \sin x dx$ using tabular integration

(alt) +/-	(diff). u	(int). dv	
+	$x^2 - 3x$	$\sin x$	the top row contains the original integrand
-	$2x - 3$	$-\cos x$	$\int (2x-3)\cos x dx$ is not known, so continue
+	2	$-\sin x$	$\int 2(-\sin x) dx = 2\cos x + C$ so stop here

The answer is then

$$\int (x^2 - 3x) \sin x dx = +(x^2 - 3x)(-\cos x) - (2x-3)(-\sin x) + (2\cos x + C) \tag{919}$$

Example 169. Find $\int e^{3x} \sin 2x dx$ using tabular integration

(alt) +/-	(diff). u	(int). dv	
+	e^{3x}	$\sin 2x$	the top row contains the original integrand
-	$3e^{3x}$	$-\frac{1}{2}\cos 2x$	$+\frac{3}{2}\int e^{3x}\cos 2x dx$, try another step
+	$9e^{3x}$	$-\frac{1}{4}\sin 2x$	$-\frac{9}{4}\int e^{3x}\sin 2x dx$, a copy of the original integral

Therefore

$$\int e^{3x} \sin 2x dx = +e^{3x}\left(-\frac{1}{2}\cos 2x\right) - (3e^{3x})\left(-\frac{1}{4}\sin 2x\right) + \int 9e^{3x}\left(-\frac{1}{4}\sin 2x dx\right) \tag{920}$$

and solving for $\int e^{3x} \sin 2x dx$ we find that

$$\int e^{3x} \sin 2x dx = \frac{4}{13}\left(-\frac{e^{3x}}{2}\cos 2x + \frac{3}{4}e^{3x}\sin 2x\right) + C \tag{921}$$

Example 170. Find $\int \sin 2x \cos 5x dx$ using tabular integration

(alt) +/-	(diff). u	(int). dv	
+	$\sin 2x$	$\cos 5x$	the top row contains the original integrand
-	$2\cos 2x$	$\frac{1}{5}\sin 5x$	$-\int (2\cos 2x)\left(\frac{1}{5}\sin 5x\right) dx$, try another step
+	$-4\sin 2x$	$-\frac{1}{25}\cos 5x$	$+\frac{4}{25}\int \sin 2x \cos 5x dx$ a copy of the original integral

Therefore

$$\int \sin 2x \cos 5x dx = +(\sin 2x)\left(\frac{1}{5}\sin 5x\right) - (2\cos 2x)\left(-\frac{1}{25}\cos 5x\right) + \left(\frac{4}{25}\int \sin 2x \cos 5x dx\right) \tag{922}$$

so if we solve for $\int \sin 2x \cos 5x dx$ we find

$$\int \sin 2x \cos 5x dx = \frac{25}{21}\left(\frac{1}{5}\sin 2x \sin 5x + \frac{2}{25}\cos 2x \cos 5x\right) + C \tag{923}$$

Example 171. Find $\int (3x^2 - x)\ln^2 x dx$ using tabular integration

This example is more interesting because we will use tabular integration twice.

(alt) +/-	(diff). u	(int). dv	
+	$\ln^2 x$	$3x^2 - x$	the top row contains the original integrand
-	$\frac{2}{x}\ln x$	$x^3 - \frac{x^2}{2}$	$-\int (2x^2 - x)\ln x dx$

Hence

$$\int (3x^2 - x)\ln^2 x dx = \left(x^3 - \frac{x^2}{2}\right)\ln^2 x - \int (2x^2 - x)\ln x dx \qquad (924)$$

Now we evaluate $\int (2x^2 - x)\ln x dx$ in a separate table

(alt) +/-	(diff). u	(int). dv	
+	$\ln x$	$2x^2 - x$	the top row contains the original integrand
-	$\frac{1}{x}$	$\frac{2x^3}{3} - \frac{x^2}{2}$	$\int \left(\frac{x}{2} - \frac{2x^2}{3}\right) dx = \frac{x^2}{4} - \frac{2x^3}{9} + C$

Therefore we find that

$$\int (2x^2 - x)\ln x dx = \left(\frac{2x^3}{3} - \frac{x^2}{2}\right)\ln x + \frac{x^2}{4} - \frac{2x^3}{9} + C \qquad (925)$$

In this way we obtain

$$\int (3x^2 - x)\ln^2 x dx = \left(x^3 - \frac{x^2}{2}\right)\ln^2 x + \left(\frac{x^2}{2} - \frac{2x^3}{3}\right)\ln x + \frac{2x^3}{9} - \frac{x^2}{4} + C \qquad (926)$$

Combining Different Techniques

The following examples use the integration by parts technique and/or the substitution rule. It should be pointed out that there is no algorithmic method to decide how to start solving the following problems: one needs to do lots of examples to learn the hidden patterns behind each of the solutions.

Example 172. Find $\int \cos\sqrt{x}dx$

We make first the substitution

$$\begin{array}{ccc} u = \sqrt{x} & & \\ \Downarrow & & \\ u^2 = x & \Longrightarrow & 2udu = dx \end{array} \tag{927}$$

so that

$$\begin{aligned} \int \cos\sqrt{x}dx &= \int (\cos u)\,(2udu) \\ &= 2\int u\cos udu \\ &= 2u\sin u + 2\cos u + C \end{aligned} \tag{928}$$

where we used integration by parts to find $\int u\cos udu$. In this way

$$\int \cos\sqrt{x}dx = 2\sqrt{x}\sin\sqrt{x} + 2\cos\sqrt{x} + C \tag{929}$$

Example 173. Find $\int \ln^2 xdx$

Use integration by parts with

$u = \ln x$	$v = x\ln x - x$
$du = \frac{1}{x}dx$	$dv = \ln xdx$

In this way

$$\begin{aligned} \int \ln^2 xdx &= (\ln x)(x\ln x - x) - \int \tfrac{1}{x}(x\ln x - x)dx \\ &= x\ln^2 x - x\ln x - \int(\ln x - 1)dx \\ &= x\ln^2 x - x\ln x - (x\ln x - x - x) + C \\ &= x\ln^2 x - 2x\ln x + 2x + C \end{aligned} \tag{930}$$

Example 174. Find $\int e^{\sqrt{x}} dx$

We make the substitution

$$\begin{array}{c} u = \sqrt{x} \\ \Downarrow \\ u^2 = x \quad \Longrightarrow \quad 2udu = dx \end{array} \tag{931}$$

in this way

$$\begin{aligned} \int e^{\sqrt{x}} dx &= \int e^u 2udu \\ &= 2\int e^u udu \\ &= 2ue^u - 2e^u + C \end{aligned} \tag{932}$$

where we used integration by parts to find $\int e^u du$. In this way

$$\int e^{\sqrt{x}} dx = 2\sqrt{x}e^{\sqrt{x}} - 2e^{\sqrt{x}} + C \tag{933}$$

Example 175. Find $\int \cos(\ln x) dx$

We make the substitution

$$\begin{array}{c} u = \ln x \\ \Downarrow \\ e^u = x \quad \Longrightarrow \quad e^u du = dx \end{array} \tag{934}$$

In this way the integral becomes

$$\begin{aligned} \int \cos(\ln x) dx &= \int (\cos u)\, e^u du \\ &= \tfrac{1}{2}(e^u \sin u + e^u \cos u) + C \end{aligned} \tag{935}$$

where we used integration by parts (twice) to find $\int (\cos u)\, e^u du$. Therefore

$$\int \cos(\ln x) dx = \frac{1}{2}(x \sin(\ln x) + x \cos(\ln x)) + C \tag{936}$$

Example 176. Find $\int \frac{12dx}{3x+x\sqrt{x}}$

We make the following substitution

$$\begin{array}{c} u = \sqrt{x} \\ \Downarrow \\ u^2 = x \quad \Longrightarrow \quad 2udu = dx \end{array} \tag{937}$$

The integral becomes

$$\begin{aligned}
\int \frac{12dx}{3x+x\sqrt{x}} &= \int \frac{24udu}{3u^2+u^3} \\
&= 24\int \frac{du}{3u+u^2} \\
&= 24\int \frac{du}{u(3+u)} \\
&= 24\int \left(\frac{1}{3u} - \frac{1}{3(3+u)}\right) du \\
&= 8\left(\ln|u| - \ln|3+u|\right) + C \\
&= 8(\ln\sqrt{x} - \ln(3+\sqrt{x}) + C
\end{aligned} \tag{938}$$

Improper Integrals

So far we have only calculated integrals $\int_a^b f(x)dx$ where the interval $[a, b]$ is of finite length. However, sometimes it is useful to consider integrals where the interval of integration is infinite, like $\int_1^\infty \frac{1}{x^2}dx$, $\int_{-\infty}^2 e^x dx$ or $\int_{-\infty}^\infty x dx$. Such integrals are known as **improper integrals**.

Such integrals occur for example in Physics when one is dealing with the gravitational or electric force, which are long-range forces, that is, their effect extends throughout all space, albeit it becomes weaker as one moves farther away from the source. As we will see later in the course, integrals with infinite intervals of integration also happen in Probability Theory. Finally, we can also use these improper integrals to extend our notions of length, area and volume, albeit sometimes these extensions seem to have paradoxical consequences as we will soon see.

First we consider the improper integral $\int_1^\infty \frac{1}{x^2}dx$. At this point it should be pointed out that so far the symbol "$\int_1^\infty \frac{1}{x^2}dx$" is meaningless, since we haven't defined what we mean by a definite integral over an infinite interval. However, in Mathematics we try to avoid reinventing the wheel as much as possible, so we would like to define $\int_1^\infty \frac{1}{x^2}dx$ in such a way that uses the integrals that we know how to compute as much as possible.

A pragmatic approach one might use is the following. In practice, when we have tried to find a definite integral like $\int_a^b f(x)dx$, we have found first the antiderivative (or indefinite integral) $\int f(x)dx$ and then evaluated the function(s) we get at the endpoints. Similarly, to find $\int_1^\infty \frac{dx}{x^2}$ one could try to find first $\int \frac{dx}{x^2}$, which is the same as

$$\int \frac{dx}{x^2} = -\frac{1}{x} + C \tag{939}$$

where C is a constant. Now we would like to say that

$$\int_1^\infty \frac{dx}{x^2} = \left(-\frac{1}{x} + C\right)\Bigg|_{x=1}^{x=\infty} \tag{940}$$

and so the entire problem has been reduced to deciding what we mean by "$x = \infty$". However, at this point it should not be surprising that we will use the limit notion to define what we mean by "$x = \infty$", in fact, we take it to mean the following:

$$\int_1^\infty \frac{dx}{x^2} = \lim_{b\to\infty} \left(-\frac{1}{x} + C\right)\Bigg|_{x=1}^{x=b} \tag{941}$$

That is, we just evaluate our expression at $x = b$, and see what happens as x approaches ∞. In this way the value of $\int_1^\infty \frac{dx}{x^2}$ is

$$\begin{aligned} \int_1^\infty \frac{dx}{x^2} &= \lim_{b\to\infty} \left(-\tfrac{1}{x} + C\right)\Big|_{x=1}^{x=b} \\ &= \lim_{b\to\infty} \left(\left(-\tfrac{1}{b} + C\right) - (-1 + C)\right) \\ &= \lim_{b\to\infty} \left(1 - \tfrac{1}{b}\right) \\ &= 1 \end{aligned} \tag{942}$$

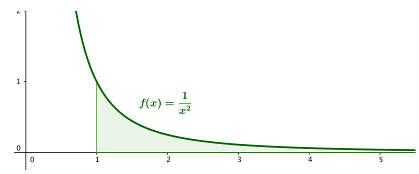


Figure 144: Geometric Interpretation of $\int_1^\infty \frac{dx}{x^2} = 1$

Therefore, we just found that $\int_1^\infty \frac{dx}{x^2} = 1$!

If we want, we can give a geometric interpretation to the previous calculation by saying that the "area" under the curve $y = \frac{1}{x^2}$ starting at $x = 1$ (and without end) is 1. Again, it is important to notice that since this region is unbounded there is no pre-existing notion of an area that we can use to compare this result with. Rather, the idea is that we can take $\int_1^\infty \frac{dx}{x^2}$ as our *definition* of what we mean by the area of this region.

Our pragmatic approach secretly used the Fundamental Theorem of Calculus, which relates the indefinite integral $\int f(x)dx$ to the definite integral $\int_a^b f(x)dx$. Typically the improper integral $\int_1^\infty \frac{dx}{x^2}$ is defined in a way that does not appeal directly to the fundamental theorem, although in practice we end up using it to compute the value of $\int_1^\infty \frac{dx}{x^2}$. The idea is to define $\int_1^\infty \frac{dx}{x^2}$ as the limit of definite integrals $\int_1^b \frac{dx}{x^2}$, that is,

$$\boxed{\int_1^\infty \frac{dx}{x^2} = \lim_{b\to\infty} \int_1^b \frac{dx}{x^2}} \tag{943}$$

Again, thanks to the fundamental theorem of Calculus, the right hand side is the same as $\lim_{b\to\infty} \left(-\frac{1}{x} + C\right)\Big|_{x=1}^{x=\infty}$, which is what we used in our computations. However, from a theoretical point of view, the definition 943 is more pleasing because it can be given even if no one had ever found out the fundamental theorem.

Definition of the Improper Integral $\int_a^\infty f(x)dx$ or $\int_{-\infty}^b f(x)dx$:

- **Definition of $\int_a^\infty f(x)dx$:** suppose that $\int_a^b f(x)dx$ exists for every number $b \geq a$. Then we define

$$\int_a^\infty f(x)dx \equiv \lim_{b\to\infty} \int_a^b f(x)dx \tag{944}$$

- **Definition of $\int_{-\infty}^b f(x)dx$:** suppose that $\int_a^b f(x)dx$ exists for every number $a \leq b$. Then we define

$$\int_{-\infty}^b f(x)dx \equiv \lim_{a\to-\infty} \int_a^b f(x)dx \tag{945}$$

The improper integrals $\int_a^\infty f(x)dx$ and $\int_{-\infty}^b f(x)dx$ are called **convergent** if the corresponding limit exists and are called **divergent** if the limit does not exist.

Example 177. Find $\int_{-\infty}^2 e^x dx$

Using the definition 945 we find $\int_{-\infty}^2 e^x dx$ as

$$\begin{aligned} \int_{-\infty}^2 e^x dx &= \lim_{a\to-\infty} \int_a^2 e^x dx \\ &= \lim_{a\to-\infty} e^x|_{x=a}^{x=2} \\ &= \lim_{a\to-\infty}(e^2 - e^a) \\ &= e^2 \end{aligned} \tag{946}$$

Example 178. Show that $\int_1^\infty \frac{dx}{x^p}$ is convergent if $p > 1$ and divergent if $0 < p \leq 1$

Again we use the definition 944. We need to make two cases: $p = 1$ and $p \neq 1$. If $p = 1$ then

$$\begin{aligned} \int_1^\infty \frac{dx}{x} &= \lim_{b\to\infty} \int_1^b \frac{dx}{x} \\ &= \lim_{b\to\infty} \ln x|_{x=1}^{x=b} \\ &= \lim_{b\to\infty} \ln b \\ &= \infty \end{aligned} \tag{947}$$

so the improper integral is divergent. If $p \neq 1$

$$\begin{aligned}\int_1^\infty \frac{dx}{x^p} &= \lim_{b\to\infty}\int_1^b \frac{dx}{x^p} \\ &= \lim_{b\to\infty} \left.\frac{x^{-p+1}}{-p+1}\right|_{x=1}^{x=b} \\ &= \lim_{b\to\infty}\left(\frac{b^{1-p}}{1-p}-\frac{1}{1-p}\right)\end{aligned} \tag{948}$$

To analyze this limit we make two cases.

If $0<p<1$ then $1-p$ is positive and so $\lim_{b\to\infty} b^{1-p}=\infty$, which means that $\int_1^\infty \frac{dx}{x^p}$ is divergent.

If $p>1$ then $1-p$ is negative and so $\lim_{b\to\infty} b^{1-p}=0$, which means that $\int_1^\infty \frac{dx}{x^p}=-\frac{1}{1-p}=\frac{1}{p-1}$.

These calculations are so important that we summarize them as follows:

Improper p integral: if $p\geq 0$ then $\int_1^\infty \frac{dx}{x^p}$ is convergent for $p>1$ and divergent otherwise:

$$\int_1^\infty \frac{dx}{x^p}=\begin{cases}\frac{1}{p-1} & \text{if } p>1\\ \infty & \text{if } 0\leq p<1\end{cases} \tag{949}$$

Example 179. Show that $\int_0^\infty \lambda e^{-\lambda x}dx = 1$ where $\lambda > 0$. This is a property that the function $f(x) = \lambda e^{-\lambda x}$ has to fulfill in order to become a probability density function, as we will see later. Then find the "expected value" of x, $\int_0^\infty \lambda x e^{-\lambda x}dx$.

The first integral is easy to compute directly

$$\begin{aligned}\int_0^\infty \lambda e^{-\lambda x}dx &= \lambda\lim_{b\to\infty}\int_0^b e^{-\lambda x}dx \\ &= \lambda\lim_{b\to\infty}\left.\frac{e^{-\lambda x}}{-\lambda}\right|_{x=0}^{x=b} \\ &= -\lim_{b\to\infty}\left(e^{-\lambda b}-1\right) \\ &= 1\end{aligned} \tag{950}$$

To find $\int_0^\infty \lambda x e^{-\lambda x}dx$ we find first $\int \lambda x e^{-\lambda x}dx$ using integration by parts.

$u=x$	$v=-e^{-\lambda x}$
$du=dx$	$dv=\lambda e^{-\lambda x}dx$

Therefore

$$\int x\lambda e^{-\lambda x}dx = -xe^{-\lambda x}+\int e^{-\lambda x}dx = -\frac{x}{e^{\lambda x}}-\frac{e^{-\lambda x}}{\lambda} \tag{951}$$

Therefore

$$\begin{aligned}
\int_0^\infty \lambda x e^{-\lambda x} dx &= - \lim_{b\to\infty} \left(\frac{x}{e^{\lambda x}} + \frac{e^{-\lambda x}}{\lambda} \right) \Big|_{x=0}^{x=b} \\
&= -\lim_{b\to\infty} \left(\frac{b}{e^{\lambda b}} + \frac{1}{\lambda e^{\lambda b}} - \frac{1}{\lambda} \right) \\
&= \frac{1}{\lambda}
\end{aligned} \tag{952}$$

Now we must discuss how to compute improper integrals like $\int_{-\infty}^{\infty} x dx$. There are actually two possibilities on how one might define this integrals. Both are perfectly acceptable, but we will write first the one we will use as our definition.

Definition of the improper integral $\int_{-\infty}^{\infty} f(x) dx$: to define it choose a real number c and compute separately the improper integrals $\int_{-\infty}^{c} f(x) dx$ and $\int_{c}^{\infty} f(x) dx$. If any of those two integrals are divergent, we say that $\int_{-\infty}^{\infty} f(x) dx$ is divergent. Otherwise, we define $\int_{-\infty}^{\infty} f(x) dx$ as

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{c} f(x) dx + \int_{c}^{\infty} f(x) dx \tag{953}$$

Using this definition we will show that $\int_{-\infty}^{\infty} x dx$ is divergent. To do this choose a real number, for example, $c = 2$. Then we have to determine if the improper integrals $\int_{-\infty}^{2} x dx$ and $\int_{2}^{\infty} x dx$ converge or diverge. It is not difficult to see that each of those integrals diverge so we will say that $\int_{-\infty}^{\infty} x dx$ is divergent.

There is an alternative definition of $\int_{-\infty}^{\infty} x dx$ for which the integral converges. This definition is called the **principal value** of the function $f(x) = x$ and it is denoted as $P.V \int_{-\infty}^{\infty} x dx$. Its definition is

$$P.V \int_{-\infty}^{\infty} x dx = \lim_{c\to\infty} \int_{-c}^{c} x dx \tag{954}$$

If we use this definition it is not hard to see that

$$P.V \int_{-\infty}^{\infty} x dx = \lim_{c\to\infty} \int_{-c}^{c} x dx = \lim_{c\to\infty} \frac{x^2}{2} \Big|_{x=-c}^{x=c} = 0 \tag{955}$$

Basically the principal value computes the integral symmetrically with respect to the y axis, while our definition of $\int_{-\infty}^{\infty} f(x) dx$ computes the integral as two separate improper integrals. It should be pointed out that there is no correct definition of $\int_{-\infty}^{\infty} f(x) dx$, each definition has its own strengths and weaknesses. Also, these two definitions do not necessarily disagree, in fact, whenever $\int_{-\infty}^{\infty} f(x) dx$ exists using formula 953, the principal value $P.V \int_{-\infty}^{\infty} f(x) dx$ will exist and it will agree with $\int_{-\infty}^{\infty} f(x) dx$.

Example 180. Find $\int_{-\infty}^{\infty} \left| x - \frac{1}{2} \right| e^{-x^2+x-1} dx$

Again we need to choose a real number c and compute $\int_{-\infty}^{c} \left|x-\frac{1}{2}\right| e^{-x^2+x-1}dx$ and $\int_{c}^{\infty} \left|x-\frac{1}{2}\right| e^{-x^2+x-1}dx$. Since the absolute value formula changes at $x = \frac{1}{2}$ we will choose $c = \frac{1}{2}$, but certainly any other number would work, for example $c = 0$, $c = \pi$, $c = -e$, although does choices would make the integral harder to compute. Therefore, using our definition we need to compute *separately* the improper integrals $\int_{-\infty}^{\frac{1}{2}} \left|x-\frac{1}{2}\right| e^{-x^2+x-1}dx = -\int_{-\infty}^{\frac{1}{2}} \left(x-\frac{1}{2}\right) e^{-x^2+x-1}dx$ and $\int_{\frac{1}{2}}^{\infty} \left|x-\frac{1}{2}\right| e^{-x^2+x-1}dx = \int_{\frac{1}{2}}^{\infty} \left(x-\frac{1}{2}\right) e^{-x^2+x-1}dx$.

In any case we will find first $\int \left(x-\frac{1}{2}\right) e^{-x^2+x-1}dx$. Using the substitution $u = -x^2+x-1$ we have $du = (-2x+1)dx = -2\left(x-\frac{1}{2}\right)dx$ and so

$$\begin{aligned} \int \left(x-\tfrac{1}{2}\right) e^{-x^2+x-1}dx &= \int -\tfrac{1}{2}e^{u}du \\ &= -\tfrac{1}{2}e^{u}+C \\ &= -\tfrac{1}{2}e^{-x^2+x-1}+C \end{aligned} \tag{956}$$

Therefore

$$\begin{aligned} \int_{-\infty}^{\frac{1}{2}} \left|x-\tfrac{1}{2}\right| e^{-x^2+x-1}dx &= -\lim_{a\to-\infty} \left(-\tfrac{1}{2}e^{-x^2+x-1}+C\right)\Big|_{x=a}^{x=\frac{1}{2}} \\ &= \tfrac{1}{2}\lim_{a\to-\infty} \left(e^{-\frac{3}{4}}-e^{-a^2+a-1}\right) \\ &= \tfrac{1}{2}e^{-\frac{3}{4}} \end{aligned} \tag{957}$$

and the second integral is

$$\begin{aligned} \int_{\frac{1}{2}}^{\infty} \left|x-\tfrac{1}{2}\right| e^{-x^2+x-1}dx &= \lim_{b\to\infty} \left(-\tfrac{1}{2}e^{-x^2+x-1}+C\right)\Big|_{x=\frac{1}{2}}^{x=b} \\ &= -\tfrac{1}{2}\lim_{b\to\infty} \left(e^{-b^2+b-1}-e^{-\frac{3}{4}}\right) \\ &= \tfrac{1}{2}e^{-\frac{3}{4}} \end{aligned} \tag{958}$$

Therefore, the value for $\int_{-\infty}^{\infty} \left|x-\frac{1}{2}\right| e^{-x^2+x-1}dx$ is

$$\begin{aligned} \int_{-\infty}^{\infty} \left|x-\tfrac{1}{2}\right| e^{-x^2+x-1}dx &= \int_{-\infty}^{\frac{1}{2}} \left|x-\tfrac{1}{2}\right| e^{-x^2+x-1}dx + \int_{\frac{1}{2}}^{\infty} \left|x-\tfrac{1}{2}\right| e^{-x^2+x-1}dx \\ &= \tfrac{1}{2}e^{-\frac{3}{4}}+\tfrac{1}{2}e^{-\frac{3}{4}} \\ &= e^{-\frac{3}{4}} \end{aligned} \tag{959}$$

Example 181. Show that $\int_{-\infty}^{\infty} xe^{-x^2}dx$ is convergent and find its value.

We compute the integral as

$$\int_{-\infty}^{\infty} xe^{-x^2}dx = \int_{-\infty}^{0} xe^{-x^2}dx + \int_{0}^{\infty} xe^{-x^2}dx \tag{960}$$

First we find the indefinite integral $\int xe^{-x^2}dx$ using the substitution $u = -x^2$, $du = -2xdx$:

$$\int xe^{-x^2}dx = -\frac{1}{2}\int e^u du = -\frac{1}{2}e^u + C = -\frac{1}{2}e^{-x^2} + C \tag{961}$$

Therefore

$$\begin{aligned} \int_{-\infty}^{0} xe^{-x^2}dx &= \lim_{a\to-\infty}\left(-\tfrac{1}{2}e^{-x^2}+C\right)\Big|_{x=a}^{x=0} \\ &= \lim_{a\to-\infty}\left(-\tfrac{1}{2}+\tfrac{1}{2}e^{-a^2}\right) \\ &= -\tfrac{1}{2} \end{aligned} \tag{962}$$

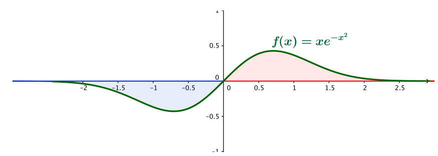


Figure 145: $\int_{-\infty}^{\infty} xe^{-x^2}dx = 0$

The other integral is computed similarly

$$\begin{aligned} \int_{0}^{\infty} xe^{-x^2}dx &= \lim_{b\to\infty}\left(-\tfrac{1}{2}e^{-x^2}+C\right)\Big|_{x=0}^{x=b} \\ &= \lim_{b\to\infty}\left(-\tfrac{1}{2}e^{-b^2}+\tfrac{1}{2}\right) \\ &= \tfrac{1}{2} \end{aligned} \tag{963}$$

Therefore

$$\int_{-\infty}^{\infty} xe^{-x^2}dx = \int_{-\infty}^{0} xe^{-x^2}dx + \int_{0}^{\infty} xe^{-x^2}dx = -\frac{1}{2}+\frac{1}{2} = 0 \tag{964}$$

We can interpret this result by saying that the area above the x axis is the same as the area below the x axis.

Example 182. Prove that $\int_{-\infty}^{\infty} x/(1+x^2)dx$ diverges but that the principal value $P.V \int_{-\infty}^{\infty}\frac{xdx}{1+x^2} = \lim_{b\longrightarrow\infty}\int_{-b}^{b}\frac{x}{1+x^2}dx$ converges

First of all, notice that

$$\int \frac{x}{1+x^2}dx = \frac{1}{2}\ln(1+x^2)+C \tag{965}$$

To compute $\int_{-\infty}^{\infty}\frac{x}{1+x^2}dx$ we break the integral into two pieces

$$\int_{-\infty}^{\infty}\frac{x}{1+x^2}dx = \int_{-\infty}^{0}\frac{x}{1+x^2}dx + \int_{0}^{\infty}\frac{x}{1+x^2}dx \tag{966}$$

To find $\int_0^\infty \frac{x}{1+x^2}\,dx$ we compute the following limit

$$\begin{aligned}
\int_0^\infty \frac{x}{1+x^2}\,dx &= \lim_{b\to\infty}\int_0^b \frac{x}{1+x^2}\,dx \\
&= \lim_{b\to\infty}\left(\frac{1}{2}\ln(1+x^2)+C\right)\Big|_{x=0}^{x=b} \\
&= \lim_{b\to\infty}\frac{1}{2}\ln(1+b^2) \\
&= \infty
\end{aligned} \tag{967}$$

and since this term diverges the entire integral $\int_{-\infty}^{\infty} \frac{x}{1+x^2}\,dx$ can't converge.

On the other hand, notice that the principal value in this case is

$$\begin{aligned}
\lim_{b\to\infty}\int_{-b}^b \frac{x}{1+x^2}\,dx &= \lim_{b\to\infty}\left(\frac{1}{2}\ln(1+x^2)+C\right)\Big|_{x=-b}^{x=b} \\
&= \lim_{b\to\infty}\frac{1}{2}\left[\ln(1+b^2)-\ln(1+(-b)^2)\right] \\
&= \lim_{b\to\infty} 0 \\
&= 0
\end{aligned} \tag{968}$$

and so it converges!

Part VII

More Applications of the Integral

Area Between Curves

Recall that our original interpretation of the definite integral $\int_a^b f(x)dx$ was as the area under the curve $y = f(x)$. Now, the way in which the definite integral was defined resulted in the area being counted as negative if the curve y is below the x axis and being counted as positive if the curve y is above the x axis.

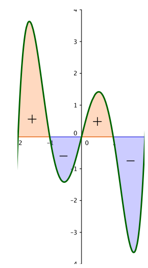


Figure 146: $\int_a^b f(x)dx$ computes the "net area"

Therefore, if we need to compute the "true" geometric area, we need to compute $\int_a^b |f(x)|dx$.

Now, we can consider $\int_a^b |f(x)|dx$ as the area between the curve $y = f(x)$ and the x axis. Since the equation for the x axis is simply $y = 0$, then we can write

$$\int_a^b |f(x) - 0|dx = \text{area between } y = f(x) \text{ and } y = 0 \qquad (969)$$

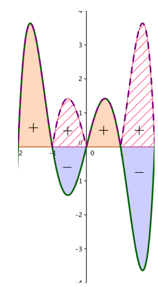


Figure 147: $\int_a^b |f(x)|dx$ computes the geometric area

Now, if we have two curves $y = f(x)$ and $y = g(x)$, it is not difficult to see that

$$\int_a^b |f(x) - g(x)|dx = \text{area between } y = f(x) \text{ and } y = g(x) \qquad (970)$$

Area between curves: Let f and g be two continuous functions on the interval $[a, b]$. Then the area of the region between the curves $y = f(x)$ and $y = g(x)$ is given by

$$\int_a^b |f(x) - g(x)|dx \qquad (971)$$

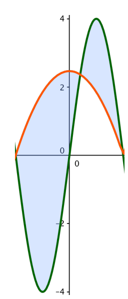


Figure 148: $\int_a^b |f(x) - g(x)|dx$ computes the area between the curves $y = f(x)$ and $y = g(x)$

Example 183. Find the area for the region bounded between the curves $y = x^2$ and $y = 2 - x$

First we need to find where the curves $y = x^2$ and $y = 2 - x$ intersect. We need to solve the equation

$$x^2 = 2 - x \qquad (972)$$

which is the same as

$$x^2 + x - 2 = 0 \implies x = -2, 1 \qquad (973)$$

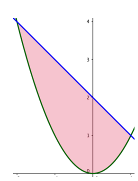


Figure 149: Area between $y = x^2$ and $y = 2 - x$

Therefore we need to find the value of the integral

$$
\begin{aligned}
& \int_{-2}^{1} |x^2-(2-x)|dx \\
= & \int_{-2}^{1}(2-x-x^2)dx \\
= & \left(2x-\tfrac{x^2}{2}-\tfrac{x^3}{3}\right)\Big|_{x=-2}^{x=1} \\
= & \left(2-\tfrac{1}{2}-\tfrac{1}{3}\right)-\left(-4-2+\tfrac{8}{3}\right) \\
= & \tfrac{9}{2}
\end{aligned}
\tag{974}
$$

Example 184. Find the area between the curves $y=e^x$ and $y=\frac{x}{e}-1$, for $-1\le x\le 1$.

Using the formula for the area between curves we need to compute

$$\int_{-1}^{1}\left|e^x-\left(\frac{x}{e}-1\right)\right|dx \tag{975}$$

Therefore, we need to find when is $e^x-\left(\frac{x}{e}-1\right)$ positive and when it is negative in order to eliminate the absolute value bars.

Consider the difference function

$$h(x)=f(x)-g(x)=e^x-\left(\frac{x}{e}-1\right)=e^x-\frac{x}{e}+1 \tag{976}$$

Observe that

$$h(-1)=e^{-1}+\frac{1}{e}+1=2e^{-1}+1 \tag{977}$$

so if we can show that $h(x)$ is increasing on $-1\le x\le 1$, we would conclude that $h(x)$ is positive and so $|h(x)|=h(x)$. To determine if $h(x)$ is increasing on $-1\le x\le 1$ consider the derivative of h:

$$h'(x)=e^x-\frac{1}{e}=\frac{e^{x+1}-1}{e} \tag{978}$$

Therefore we need to solve the equation $h'(x)\ge 0$, which is equivalent to solving the inequality $e^{x+1}-1\ge 0$, which is equivalent to

$$e^{x+1}\ge 1 \tag{979}$$

Applying ln to both sides of the inequality

$$\underbrace{\ln(e^{x+1})}_{x+1}\ge\underbrace{\ln(1)}_{0} \tag{980}$$

which implies that

$$x\ge -1 \tag{981}$$

Therefore, $h'(x)$ is positive on the interval $-1\le x\le 1$ and so $h(x)$ is increasing on the interval $-1\le x\le 1$.

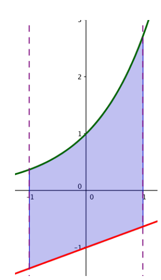


Figure 150: Area between $f(x)=e^x$ and $g(x)=\frac{x}{e}-1$ on the interval $[-1,1]$

This means that

$$\left|e^x - \left(\frac{x}{e} - 1\right)\right| = e^x - \left(\frac{x}{e} - 1\right) \text{ on the interval } -1 \le x \le 1 \quad (982)$$

Therefore the area between the curves $y = e^x$ and $y = \frac{x}{e} - 1$ on the interval $-1 \le x \le 1$ is

$$\begin{aligned} & \int_{-1}^{1} \left(e^x - \left(\tfrac{x}{e} - 1\right)\right) dx \\ = & \int_{-1}^{1} \left(e^x - \tfrac{x}{e} + 1\right) dx \\ = & \left(e^x - \tfrac{x^2}{2e} + x\right)\Big|_{x=-1}^{x=1} \\ = & \left(e - \tfrac{1}{2e} + 1\right) - \left(e^{-1} - \tfrac{1}{2e} - 1\right) \\ = & e - e^{-1} + 2 \end{aligned} \quad (983)$$

Example 185. Find the area between the curves $y = \tan x$ and $y = 2\sin x$, for $-\frac{\pi}{3} \le x \le \frac{\pi}{3}$

Using the formula for the area between curves we need to compute

$$\int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} |\tan x - 2\sin x|\, dx \quad (984)$$

Therefore, we need to find when is $\tan x - 2\sin x$ positive and when is it negative in order to eliminate the absolute values. Observe that

$$\begin{aligned} & \tan x - 2\sin x \\ = & \tfrac{\sin x}{\cos x} - 2\sin x \\ = & (\sin x)\left(\tfrac{1}{\cos x} - 2\right) \\ = & (\sin x)\left(\tfrac{1-2\cos x}{\cos x}\right) \end{aligned} \quad (985)$$

Therefore, to determine when is $h(x) = \tan x - 2\sin x$ positive and negative we will do a table of signs for $(\sin x)\left(\frac{1-2\cos x}{\cos x}\right)$ on the interval $-\frac{\pi}{3} \le x \le \frac{\pi}{3}$.

Observe that $\sin x$ is 0 for $x = 0$, $1 - 2\cos x$ is 0 when $\cos x = \frac{1}{2}$, that is, when $x = -\frac{\pi}{3}, \frac{\pi}{3}$ and finally, $\cos x$ never vanishes on $-\frac{\pi}{3} \le x \le \frac{\pi}{3}$. Therefore, the table of signs looks like

	$\left[-\frac{\pi}{3}, 0\right]$	$\left[0, \frac{\pi}{3}\right]$
$\sin x$	$-$	$+$
$1 - 2\cos x$	$-$	$-$
$\cos x$	$+$	$+$
sign of $h(x)$	$+$	$-$

Therefore, $h(x) = \tan x - 2\sin x$ is positive on $\left[-\frac{\pi}{3}, 0\right]$ and negative on $\left[0, \frac{\pi}{3}\right]$ and so

$$|\tan x - 2\sin x| = \begin{cases} \tan x - 2\sin x & \text{on } -\frac{\pi}{3} \le x \le 0 \\ -(\tan x - 2\sin x) & \text{on } 0 \le x \le \frac{\pi}{3} \end{cases} \quad (986)$$

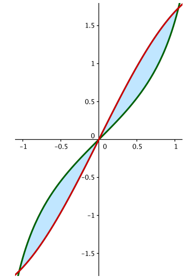


Figure 151: Area between $y = \tan x$ and $y = 2\sin x$ for $-\frac{\pi}{3} \le x \le \frac{\pi}{3}$

The formula for the area becomes

$$\begin{aligned} & \int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} |\tan x - 2\sin x| \, dx \\ = \quad & \int_{-\frac{\pi}{3}}^{0} |\tan x - 2\sin x| \, dx + \int_{0}^{\frac{\pi}{3}} |\tan x - 2\sin x| \, dx \\ = \quad & \int_{-\frac{\pi}{3}}^{0} (\tan x - 2\sin x) \, dx - \int_{0}^{\frac{\pi}{3}} (\tan x - 2\sin x) \, dx \end{aligned} \tag{987}$$

We will find first the indefinite integral

$$\begin{aligned} & \int (\tan x - 2\sin x) dx \\ = \quad & \int \tan x dx - 2\int \sin x dx \\ = \quad & \int \frac{\sin x}{\cos x} dx - 2\int \sin x dx \\ = \quad & -\ln|\cos x| + 2\cos x + C \end{aligned} \tag{988}$$

where we used the substitution $u = \cos x$ for the first integral. In this way the values of the definite integrals are

$$\begin{aligned} & \int_{-\frac{\pi}{3}}^{0} (\tan x - 2\sin x) \, dx - \int_{0}^{\frac{\pi}{3}} (\tan x - 2\sin x) \, dx \\ = \quad & (-\ln|\cos x| + 2\cos x + C)|_{x=-\frac{\pi}{3}}^{x=0} - (-\ln|\cos x| + 2\cos x + C)|_{x=0}^{x=\frac{\pi}{3}} \\ = \quad & \left(-\ln(1) + 2\cos 0 + \ln\left(\cos\left(\frac{\pi}{3}\right)\right) - 2\cos\left(\frac{\pi}{3}\right)\right) - \left(-\ln\left(\cos\left(\frac{\pi}{3}\right)\right) + 2\cos\left(\frac{\pi}{3}\right) + \ln(1) - 2\cos 0\right) \\ = \quad & 2 + \ln\left(\frac{1}{2}\right) - 2\left(\frac{1}{2}\right) - \left(-\ln\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) - 2\right) \\ = \quad & 2 + 2\ln\left(\frac{1}{2}\right) \\ = \quad & 2 + 2(\ln 1 - \ln 2) \\ = \quad & 2 - 2\ln 2 \end{aligned} \tag{989}$$

Solids of Revolution

As another application of the definite integral, we consider the problem of finding the volume V of the solid of revolution obtained by revolving the region below the graph of a nonnegative function $y = f(x)$ from $x = a$ to $x = b$ about the x axis. For example, the next image shows the solid of revolution $y = \sqrt{x}$ from $x = 0$ to $x = 4$.

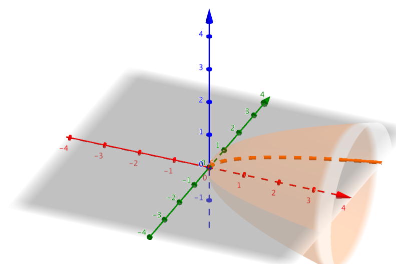


Figure 152: Solid of revolution obtained from rotating $y = \sqrt{x}$ about the x axis for $0 \le x \le 4$

To find the volume V we use a similar approach to the one we used to find the area under the curve. In the latter case, we approximated the area by the area of rectangles of width $\triangle x$ and height $f(x)$. Then the area was computed as the limit of such approximations, as the width of the rectangles shrunk to zero.

The strategy we use to find V is the following:

1. Divide the interval $[a, b]$ into n subintervals of length $\triangle x = \frac{b-a}{n}$. These intervals are $[a, a + \triangle x]$, $[a + \triangle x, a + 2\triangle x]$, ..., $[a + (n - 1)\triangle x, a + n\triangle x]$.

2. Construct n cylinders of thickness $\triangle x$ and radius $f(x)$, where x is the final point of each interval [56]

3. The volume of each cylinder is $\pi(f(x))^2\triangle x$ and so the volume of the n cylinders is

$$\pi \left((f(a + \triangle x))^2 + (f(a + 2\triangle x))^2 + \cdots + (f(a + n\triangle x))^2 \right) \triangle x \tag{990}$$

   As an example, the next image shows this procedure for $f(x) = \sqrt{x}$ for $0 \le x \le 4$ with $n = 4$.

4. Now the volume should be the volume of the previous approximation as $\triangle x \to 0$. If we recall that the general interpretation of the definite integral is

$$\int_a^b q(u)du \simeq \text{sum of contributions of magnitude } q(u)\triangle u \tag{991}$$

   It is clear from the approximation 990 that the volume should be

$$V = \int_a^b \pi \left[f(x)\right]^2 dx \tag{992}$$

   Therefore, we just found the following result:

[56] In fact, the choice of x is arbitrary, we chose the initial point to simplify the notation

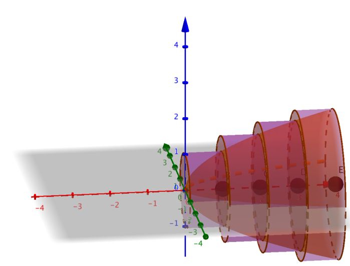


Figure 153: Approximation of the volume by cylinders

Volume of a solid of revolution: the volume V of the solid of revolution obtained by revolving the region below the graph of a nonnegative function $y = f(x)$ from $x = a$ to $x = b$ about the x axis is

$$V = \pi \int_a^b [f(x)]^2 \, dx \tag{993}$$

We will now apply this formula to find the volume of our example with $f(x) = \sqrt{x}$ for $0 \leq x \leq 4$. In this case we have to compute

$$\begin{aligned} V &= \pi \int_0^4 [\sqrt{x}]^2 dx \\ &= \pi \int_0^4 x dx \\ &= \pi \left. \frac{x^2}{2} \right|_{x=0}^{x=4} \\ &= 8\pi \end{aligned} \tag{994}$$

Example 186. Find the volume of a frustum of a right circular cone with height h, lower bases radius R, and top radius r

We regard the frustum as a solid of revolution by letting the x axis go through the center of the cone as shown in the figure.

In this case the function $f(x)$ is the segment at the top of the figure. We can set the origin of the x axis so that the segment goes through the points $(0, R)$ and (h, r). The equation has slope

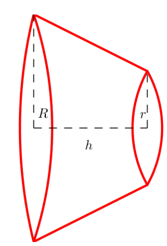


Figure 154: Volume frustum of a right circular cone

$$m = \frac{r - R}{h} \tag{995}$$

and so the equation of the line is

$$y - R = \left(\frac{r - R}{h}\right) x \tag{996}$$

In this case we are integration over the interval $0 \leq x \leq h$ and so the volume is

$$V = \pi \int_0^h \left[\left(\frac{r - R}{h}\right) x + R\right]^2 dx \tag{997}$$

To find the integral we make the substitution

$$\begin{aligned} u &= \left(\tfrac{r-R}{h}\right) x + R \\ du &= \left(\tfrac{r-R}{h}\right) dx \end{aligned} \tag{998}$$

and so the integral becomes

$$\begin{aligned} V &= \pi \int_R^r \left(\frac{h}{r-R}\right) u^2 du \\ &= \pi \left(\frac{h}{r-R}\right) \frac{u^3}{3}\Big|_{u=R}^{u=r} \\ &= \frac{\pi}{3}\left(\frac{h}{r-R}\right)\left(r^3 - R^3\right) \end{aligned} \tag{999}$$

If we want we can simplify this formula further using the identity $a^3 - b^3 = (a-b)(a^2+ab+b^2)$ and so the volume is

$$V = \frac{\pi h}{3}(r^2 + rR + R^2) \tag{1000}$$

Example 187. The upper half of the ellipse $\frac{x^2}{16} + \frac{y^2}{4} = 1$ is rotated around the x axis. The surface it generates is a special kind of ellipsoid called a prolate spheroid and has the shape of a football. Find the volume of this spheroid.

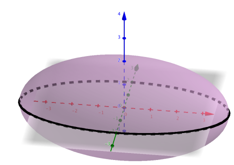


Figure 155: Volume prolate spheroid

If we take the equation of the ellipse $\frac{x^2}{16} + \frac{y^2}{4} = 1$ and solve for y^2 (which is what we need in the integrand of 993) we find that

$$y^2 = 4 - \frac{x^2}{4} \tag{1001}$$

To find the limits of integration for x, we set $y = 0$ in the previous equation to find $x = \pm 4$ and so the volume of the spheroid is

$$\begin{aligned} V &= \pi \int_{-4}^{4} \left(4 - \frac{x^2}{4}\right) dx \\ &= \pi \left(4x - \frac{x^3}{12}\right)\Big|_{x=-4}^{x=4} \\ &= \pi \left[\left(16 - \frac{16}{3}\right) - \left(-16 + \frac{16}{3}\right)\right] \\ &= \frac{64\pi}{3} \end{aligned} \tag{1002}$$

We finish mentioning some paradoxes (or counterintuitive results if you prefer) that may arise when you combine improper integrals with the formula for the volume of a solid of revolution.

Consider the curve $y = f(x) = \frac{1}{x}$. The volume of the solid of revolution obtained by rotated about the x axis from $x = 1$ to $x = b$ for $b > 1$ is [57]

$$\begin{aligned} V(b) &= \pi \int_1^b \frac{1}{x^2} dx \\ &= \pi \left(-\frac{1}{x}\right)\Big|_{x=1}^{x=b} \\ &= \pi \left(-\frac{1}{b} + 1\right) \end{aligned} \tag{1003}$$

[57] Since we haven't specified a particular value for b we consider V as a function of b, which is why we use the functional notation

Observe that $\lim_{b\to\infty} V(b) = \pi$ and so we can interpret this by saying that the volume of the solid of revolution obtained by revolving the hyperbola $y = \frac{1}{x}$ for $x \geq 1$ is finite. On the other hand, the area under the curve of $y = \frac{1}{x}$ for $x \geq 1$ is

$$\begin{aligned} A &= \int_1^\infty \frac{dx}{x} \\ &= \lim_{b\to\infty} \int_1^b \frac{dx}{x} \\ &= \lim_{b\to\infty} \ln x|_{x=1}^{x=b} \\ &= \lim_{b\to\infty} \ln b \\ &= \infty \end{aligned} \tag{1004}$$

Therefore, the area under the curve is infinite and so we just found an example of an infinite area generating a finite volume of revolution! Clearly this may seem disturbing, but it should be pointed out that this paradox is happening for an example for which we have no previous experience, after all, it is empirically impossible to take an infinite hyperbola and rotate it to compute its volume. On the other hand, there are many curves (for example, $y = \frac{1}{x^2}$ for $x \geq 1$) for which both the area and the volume will be finite, so there are benefits in using improper integrals in computing the volume of solids of revolution.

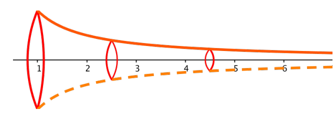


Figure 156: Gabriel's Horn: An infinite area generates a finite volume of revolution

Solving Differential Equations

As we are about to see, integrals are the major tool in solving **differential equations.** An ordinary (algebraic) equation relates different (algebraic) operations performed on a variable x, for example, the equation

$$x + \cos(x) = 1 \tag{1005}$$

says that a variable x plus its cosine equals 0. It is easy to see that $x = 0$ solves the equation but perhaps it is not obvious if there are other solutions. A way to see that this is the only solution is to reinterpret the previous equation geometrically. The idea is to consider the left hand side as defining a function, in our case $f(x) = x + \cos(x)$, and the right hand side as defining another function, in this case $g(x) = 1$. Therefore the equation is equivalent to

$$f(x) = g(x) \tag{1006}$$

and so what we are trying to determine is those values of x for which the graphs of f and g intersect.

Therefore, for us **solving an algebraic equation is equivalent to determining the points of intersection between two curves**. Looking at the graph of the two functions it is reasonable to expect that the only intersect when $x = 0$, and to justify this rigorously we can use calculus. Observe that

$$f'(x) = 1 - \sin(x) \tag{1007}$$

and because $\sin(x) \leq 1$ for any value of x we have that $f'(x) \geq 0$ for all values of x, that is $f(x)$ is never decreasing. This means in particular that it can only intersect any given line at most one time and so the graphs of $f(x) = x + \cos(x)$ and $g(x) = 1$ only intersect when $x = 0$, which is what we wanted to show.

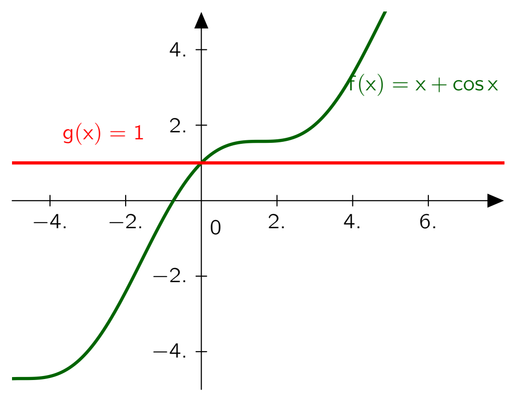


Figure 157: Geometric interpretation of the equation $x + \cos x = 1$

A differential equation is similar to an algebraic equation, but now the equation relates a variable with its derivatives. For example the equation

$$\frac{dy}{dx} = 2y \tag{1008}$$

Relates a function $y(x)$ to its derivative $\frac{dy}{dx}$. Again, we can see that $y(x) = e^{2x}$ satisfies the previous equation by computing the derivative of $y(x)$, that is, $\frac{dy}{dx} = 2e^{2x}$ and noticing that this is the same as $2y(x)$. Now we can ask the same question as before, that is, is $y(x) = e^{2x}$ the only solution of the differential equation $\frac{dy}{dx} = 2y$. Unfortunately this is no longer true, for example, $y_2(x) = 2e^{2x}$ also satisfies the differential equation and in fact, $y_c(x) = ce^{2x}$ satisfies the differential equation where c is an arbitrary constant. Then the next question one can ask is if we can find solutions of $\frac{dy}{dx} = 2y$ which are different from an exponential. To see this will not be the case we will interpret the differential equation geometrically, just as we did for the algebraic equation.

The equation $\frac{dy}{dx} = 2y$ can be interpreted in the following way: find a function $y(x)$ whose derivative is $2y(x)$. Since the derivative of a function is the slope of the tangent line, we can restate the differential equation $\frac{dy}{dx} = 2y$ as: find a function $y(x)$ such that the slope of the tangent line to the graph of y as the point $(x, y(x))$ is $2y(x)$. This can be represented using a **direction field**: to construct it draw the xy plane and at each point (x, y) we draw a small straight segment whose slope is $y'(x)$. A solution of the differential equation is the a curve that follows the direction field, that is, a curve whose tangent lines coincide with the segments we drew on the direction field. The following picture show the direction field for $\frac{dy}{dx} = 2y$ and three of its solutions: $y(x) = e^{2x}$, $y_{-3}(x) = -3e^{2x}$ and $y_0(x) = 0 = 0e^{2x}$.

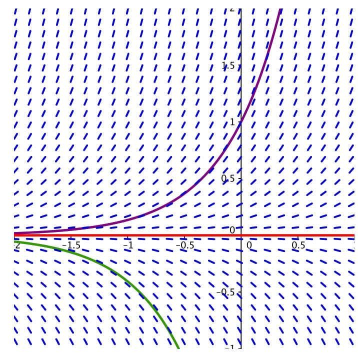


Figure 158: Direction field of $\frac{dy}{dx} = 2y$

From the direction field it looks like all the solutions will be exponential functions (and the line $y = 0$). To show this must be the case we go back to our differential equation

$$\frac{dy}{dx} = 2y \tag{1009}$$

If $y \neq 0$ then we can divide by y both sides of the equation to obtain

$$\frac{1}{y}\frac{dy}{dx} = 2 \tag{1010}$$

and we "multiply" by dx to obtain [58]

$$\frac{dy}{y} = 2dx \tag{1011}$$

[58] Strictly speaking one can't multiply by dx but the Leibniz notation is set up in such a way that the solution can be found as if multiplication by dx were possible.

and if we integrate both sides of the equation we obtain

$$\int \frac{dy}{y} = \int 2dx \tag{1012}$$

Each of these integrals is easy to compute so we find that

$$\ln|y| = 2x + C \tag{1013}$$

which we can solve for $|y|$ as

$$|y| = e^{2x+C} = e^C e^{2x} \tag{1014}$$

This means that $y = e^C e^{2x}$ or $y = -e^C e^{2x}$ and since e^C is a positive constant we can simply write $y = ce^{2x}$ where $c = e^C$ or $c = -e^C$. Therefore we found that every solution is of the form

$$y = ce^{2x} \tag{1015}$$

where c is a positive or negative constant. Notice that we assumed that $y \neq 0$. If $y = 0$ we find another solution (the x axis) which can be written in the form ce^{2x} by taking $c = 0$. Therefore, we just proved our claim that all the solutions of the differential equations are exponential functions. Finally, although we won't use this fact, **one of the properties of the solutions of the same differential equation is that the corresponding curves do not intersect**.

Separable Equations

Solving differential equations is a lot harder than solving algebraic equations, which is why we will simply focus on the easiest type of differential equations, the so called **separable equations.** These are differential equations of the form

$$\frac{dy}{dx} = f(x)g(y) \tag{1016}$$

That is, the derivative $\frac{dy}{dx}$ depends on x and y through separate functions. We will solve them as we did with the previous example. Assuming that $g(y) \neq 0$ we can divide both sides of the equation by $g(y)$ and multiply both sides by dx so that the differential equations becomes

$$\frac{dy}{g(y)} = f(x)dx \tag{1017}$$

Now we integrate both sides of the equation and the solution will be given by

$$\int \frac{dy}{g(y)} = \int f(x)dx \tag{1018}$$

There will be an arbitrary constant of integration. To eliminate the constant we need an **initial condition,** that is, the value of $y(x)$ at a particular point x_0. When an initial condition is specified, we call the differential equation an **initial valued problem.** These will be seen in the following examples.

Solution of a Separable Differential Equation: suppose we are given a first order differential in the form

$$\frac{dy}{dx} = f(x)g(y) \tag{1019}$$

To solve it write it in the form

$$\frac{dy}{g(y)} = f(x)dx \tag{1020}$$

and integrate each side of the equation as

$$\int \frac{dy}{g(y)} = \int f(x)dx \tag{1021}$$

Example 188. Solve the differential equation

$$\frac{dy}{dx} = \frac{2y+3}{x^2} \tag{1022}$$

We rewrite it as

$$\frac{dy}{2y+3} = \frac{dx}{x^2} \tag{1023}$$

and we integrate both sides

$$\int \frac{dy}{2y+3} = \int x^{-2} dx \tag{1024}$$

The left hand side can be integrated using $u = 2y + 3$ and so the answer is

$$\frac{1}{2} \ln |2y+3| = -x^{-1} + C \tag{1025}$$

To solve for y multiply both sides by 2 so

$$\ln |2y+3| = 2C - \frac{2}{x} \tag{1026}$$

and exponentiate both sides of the equation to obtain

$$|2y+3| = e^{2C-\frac{2}{x}} = e^{2C} e^{-\frac{2}{x}} \tag{1027}$$

Therefore $2y + 3 = ce^{-\frac{2}{x}}$ where $c = \pm e^{2C}$ and so the solution becomes

$$y = \frac{ce^{-\frac{2}{x}} - 3}{2} \tag{1028}$$

Example 189. Solve the initial value problem

$$\begin{cases} y' = xye^x \\ y(1) = 3 \end{cases} \tag{1029}$$

This differential equation is separable and can be written as

$$\frac{dy}{y} = xe^x dx \tag{1030}$$

and we integrate both sides of the equation

$$\int \frac{dy}{y} = \int xe^x dx \tag{1031}$$

The right integral can be found using integration by parts to obtain

$$\ln |y| = e^x x - e^x + C \tag{1032}$$

and so the solution is

$$|y| = e^{e^x x - e^x + C} = e^C e^{e^x x - e^x} \tag{1033}$$

Therefore $y = ce^{e^x x - e^x}$ where $c = \pm e^C$ and so the solution is

$$y = ce^{e^x x - e^x} \tag{1034}$$

Now we must use the initial condition $y(1) = 3$ which gives the equation

$$3 = ce^0 \tag{1035}$$

therefore $c = 3$ and the solution becomes

$$y = 3e^{e^x x - e^x} \tag{1036}$$

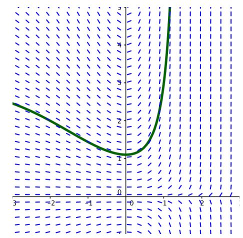


Figure 159: Solution of the initial value problem $y' = xye^x$, $y(1) = 3$

Example 190. Solve the first order initial value problem

$$\begin{cases} \frac{dy}{dx} = \cos^2(\pi - x) \\ y(0) = 1 \end{cases} \tag{1037}$$

The solution is

$$\int dy = \int \cos^2(\pi - x)dx \tag{1038}$$

To integrate the right hand side, we use the identity $\cos^2\theta = \frac{1}{2}(1 + \cos(2\theta))$ and so

$$\int \cos^2(\pi - x)dx = \frac{1}{2}\int(1 + \cos(2\pi - 2x))dx \tag{1039}$$

Using the identity $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ or the fact that cos is of period 2π we can see that $\cos(2\pi - 2x) = \cos(2x)$ and so we must integrate

$$\frac{1}{2}\int(1 + \cos(2x))dx = \frac{1}{2}(x + \frac{\sin(2x)}{2}) + C \tag{1040}$$

and so the solution of the differential equation is

$$y = \frac{x}{2} + \frac{\sin(2x)}{4} + C \tag{1041}$$

Using the initial condition $y(0) = 1$ we find that $C = 1$ and so the solution is

$$y(x) = \frac{x}{2} + \frac{\sin(2x)}{4} + 1 \tag{1042}$$

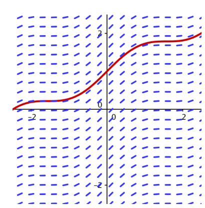


Figure 160: Solution of the initial value problem $\frac{dy}{dx} = \cos^2(\pi - x)$, $y(0) = 1$

Example 191. Solve the second order initial value theorem

$$\begin{cases} \frac{d^2y}{dt^2} = \sin(2t - \pi) \\ y(0) = 1 \\ y'(0) = -1 \end{cases} \tag{1043}$$

using the substitution $v = \frac{dy}{dt}$.

With the substitution $v = \frac{dy}{dt}$ we have $\frac{dv}{dt} = \frac{d^2y}{dt^2}$ the differential equation can be rewritten as

$$\frac{dv}{dt} = \sin(2t - \pi) \tag{1044}$$

and so the solution is

$$\int dv = \int \sin(2t - \pi) dt \tag{1045}$$

The second integral can be used using the substitution $u = 2t - \pi$ and so the solution is

$$v = -\frac{\cos(2t - \pi)}{2} + C \tag{1046}$$

The initial condition $y'(0) = -1$ translates into $v(0) = -1$ and so

$$-1 = -\frac{\cos(-\pi)}{2} + C \implies C = -\frac{3}{2} \tag{1047}$$

Therefore the formula for v becomes

$$v = -\frac{\cos(2t - \pi)}{2} - \frac{3}{2} \tag{1048}$$

To find y we use that $v = \frac{dy}{dt}$ and so

$$\frac{dy}{dt} = -\frac{\cos(2t - \pi)}{2} - \frac{3}{2} \tag{1049}$$

The differential equation is again separable so we integrate it one more time to find

$$y = \int dy = \int \left(-\frac{\cos(2t - \pi)}{2} - \frac{3}{2}\right) dt = -\frac{\sin(2t - \pi)}{4} - \frac{3}{2}t + C_2 \tag{1050}$$

where C_2 is another constant of integration. Using the initial condition $y(0) = 1$ we find that $C_2 = 1$ and so the solution is

$$y(t) = -\frac{\sin(2t - \pi)}{4} - \frac{3}{2}t + 1 \tag{1051}$$

Applications of Separable Equations

Example 192. Suppose that the population P of bacteria in a certain culture is increasing at a rate directly proportional to the square of the population. Find the constant of proportionality assuming that at time $t = 0$ minutes, the population is 10 and is increasing at a rate of 5 bacteria per minute.

Let $P(t)$ denote the population at time t. Because it increases at a rate directly proportional to the square of itself , $P(t)$ satisfies the differential equation

$$\frac{dP}{dt} = kP^2 \tag{1052}$$

where k is the proportionality constant we want to determine. The equation is separable and the solution is

$$\int \frac{dP}{P^2} = \int k dt \tag{1053}$$

and so

$$-P^{-1}(t) = kt + c \tag{1054}$$

Using the initial condition $P(0) = 10$ we obtain that $c = 10$ and so

$$-\frac{1}{P(t)} = kt + 10 \implies \boxed{P(t) = -\frac{1}{kt+10}} \tag{1055}$$

We also use that at time $t = 0$, $\frac{dP}{dt} = 5\frac{\text{bacteria}}{\text{minute}}$ and so using the equation $\frac{dP}{dt} = kP^2 = k\left(-\frac{1}{kt+10}\right)^2$ we find that

$$5 = \frac{k}{100} \tag{1056}$$

and so the value of the constant of proportionality is

$$\boxed{k = 500} \tag{1057}$$

Example 193. A tank initially contains 10 gallons of pure water. Brine containing 3 pounds of salt per gallon flows into the tank at a rate of 2 gallons per minute, and the well-stirred mixture flows out of the tank at the same rate. How much salt is present at the end of 10 minutes? How much salt is present in the long run?

Let $Q(t)$ denote the amount of salt in the tank at time t. Then

$$\frac{dQ}{dt} = \text{rate of salt flowing in } - \text{ rate of salt flowing out} \tag{1058}$$

The rate of salt flowing in is given by

$$(\text{rate of flow})(\text{concentration}) = (2\text{gal}/\text{min})(3\text{lb}/\text{text}) = 6\text{pounds}/\text{minute} \tag{1059}$$

Since the rate at which the solution leaves the tank is the same as the rate at which the brine is poured into it, the tank contains 10 gallons of the mixture at any time t. Since the salt content at any time t is Q pounds, the concentration of the salt in the mixture is $(Q/10)$ pounds per gallon. Therefore, the rate at which salt flows out of the tank is given by

$$(2\text{gal}/\text{min})\left(\frac{Q}{10}\text{lb}/\text{gal}\right) = \frac{Q}{5}\text{pounds per minute} \tag{1060}$$

Therefore, we obtain the differential equation

$$\frac{dQ}{dt} = 6 - \frac{Q}{5} \tag{1061}$$

The right hand side can be rewritten as

$$\frac{30-Q}{5} \tag{1062}$$

we can see that the differential equation is equivalent to

$$\frac{dQ}{30-Q} = \frac{dt}{5} \tag{1063}$$

The solution is given as

$$\int \frac{dQ}{30-Q} = \int \frac{dt}{5} \tag{1064}$$

the left hand side can be found with the substitution $u = 30 - Q$, we obtain

$$-\ln|30-Q| = \frac{1}{5}t + C \tag{1065}$$

which is the same as

$$\ln|30-Q| = -\frac{1}{5}t - C \tag{1066}$$

Exponentiate both sides to find Q

$$|30-Q| = e^{-\frac{1}{5}t}e^{-C} \implies 30-Q = ce^{-\frac{1}{5}t} \tag{1067}$$

and so

$$Q(t) = 30 - ce^{-\frac{1}{5}t} \tag{1068}$$

To find c we use that initially there was no salt in the tank so $Q(0) = 0$, therefore $c = 30$ and the solution is

$$Q(t) = 30(1 - e^{-\frac{t}{5}}) \tag{1069}$$

Example 194. a) Several models of plant growth have been developed. The simplest one assumes that the rate of growth of the dry mass of a plant, $W'(t)$, is only proportional to the amount of dry mass, that is, $W'(t) = \alpha W(t)$; α is a positive parameter. Find the formula for $W(t)$.

The equation is separable

$$\frac{dW}{W} = \alpha dt \tag{1070}$$

and so

$$\int \frac{dW}{W} = \int \alpha dt \tag{1071}$$

which gives as solution

$$\ln|W| = \alpha t + C \tag{1072}$$

and this gives

$$W(t) = e^{\alpha t}e^{C} = ce^{\alpha t} \tag{1073}$$

b) For a more realistic model consider the possibility that the rate of growth of the dry mass is proportional to the amount of substrate present ($S(t)$, the amount of nutrients), but is independent of the amount of dry mass, that is, $W'(t) = kS(t)$, where k is a positive parameter. Taking into account that, at any time t, $W(t) + S(t) = W_0 + S_0 = a$, the differential equation becomes $W'(t) = k(a - W(t))$. Find the new solution.

Now the differential equation is

$$\frac{dW}{a-W} = kdt \tag{1074}$$

Integrating both sides of the equation

$$\int \frac{dW}{a-W} = \int kdt \tag{1075}$$

and if we use the substitution $u = a - W$ on the left hand side we find

$$-\ln|a-W| = kt + C \tag{1076}$$

and so

$$|a-W| = e^{-kt-C} = e^{-C}e^{-kt} \tag{1077}$$

Therefore

$$a - W = ce^{-kt} \implies W(t) = a - ce^{-kt} \tag{1078}$$

Since $W(0) = W_0$ we have that $W_0 = a - c$ or $c = a - W_0$ and so the solution is

$$\boxed{W(t) = a - (a - W_0)e^{-kt}} \tag{1079}$$

Example 195.

Euler's Method

So far we have solved some examples of separable differential equations. The solutions of these equations reduced to a problem of finding antiderivatives. Since antiderivatives can be hard to find, it is unreasonable to try to solve these differential equations directly. Worse yet, separable equations are among the easiest equations to solve, so in general solving differential equations explicitly is a lost cause. Therefore, we need to come up with methods to approximate the solution of a differential equation. There are many different methods that try to find such approximations, but perhaps one of the easiest to explain is **Euler's Method.**

We will explain how to use Euler's Method by applying it to the differential equation

$$\begin{cases} y' = x - y \\ y(0) = 1 \end{cases} \tag{1080}$$

Notice in particular that this equation is not separable, so in fact with what we have discussed so far, we would not be able to solve this equation though this is actually one of the easy equations to solve since it is linear.

The idea is to approximate the actual solution of the differential equation (which is a curve in the xy plane) by a polygonal curve. Suppose that our initial condition is (x_0, y_0). In our example, $x_0 = 0$ and $y_0 = 1$. The idea is to find an approximation of $y(b)$, where b is a number greater than x_0.

We divide the interval $[x_0, b]$ into n subintervals, each of length

$$h = \frac{b - x_0}{n} \tag{1081}$$

This gives us n points

$$x_1 = x_0 + h, \quad , x_2 = x_0 + 2h, \quad x_3 = x_0 + 3h, \quad , x_n = x_0 + nh = b \tag{1082}$$

Euler's method consists in approximating the graph of f on the interval $[x_0, x_1]$ by the straight line segment that is tangent to the graph of f at (x_0, y_0). If we call $F(x, y)$ the right hand side of the differential equation, in our example, $F(x, y) = x - y$, then $F(x_0, y_0)$ is the slope of the tangent line going through (x_0, y_0) and so the equation of this tangent line is

$$y - y_0 = F(x_0, y_0)(x - x_0) \tag{1083}$$

The approximation y_1 to $y(x)$ is obtained by replacing x by x_1, that is

$$y_1 = y_0 + F(x_0, y_0)(x_1 - x_0) = y_0 + F(x_0, y_0)h \tag{1084}$$

The idea is now to iterate this procedure, namely, y_2 is related to y_1 in the same way as y_1 is related to y_0

$$y_2 = y_1 + F(x_1, y_1)h \tag{1085}$$

we continue in this way and thus Euler's Method can be stated as:

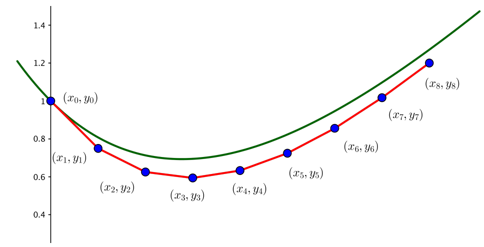


Figure 161: Example of Euler's Method for $y' = x - y$, $y(0) = 1$, $n = 8$, $b = 2$

Euler's Method: suppose we are given the differential equation

$$\frac{dy}{dx} = F(x, y) \tag{1086}$$

subject to the initial condition

$$y(x_0) = y_0 \tag{1087}$$

and we wish to find an approximation of $y(b)$, where b is a number greater than x_0 and n is a positive integer. Compute

$$h = \frac{b - x_0}{n} \tag{1088}$$

and

$$x_1 = x_0 + h, \quad x_2 = x_1 + h = x_0 + 2h, \ \cdots \ x_n = x_{n-1} + h = x_0 + nh = b \tag{1089}$$

Define

$$y_1 = y_0 + hF(x_0, y_0), \quad y_2 = y_1 + hF(x_1, y_1), \quad y_n = y_{n-1} + hF(x_{n-1}, y_{n-1}) \tag{1090}$$

Then y_n gives an approximation of the true value $y(b)$ of the solution to the initial value problem at $x = b$.

We will use Euler's method to solve our example, with $b = x = 2$ and $n = 8$. In this case

$$h = \frac{2 - 0}{8} = \frac{1}{4} \tag{1091}$$

and so

$$\begin{aligned}
x_1 &= x_0 + h = 0 + \tfrac{1}{4} = \tfrac{1}{4} \\
x_2 &= x_1 + h = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2} \\
x_3 &= x_2 + h = \tfrac{1}{2} + \tfrac{1}{4} = \tfrac{3}{4} \\
x_4 &= x_3 + h = \tfrac{3}{4} + \tfrac{1}{4} = 1 \\
x_5 &= x_4 + h = 1 + \tfrac{1}{4} = \tfrac{5}{4} \\
x_6 &= x_5 + h = \tfrac{5}{4} + \tfrac{1}{4} = \tfrac{3}{2} \\
x_7 &= x_6 + h = \tfrac{3}{2} + \tfrac{1}{4} = \tfrac{7}{4} \\
x_8 &= x_7 + h = \tfrac{7}{4} + \tfrac{1}{4} = 2
\end{aligned} \tag{1092}$$

Now we find the corresponding y values using $F(x, y) = x - y$

$$
\begin{aligned}
y_1 &= y_0 + hF(x_0, y_0) = 1 + \tfrac{1}{4}(0 - 1) = \tfrac{3}{4} \\
y_2 &= y_1 + hF(x_1, y_1) = \tfrac{3}{4} + \tfrac{1}{4}(\tfrac{1}{4} - \tfrac{3}{4}) = \tfrac{5}{8} \\
y_3 &= y_2 + hF(x_2, y_2) = \tfrac{5}{8} + \tfrac{1}{4}(\tfrac{1}{2} - \tfrac{5}{8}) = \tfrac{19}{32} \\
y_4 &= y_3 + hF(x_3, y_3) = \tfrac{19}{32} + \tfrac{1}{4}(\tfrac{3}{4} - \tfrac{19}{32}) = \tfrac{81}{128} \\
y_5 &= y_4 + hF(x_4, y_4) = \tfrac{81}{128} + \tfrac{1}{4}(1 - \tfrac{81}{128}) = \tfrac{371}{512} \\
y_6 &= y_5 + hF(x_5, y_5) = \tfrac{371}{512} + \tfrac{1}{4}(\tfrac{5}{4} - \tfrac{371}{512}) = \tfrac{1753}{2048} \\
y_7 &= y_6 + hF(x_6, y_6) = \tfrac{1753}{2048} + \tfrac{1}{4}(\tfrac{3}{2} - \tfrac{1753}{2048}) = \tfrac{8331}{8192} \\
y_8 &= y_7 + hF(x_7, y_7) = \tfrac{8331}{8192} + \tfrac{1}{4}(\tfrac{7}{4} - \tfrac{8331}{8192}) = \tfrac{39329}{32768}
\end{aligned}
\tag{1093}
$$

Part VIII

Sequences and (Power) Series

Sequences

In Calculus most of our attention has been focused to studying quantities that are continuous in some form. For example, all the functions we have studied so far had continuous domain and range (intervals of the real line) and changed continuously (continuity/differentiability). Also, the definite integral of a function could be considered as a continuous sum.

However, quantities that are discrete in nature are also useful in Calculus, as we are about to see. First we will start studying the discrete version of a function, which is known as a **sequence**. A sequence is basically a list of numbers. It is usually denoted by $\{a_n\}_{n=1}^{\infty}$, where a_1 is the first number in the list, a_2 the second number and so on. For example, the sequence $\left\{\frac{1}{n+1}\right\}_{n=1}^{\infty}$ represents the list of numbers $\frac{1}{1+1}$, $\frac{1}{2+1}$, $\frac{1}{3+1}$, etc. In this case we would write the sequence as $\{a_n\}_{n=1}^{\infty}$ where $a_n = \frac{1}{n+1}$. It is easy to see that successive term of the sequence gets smaller and smaller, so we would like to say that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n+1} = 0 \tag{1094}$$

This should not be surprising because for the function $f(x) = \frac{1}{x+1}$, we have that $\lim_{x\to\infty} \frac{1}{x+1} = 0$.

Limit of a Sequence: let $\{a_n\}$ be a sequence. We say that the sequence $\{a_n\}$ **converges** to the limit L, and we write

$$\lim_{n\to\infty} a_n = L \tag{1095}$$

if the terms of the sequence a_n can be made as close to L as we please by taking n sufficiently large. More precisely, $\lim_{n\to\infty} a_n = L$ if for every number $\epsilon > 0$ we can find a number N (depending on ϵ), so that if $n \geq N$ then $|a_n - L| < \epsilon$.
If a sequence is not convergent, we say that it is **divergent**.

As it is to be expected, limits of sequences behave very similarly to limits of functions. In fact, we have the following:

Properties of Limits of Sequences:

- Suppose that $\{a_n\}$ is a sequence and there is a function $f(x)$ such that $a_n = f(n)$. If $\lim_{x\to\infty} f(x) = L$, then $\lim_{n\to\infty} a_n = L$.
- **Squeeze Theorem**: suppose $\{a_n\}, \{b_n\}, \{c_n\}$ are sequences and $a_n \le b_n \le c_n$ for all n bigger than some index n_0. If $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$ then $\lim_{n\to\infty} b_n = L$.

If $\{a_n\}, \{b_n\}$ are two sequences with limits

$$\lim_{n\to\infty} a_n = A \qquad \lim_{n\to\infty} b_n = B \tag{1096}$$

Then

- $\lim_{n\to\infty} ca_n = c\lim_{n\to\infty} a_n = cA$ where c is a constant
- $\lim_{n\to\infty}(a_n \pm b_n) = \lim_{n\to\infty} a_n \pm \lim_{n\to\infty} b_n = A \pm B$
- $\lim_{n\to\infty} a_n b_n = (\lim_{n\to\infty} a_n)(\lim_{n\to\infty} b_n) = AB$
- $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = \frac{A}{B}$ provided $B \neq 0$ and $b_n \neq 0$
- $\lim_{n\to\infty} a_n^p = (\lim_{n\to\infty} a_n)^p$ if $p > 0$ and $a_n > 0$
- $\lim_{n\to\infty} f(a_n) = f(\lim_{n\to\infty} a_n)$ whenever f is continuous at the number $\lim_{n\to\infty} a_n$ and the terms of the sequence lie on the domain of the function f
- If $\lim_{n\to\infty} |a_n| = 0$ then $\lim_{n\to\infty} a_n = 0$

Example 196. Determine if the following sequences have a limit or not.

- $a_n = \sqrt{n+3} - \sqrt{n}$

Just as for limits of functions, we rationalize the expression

$$\begin{aligned} a_n &= \sqrt{n+3} - \sqrt{n} \\ &= \left(\sqrt{n+3} - \sqrt{n}\right)\left(\tfrac{\sqrt{n+3}+\sqrt{n}}{\sqrt{n+3}+\sqrt{n}}\right) \\ &= \tfrac{n+3-n}{\sqrt{n+3}+\sqrt{n}} \\ &= \tfrac{3}{\sqrt{n+3}+\sqrt{n}} \end{aligned} \tag{1097}$$

Therefore we find that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{3}{\sqrt{n+3}+\sqrt{n}} = 0 \tag{1098}$$

- $a_n = (-\frac{1}{4})^n$

Since $a_n = \frac{(-1)^n}{4^n}$ we have $-\frac{1}{4^n} \le a_n \le \frac{1}{4^n}$ and so by the squeeze theorem it is clear that $\lim_{n\to\infty} a_n = 0$

- $a_n = \frac{\sin n}{n^2}$

We have for all n

$$-\frac{1}{n^2} \leq \frac{\sin n}{n^2} \leq \frac{1}{n^2} \tag{1099}$$

Therefore, because of the squeeze theorem for sequences we can conclude that

$$\underbrace{\lim_{n\to\infty} -\frac{1}{n^2}}_{0} \leq \lim_{n\to\infty} \frac{\sin n}{n^2} \leq \underbrace{\lim_{n\to\infty} \frac{1}{n^2}}_{0} \tag{1100}$$

and so the limit is zero.

Series: Definitions and Some Properties

Just as a sequence is a discrete version of a function, we can think of a **series** as a discrete version of an integral[59]. A series is the technical tool required to “add” infinitely many numbers.

For example, suppose we begin with a rod of length 1 and divide it into two pieces, each of length 1/2. We can now take one of these pieces and bisect it again, giving us two more pieces, each of length 1/4. We iterate this process: that is, we take one of the pieces of length 1/4 and bisect it again. The following diagram illustrates this process:

We can define the sequence $a_n = \frac{1}{2^n}$ for $n \geq 1$: this sequence lists the pieces of the rod that are no longer bisected. At every step we can add the pieces that are no longer bisected as the table shows for the first steps

Step n	$a_1 + \cdots + a_n$
1	$a_1 = \frac{1}{2}$
2	$a_1 + a_2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$
3	$a_1 + a_2 + a_3 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$

Clearly this process of bisecting the rod can be continued indefinitely [60] because at the end of step n there is a remaining piece of length $\frac{1}{2^n}$ which is used to generate two more rectangles. Also, at every step we will have that the sum of the pieces is less than one, that is, for every n we have

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n} < 1 \qquad (1101)$$

The idea of a series is to try to add all the pieces simultaneously, that is, consider the sum at “step infinity”

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n} + \cdots + \qquad (1102)$$

which we will write in a convenient notation as

$$\sum_{i=1}^{\infty} \frac{1}{2^i} \qquad (1103)$$

where the symbol $\sum$ is used to indicate that we are adding the terms that appear next to this symbol. Below the symbol $\sum$ we write the counting index i: it tells us at what value our sum starts. Above the

[59] In fact, one could argue that it makes more sense to begin with series and then consider integrals as a continuous version of a series

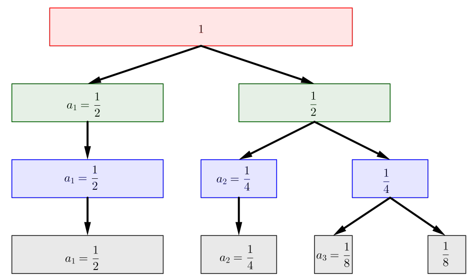


Figure 162: Bisecting a rod of length 1

[60] This is reminiscent of Zeno’s paradoxes

sum $\sum$ we write the last value that our counting index will take. For example,

$$\begin{aligned} \textstyle\sum_{i=1}^{3} (3i-1) = (3\cdot 1-1)+(3\cdot 2-1)+(3\cdot 3-1) = 15 \\ \textstyle\sum_{i=0}^{2} \cos(\pi i) = \cos(0)+\cos(\pi)+\cos(2\pi) = 1 \end{aligned} \tag{1104}$$

Our objective is to *define* the infinite sum $\sum_{i=1}^{\infty} \frac{1}{2^i}$ in such a way that we obtain

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = 1 \tag{1105}$$

which is what we would expect from our rod construction. Following the philosophy of Calculus, we will define $\sum_{i=1}^{\infty} \frac{1}{2^i}$ as a limit. More concretely, consider the **partial sums**

$$\boxed{S_n = \sum_{i=1}^{n} \frac{1}{2^i}} \tag{1106}$$

For example,

$$\begin{aligned} S_1 = \textstyle\sum_{i=1}^{1} \frac{1}{2^i} = \frac{1}{2} \\ S_2 = \textstyle\sum_{i=1}^{2} \frac{1}{2^i} = \frac{1}{2}+\frac{1}{4} = \frac{3}{4} \\ S_3 = \textstyle\sum_{i=1}^{3} \frac{1}{2^i} = \frac{1}{2}+\frac{1}{4}+\frac{1}{8} = \frac{7}{8} \end{aligned} \tag{1107}$$

We then define the series $\sum_{i=1}^{\infty} \frac{1}{2^i}$ as the limit of the partial sums, that is,

$$\boxed{\sum_{i=1}^{\infty} \frac{1}{2^i} = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{2^i}} \tag{1108}$$

From the rod model it is not difficult to see that

$$S_n = 1 - \frac{1}{2^n} \tag{1109}$$

and so

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = \lim_{n\to\infty} \left(1 - \frac{1}{2^n}\right) = 1 \tag{1110}$$

which is what we wanted to see happen.

Convergence/ Divergence of an Series: given a series

$$\sum_{i=1}^{\infty} a_i = a_1 + a_2 + \cdots \tag{1111}$$

let S_n denote its nth partial sum

$$S_n = \sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n \tag{1112}$$

If the sequence $\{S_n\}$ is convergent and

$$\lim_{n\to\infty} S_n = S \tag{1113}$$

exists as a real number then the series $\sum a_n$ is called **convergent** and we write

$$S = \sum_{i=1}^{\infty} a_i \tag{1114}$$

If the limit does not exist, we say that the series is **divergent.**

As an example of an divergent series, consider $\sum_{i=1}^{\infty}(-1)^i$. To show that this series diverges, we must analyze the partial sums. The sequence is $a_n = (-1)^n$ and the first partial sums are

$$\begin{gathered} S_1 = \textstyle\sum_{i=1}^{1}(-1)^i = -1 \\ S_2 = \textstyle\sum_{i=1}^{2}(-1)^i = (-1)^1 + (-1)^2 = -1 + 1 = 0 \\ S_3 = \textstyle\sum_{i=1}^{3}(-1)^i = -1 + 1 - 1 = -1 \\ S_4 = \textstyle\sum_{i=1}^{4}(-1)^4 = -1 + 1 - 1 + 1 = 0 \end{gathered} \tag{1115}$$

It is not difficult to extrapolate these calculations and conclude that

$$S_n = \sum_{i=1}^{n}(-1)^i = \begin{cases} -1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases} \tag{1116}$$

Therefore the partial sums keep jumping between -1 and 0 and so there is no possibility for the limit to exist, that is, the series $\sum_{i=1}^{\infty}(-1)^i$ is divergent!

In fact, it was not necessary to make such a detailed analysis to conclude that $\sum_{i=1}^{\infty}(-1)^i$ is a divergent series. The fact that the sequence $(-1)^n$ does not limit 0 is sufficient to conclude that the series diverges! To be more precise, suppose that

$$S = \sum_{i=1}^{\infty} a_i \tag{1117}$$

is a series which converges to the number S. This means that

$$S = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=1}^{n} a_i \tag{1118}$$

It is not difficult to see that

$$a_n = S_n - S_{n-1} \tag{1119}$$

and that $\lim_{n\to\infty} S_{n-1} = S$ therefore

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (S_n - S_{n-1}) = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = S - S = 0 \tag{1120}$$

This result is known as the Nth term test and we restate it together with some other properties of series.

Some properties of series:

- ***N*th Term Test:** if the series $\sum_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n\to\infty} a_n = 0$. Therefore, if $\lim_{n\to\infty} a_n \neq 0$, then the series $\sum_{n=1}^{\infty} a_n$ will be divergent.

If $\sum a_n$ and $\sum b_n$ are convergent series, then

- $\sum c a_n = c \sum a_n$ where c is a constant
- $\sum (a_n \pm b_n) = \sum a_n \pm \sum b_n$

Example 197. Does the series $\sum_{n=0}^{\infty} \cos(\frac{1}{n+3})$ converge?

If we let $a_n = \cos\left(\frac{1}{n+3}\right)$ we can see that $\lim_{n\to\infty} a_n$ does not exist, in particular, it is not 0 so the series $\sum_{n=0}^{\infty} \cos\left(\frac{1}{n+3}\right)$ does not converge by the Nth term test.

Example 198. Is $\sum_{n=0}^{\infty} \frac{n}{n+1}$ convergent or divergent?

In this case the sequence $a_n = \frac{n}{n+1}$ has limit

$$\lim_{n\to\infty} \frac{n}{n+1} = 1 \tag{1121}$$

and so by the Nth term test the series will be divergent.

It should be pointed out that the Nth term test does not imply that any series $\sum a_n$ whose nth term has limit zero is convergent. For example, the **harmonic series**

$$\sum_{n=1}^{\infty} \frac{1}{n} \tag{1122}$$

is divergent, despite the fact that $\lim_{n\to\infty} \frac{1}{n} = 0$. We will see in the next section how to show that the harmonic series is divergent, for now we will focus on finding the values of a family of series known as **geometric series**. This series is of the form

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + \cdots \tag{1123}$$

where r is a fixed number. Observe that if $|r| \geq 1$, then the series is divergent because of the N th term test. Therefore, we will focus on the case in which $|r| < 1$. Consider the partial sums

$$S_N = \sum_{n=0}^{N} r^n = 1 + r + r^2 + \cdots + r^{N-1} + r^N \tag{1124}$$

The key to computing the value of the geometric series is realizing that S_N and S_{N-1} are related not only additively, that is

$$\boxed{S_N = S_{N-1} + r^N} \tag{1125}$$

, but also multiplicatively. What we mean by this is that

$$\begin{aligned} rS_{N-1} &= r\left(1 + r + r^2 + \cdots + r^{N-2} + r^{N-1}\right) \\ &= r + r^2 + r^3 + \cdots + r^{N-1} + r^N \\ &= S_N - 1 \end{aligned} \tag{1126}$$

and so

$$\boxed{S_N = 1 + rS_{N-1}} \tag{1127}$$

If we equate the two formulas for S_N (that is, the boxed formulas) we find that

$$S_{N-1} + r^N = 1 + rS_{N-1} \tag{1128}$$

and so we can solve for S_{N-1} to conclude that

$$\boxed{S_{N-1} = \frac{1 - r^N}{1 - r}} \tag{1129}$$

If $|r| < 1$ then $\lim_{N\to\infty} r^N = 0$ and so $\lim_{N\to\infty} S_{N-1} = \frac{1}{1-r}$. Therefore, we just found the following:

Geometric Series: the geometric series

$$\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + \cdots \tag{1130}$$

is convergent if $|r| < 1$ and its sum is

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \quad |r| < 1 \tag{1131}$$

If $|r| \geq 1$, the geometric series is divergent.

Example 199. Is the series $\sum_{n=0}^{\infty} \frac{3^{n+1}}{7^n}$ convergent? If it were convergent, what is its limit?

The series $\sum_{n=0}^{\infty} \frac{3^{n+1}}{7^n}$ is almost a geometric series, in fact, we can see that

$$\sum_{n=0}^{\infty} \frac{3^{n+1}}{7^n} = \sum_{n=0}^{\infty} \frac{3 \cdot 3^n}{7^n} = 3\left(\sum_{n=0}^{\infty} \frac{3^n}{7^n}\right) = 3\left(\sum_{n=0}^{\infty} \left(\frac{3}{7}\right)^n\right) \tag{1132}$$

and $\sum_{n=0}^{\infty}\left(\frac{3}{7}\right)^n$ is a geometric series with $r=\frac{3}{7}$. Since $r<1$ the series converges to

$$\sum_{n=0}^{\infty}\left(\frac{3}{7}\right)^n=\frac{1}{1-\frac{3}{7}}=\frac{7}{4} \tag{1133}$$

Therefore, the value of the original series is

$$\sum_{n=0}^{\infty}\frac{3^{n+1}}{7^n}=3\cdot\left(\frac{7}{4}\right)=\frac{21}{4} \tag{1134}$$

Example 200. Is the series $\sum_{n=0}^{\infty}\frac{\cos(n\pi)}{2^n}$ convergent? If it were convergent, to what number would it converge?

Observe that

$$\cos(n\pi)=(-1)^n \tag{1135}$$

and so

$$\begin{aligned}\textstyle\sum_{n=0}^{\infty}\frac{\cos(n\pi)}{2^n} &= \textstyle\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n}\\ &= \textstyle\sum_{n=0}^{\infty}\left(\frac{-1}{2}\right)^n\\ &= \frac{1}{1-\left(\frac{-1}{2}\right)}\\ &= \frac{2}{3}\end{aligned} \tag{1136}$$

where we computed the series as a geometric series with $r=-\frac{1}{2}$.

Testing for Convergence/Divergence of Series

Integral Test

In general finding the value of a series $\sum_{n=1}^{\infty} a_n$ can be quite difficult. Therefore, sometimes we have to be satisfied with determining whether a series is convergent or divergent, without trying to find its value (it case it converges).

Since the definition of an integral in terms of a Riemann sums was very similar to computing a series (the areas of the rectangles approximating the area under the curve), it makes sense to try to relate series with integrals. A situation where we can find quickly a relationship is the following.

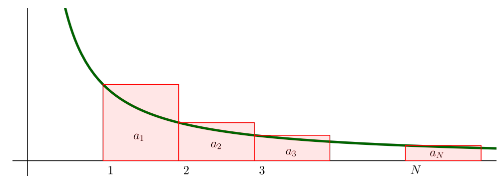


Figure 163: $\int_1^{N+1} f(x)dx \leq S_N$

Suppose that $f(x)$ is a continuous, positive and decreasing function defined (at least) for $x \geq 1$. If $\sum_{n=1}^{\infty} a_n$ is a series for which $a_n = f(n)$ then we can think of each term a_n as a rectangle of width 1 and height $f(n)$, as the images show. This resembles the Riemann sums we used to define $\int_a^b f(x)dx$, except that in this case we need to work with the improper integral $\int_1^{\infty} f(x)dx$. In fact, if we use the left-endpoint and right endpoint Riemann sums for the definition of the integral it is not difficult to see that for all $N \geq 1$

$$S_{N-1} - a_1 \leq \int_1^{N+1} f(x)dx \leq S_N \tag{1137}$$

and so if $\int_1^{\infty} f(x)dx$ converges the first inequality says (taking $N \to \infty$) that $\sum_{n=1}^{\infty} a_n \leq \int_1^{\infty} f(x)dx + a_1$ so that the series is convergent. Conversely, if $\int_1^{\infty} f(x)dx$ diverges the second inequality says (taking $N \to \infty$) that $\int_1^{\infty} f(x)dx \leq \sum_{n=1}^{\infty} a_n$. Therefore, it is not difficult to see the following:

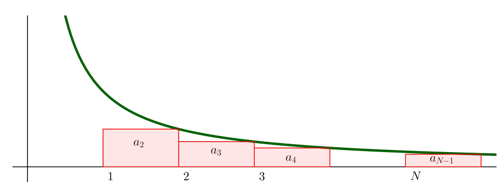


Figure 164: $S_{N-1} - a_1 \leq \int_1^{N+1} f(x)dx$

Integral Test: if $a_n = f(n)$ where f is positive valued, continuous and decreasing for $x \geq 1$ then

$$\int_1^{\infty} f(x)dx \text{ and } \sum_{n=1}^{\infty} a_n \tag{1138}$$

either both converge or both diverge.

Example 201. Show that the p series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is convergent if $p > 1$ and divergent if $0 < p \leq 1$.

Consider the function $f(x) = \frac{1}{x^p} = x^{-p}$. Observe that f is positive, is defined for $x \geq 1$ and it is decreasing because

$$f'(x) = -px^{-p-1} \tag{1139}$$

is negative on $[1, \infty)$. Therefore, by the integral test, the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges when $\int_1^{\infty} \frac{dx}{x^p}$ and diverges when $\int_1^{\infty} \frac{dx}{x^p}$ diverges. In the section on improper integrals we found that $\int_1^{\infty} \frac{dx}{x^p}$ converges for $p > 1$ and diverges for $0 < p \leq 1$ so the result follows.

Example 202. Show that $\sum_{n=3}^{\infty} \frac{\ln n}{n^2}$ is divergent.

Consider the function $f(x) = \frac{\ln x}{x^2}$. It is positive, defined for $x \geq 3$ and decreasing because

$$f'(x) = \frac{x - 2x\ln x}{x^4} = \frac{1 - 2\ln x}{x^3} \tag{1140}$$

is negative [61]. Therefore the divergence/convergence of the series $\sum_{n=3}^{\infty} \frac{\ln n}{n^2}$ is equivalent to the divergence/convergence of the integral $\int_3^{\infty} \frac{\ln x}{x^2} dx$. Using the substitution $u = \ln x$, $du = \frac{dx}{x}$ the integral is equivalent to

[61] Observe that since $e \leq x$ then $1 = \ln e \leq \ln x$ and so $1 - 2\ln x$ is negative.

$$\int_3^{\infty} \frac{\ln x}{x^2} dx = \int_3^{\infty} \frac{\ln x}{x} \frac{dx}{x} = \int_{\ln 3}^{\infty} \frac{u}{e^u} du = \int_{\ln 3}^{\infty} ue^{-u} du \tag{1141}$$

Using integration by parts we have $\int ue^{-u} du = -e^{-u}u - e^{-u}$ and so the improper integral is convergent. Therefore the series $\sum_{n=3}^{\infty} \frac{\ln n}{n^2}$ is convergent.

Example 203. Does the series $\sum_{n=2}^{\infty} \frac{1}{n\sqrt{\ln n}}$ converge?

Define $f(x) = \frac{1}{x\sqrt{\ln x}}$ for $x \geq 2$. Clearly $a_n = f(n)$ and f is positive valued, continuous and decreasing. So we can use the integral test to determine if the series converges or diverges. The integral in this case is

$$\int_2^{\infty} \frac{dx}{x\sqrt{\ln x}} \tag{1142}$$

and if we make the substitution $u = \ln x$, $du = \frac{1}{x} dx$ this integral is the same as

$$\int_{\ln(2)}^{\infty} \frac{du}{\sqrt{u}} = 2\sqrt{u}\Big|_{u=\ln 2}^{u=\infty} \tag{1143}$$

and this integral diverges. Therefore, the original series diverges.

Comparison Test

If we are only interested in deciding whether a series converges or diverges, it is also reasonable to try to compare the series with another series to determine if it converges or diverges. The easiest method that compares two series is the **comparison test**:

Comparison Test: suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms.

1. If $\sum b_n$ is convergent and $a_n \leq b_n$ for all n bigger than some index n_0, then $\sum a_n$ is also convergent.
2. If $\sum b_n$ is divergent and $a_n \geq b_n$ for all n bigger than some index n_0, then $\sum a_n$ is also divergent.

In order to use the comparison test, we need to know some basic facts about the growth of functions:

For two sequences $\{a_n\}, \{b_n\}$, write

$$a_n \prec b_n \tag{1144}$$

to mean

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0 \tag{1145}$$

We say that a_n **grows substantially slower** than b_n .

The notation $a_n \prec b_n$ does not mean that $a_n < b_n$ for all n. Maybe some initial terms in the a_n sequence are larger than the corresponding ones in the b_n sequence, but this will eventually stop and the long term growth of b_n dominates. For example, if $a_n = 100n$ and $b_n = n^2$ then

$$a_n \prec b_n \tag{1146}$$

because

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{100n}{n^2} = \lim_{n\to\infty} \frac{100}{n} = 0 \tag{1147}$$

This means that after a while, every element in the sequence b_n will be bigger than every element in the sequence a_n, despite the fact that at the beginning a_n might start out being bigger. The growth of functions we will mainly use are the following ones (there is some repetition in this list but we do this for emphasis).

- For each $r > 0$, $n^r \prec e^n$
- For each $r > 0$, $\ln n \prec n^r$
- For any polynomial $p(x)$, $p(n) \prec e^n$
- For any $r > 0$ and $k > 0$, $(\ln n)^k \prec n^r$
- For any non-constant polynomial $p(x)$ and $k > 0$, $(\ln n)^k \prec p(n)$
- For $a, b > 1$, $r > 0$, $\log_b n \prec n^r \prec a^n$

Example 204. Determine if the following series converge or diverge:

a) $\sum_{n=1}^{\infty} \frac{1+n^2}{n^3+n^2}$

The series is a series of positive terms. For large n, the leading term in the numerator is n^2 and the leading term in the denominator is n^3 so we expect $\frac{1+n^2}{n^3+n^2}$ to behave like $\frac{n^2}{n^3} = \frac{1}{n}$. Therefore, we expect the series to diverge. Therefore we would like to compare this series with a smaller one that diverges. To be more precise, we can say that

$$\frac{n^2}{2} \prec 1 + n^2 \tag{1148}$$

and so

$$\frac{n^2}{2(n^3+n^2)} \prec \frac{1+n^2}{n^3+n^2} \tag{1149}$$

Since

$$n^3 + n^2 \prec 2n^3 \implies \frac{1}{2n^3} \prec \frac{1}{n^3+n^2} \implies \frac{n^2}{4n^3} \prec \frac{n^2}{2(n^3+n^2)} \tag{1150}$$

So we can say that

$$\frac{1}{4n} = \frac{n^2}{4n^3} \prec \frac{1+n^2}{n^3+n^2} \tag{1151}$$

and because $\sum_{n=1}^{\infty} \frac{1}{4n}$ diverges, we conclude that $\sum_{n=1}^{\infty} \frac{1+n^2}{n^3+n^2}$ diverges.

b) $\sum_{n=1}^{\infty} \frac{n+\sqrt[3]{n}}{n^{7/2}+n^2}$

First of all, the series is a series of positive terms. Also, observe that for n large, the leading term in the numerator is n and the leading term in the denominator is $n^{7/2}$, so we expect $\frac{n+\sqrt[3]{n}}{n^{7/2}+n^2}$ to behave like $\frac{n}{n^{7/2}} = \frac{1}{n^{5/2}}$ so we expect the series to converge. To be more precise, we use the comparison test. For n large, it is clear that

$$n + \sqrt[3]{n} \leq 2n \tag{1152}$$

and so

$$\frac{n+\sqrt[3]{n}}{n^{7/2}+n^2} \leq \frac{2n}{n^{7/2}+n^2} \tag{1153}$$

Now, it is also clear that

$$\begin{aligned} & n^{7/2} \leq n^{7/2} + n^2 \\ \implies \quad & \tfrac{1}{n^{7/2}+n^2} \leq \tfrac{1}{n^{7/2}} \\ \implies \quad & \tfrac{2n}{n^{7/2}+n^2} \leq \tfrac{2n}{n^{7/2}} \end{aligned} \tag{1154}$$

Putting these two inequalities together, we conclude that for n sufficiently large

$$\frac{n+\sqrt[3]{n}}{n^{7/2}+n^2} \leq \frac{2n}{n^{7/2}} = \frac{2}{n^{5/2}} \tag{1155}$$

That is,

$$\frac{n+\sqrt[3]{n}}{n^{7/2}+n^2} \prec \frac{2}{n^{5/2}} \tag{1156}$$

Since $\sum_{n=1}^{\infty} \frac{2}{n^{5/2}}$ is convergent because $5/2 > 1$, we conclude that $\sum_{n=1}^{\infty} \frac{n+\sqrt[3]{n}}{n^{7/2}+n^2}$ is convergent as well.

Example 205. Determine if the following series converge or diverge:
a) $\sum_{n=2}^{\infty} \frac{1}{\ln(\ln n)}$
First of all, we have that

$$\ln n \leq n \tag{1157}$$

and because ln is a increasing function we have that

$$\ln(\ln n) \leq \ln(n) \tag{1158}$$

Therefore

$$\frac{1}{\ln(n)} \leq \frac{1}{\ln(\ln(n))} \tag{1159}$$

Since $\sum_{n=2}^{\infty} \frac{1}{\ln n}$ diverges (you can compare it with $\sum_{n=2}^{\infty} \frac{1}{n}$ to verify this!) we conclude that $\sum_{n=2}^{\infty} \frac{1}{\ln(\ln(n))}$ must diverge as well.
b) $\sum_{n=1}^{\infty} \frac{(\ln n)^2}{n^2}$
In this case we can use that for large n

$$\ln n \leq n^{1/4} \tag{1160}$$

This can be seen because

$$(x^{1/4} - \ln x)' = \frac{1}{4}x^{1/4-1} - \frac{1}{x} = \frac{x^{1/4} - 4}{4x} \tag{1161}$$

and so for large x, the function $x^{1/4} - \ln x$ is increasing. It will also be eventually positive since by L'Hospital

$$\lim_{x\to\infty} \frac{x^{1/4}}{\ln x} = \lim_{x\to\infty} \frac{1}{4}x^{1/4} = \infty \tag{1162}$$

Dividing by n we conclude that

$$\frac{\ln n}{n} \leq \frac{1}{n^{3/4}} \tag{1163}$$

and squaring both sides of the inequality

$$\left(\frac{\ln n}{n}\right)^2 \leq \frac{1}{n^{3/2}} \tag{1164}$$

since $\sum_{n=1}^{\infty} \frac{1}{n^{3/2}}$ converges (it is a p series with $p = 3/2 > 1$) we conclude that $\sum_{n=1}^{\infty} \frac{(\ln n)^2}{n^2}$ converges.

Power and Taylor Series

We saw earlier that if $|r| < 1$, then the power series $\sum_{n=0}^{\infty} r^n$ converges to the number $\frac{1}{1-r}$. If we change (for psychological reasons) r to x then for $x \in (-1, 1)$ we have the formula

$$\boxed{\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n} \tag{1165}$$

Now, the left hand side defines a function $f(x) = \frac{1}{1-x}$ and so the previous identity says that

$$f(x) = \lim_{N\to\infty} P_N(x) \tag{1166}$$

where $P_N(x)$ represents the polynomial

$$P_N(x) = \sum_{n=0}^{N} x^n = 1 + x + x^2 + \cdots + x^N \tag{1167}$$

This innocent looking result is actually one of the most important results of all Calculus, since it shows that a function (in this case $\frac{1}{1-x}$) can be approximated by polynomials! Since polynomials are easier to understand than arbitrary functions, this opens the door to understanding arbitrary [62] functions by studying polynomials.

[62] In fact the functions must satisfy certain conditions as we are about to see.

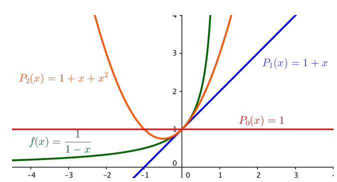


Figure 165: Approximating $\frac{1}{1-x}$ by polynomials

In general, we say that an expression of the form

$$\sum_{n=0}^{\infty} a_n(x-a)^n \tag{1168}$$

is a **power series** centered at a. Here the a_n are constants, and we would like to consider the power series as defining a function $f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$. We would then like to find the domain of f , that is, for which values of x does the (power) series converge. Interestingly enough, the domain of f is of three types:

a) either it is the single point $\{a\}$, which happens for example when we take $a_n = n^n$ and $a = 0$

b) the infinite interval $(-\infty, \infty)$, which happens for example when we take $a_n = 0$ for all n , or

c) an interval centered at a of the form $(a - R, a + R)$, where the endpoints could be included or excluded in the domain. This is the case of the previous geometric series $\sum_{n=0}^{\infty} x^n$ where $a_n = 1$ and $a = 0$. The number R is known as the **radius of convergence**.

The following test, known as the **ratio test**, gives a criteria to determine which of the three cases we are working with.

Ratio Test for Convergence of Power Series : Suppose we are given the power series

$$\sum_{n=0}^{\infty} a_n(x-a)^n \tag{1169}$$

Let

$$R = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right| \tag{1170}$$

1. If $R = 0$ then the series converges only for $x = a$,
2. If $0 < R < \infty$ the series converges for $x \in (a-R, a+R)$ and diverges for x outside the interval $[a-R, a+R]$. The convergence or divergence at the endpoints of the interval must be determined by alternative methods.
3. If $R = \infty$ the series converges for all x

Example 206. Find the largest interval of convergence of the Bessel function of order 0 defined by

$$J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n}(n!)^2} \tag{1171}$$

Let

$$a_n = \frac{(-1)^n x^{2n}}{2^{2n}(n!)^2} \tag{1172}$$

Then

$$\begin{aligned}
\left|\frac{a_{n+1}}{a_n}\right| &= \left|\frac{\frac{(-1)^{n+1}x^{2(n+1)}}{2^{2(n+1)}((n+1)!)^2}}{\frac{(-1)^n x^{2n}}{2^{2n}(n!)^2}}\right| \\
&= \left|\frac{(-1)^n(-1)x^{2n}x^2 2^{2n}(n!)^2}{2^{2n}2^2(n+1)^2(n!)^2(-1)^n x^{2n}}\right| \\
&= \left|\frac{(-1)x^2}{4(n+1)^2}\right| \\
&= \frac{x^2}{4(n+1)^2}
\end{aligned} \tag{1173}$$

Clearly $\lim_{n\to\infty} \frac{x^2}{4(n+1)^2} = 0$ and so by the ratio test the series converges for all values of x.

Example 207. Find the largest interval around the origin where the power series $\sum_{n=0}^{\infty} \frac{x^n}{\sqrt{n^2+1}}$ is convergent

Let $a_n = \frac{x^n}{\sqrt{n^2+1}}$. To use the ratio test we find $\left|\frac{a_{n+1}}{a_n}\right|$

$$\begin{aligned}
\left|\frac{a_{n+1}}{a_n}\right| &= \left|\frac{\frac{x^{n+1}}{\sqrt{(n+1)^2+1}}}{\frac{x^n}{\sqrt{n^2+1}}}\right| \\
&= \left|\frac{x\sqrt{n^2+1}}{\sqrt{(n+1)^2+1}}\right| \\
&= |x|\sqrt{\frac{n^2+1}{n^2+2n+2}} \\
&= |x|\sqrt{\frac{1+\frac{1}{n^2}}{1+\frac{2}{n}+\frac{2}{n^2}}}
\end{aligned} \tag{1174}$$

Clearly $\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = |x|$ so we need $|x| < 1$ to guarantee convergence if we want to use the ratio test. Therefore, the series converges inside the interval $(-1, 1)$.

Example 208. Express $\frac{x^5}{7-9x^3}$ as a power series and find the interval of convergence

We have

$$\frac{x^5}{7-9x^3} = \frac{x^5}{7\left(1-\frac{9x^3}{7}\right)} = \frac{x^5}{7}\frac{1}{1-\frac{9x^3}{7}} \tag{1175}$$

Now put $r = \frac{9x^3}{7}$ and comparing the series with $\frac{1}{1-r} = 1 + r + r^2 + \cdots$ for $|r| < 1$ we get

$$\frac{1}{1-\frac{9x^3}{7}} = \sum_{n=0}^{\infty}\left(\frac{9x^3}{7}\right)^n \tag{1176}$$

This expansion is valid whenever $\left|\frac{9x^3}{7}\right| < 1$ so $-\sqrt[3]{\frac{7}{9}} < x < \sqrt[3]{\frac{7}{9}}$. Therefore

$$\frac{x^5}{7-9x^3} = \frac{x^5}{7}\sum_{n=0}^{\infty}\left(\frac{9x^3}{7}\right)^n = \sum_{n=0}^{\infty}\frac{9^n x^{3n+5}}{7^{n+1}} \tag{1177}$$

Power series have very nice properties as functions. In fact, for any point in their domain, the power series $\sum_{n=0}^{\infty} c_n(x-a)^n$ can be differentiated as often as we want and integrated as often as we want. Moreover, the derivative and integral can be found in the naive way, that is, differentiating or integrating individually each term of the power series. [63]

[63] We should point out that in general one must be careful when differentiating or integrating an infinite sum. For example, $f(x) = \sum_{n=1}^{\infty}\frac{1}{2^n}\cos(3^n x)$ defines a continuous function on all of $\mathbb{R}$ which is not differentiable at any point of $\mathbb{R}$, despite the fact that any finite approximation $\sum_{n=1}^{N}\frac{1}{2^n}\cos(3^n x)$ can be differentiated as often as we want!

Differentiating and Integrating Power Series:
If the power series

$$\sum_{n=0}^{\infty} c_n(x-a)^n \tag{1178}$$

has radius of convergence $R > 0$ then the function f defined by

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n \tag{1179}$$

is differentiable (and therefore continuous) on the interval $(a-R, a+R)$ and

1. $f'(x) = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1}$
2. $\int f(x)dx = C + \sum_{n=0}^{\infty} c_n \frac{(x-a)^{n+1}}{n+1}$, where C is the constant of integration.

The radii of convergence of the power series are both R.

Example 209. Find a power series representation for the given function and determine the radius of convergence: $f(x) = \frac{1}{(3+x)^2}$
We use the fact that

$$f(x) = \frac{d}{dx}\int f(x)dx \tag{1180}$$

The antiderivative of $f(x)$ is easy to find:

$$\begin{aligned}
\int f(x)dx &= \int \frac{dx}{(3+x)^2} \\
&= -\frac{1}{3+x} + C \\
&= -\frac{1}{3}\left(\frac{1}{1+\frac{x}{3}}\right) + C \\
&= -\frac{1}{3}\sum_{n=0}^{\infty}\left(-\frac{x}{3}\right)^n + C \\
&= \sum_{n=0}^{\infty}\left(-\frac{1}{3}\right)^{n+1} x^n + C
\end{aligned} \tag{1181}$$

Notice that we can only use the previous expansion whenever $\left|-\frac{x}{3}\right| < 1$ which means $-3 < x < 3$. In that interval we have that

$$f(x) = \frac{d}{dx}\left(\sum_{n=0}^{\infty}\left(-\frac{1}{3}\right)^{n+1} x^n + C\right) = \sum_{n=1}^{\infty}\left(-\frac{1}{3}\right)^{n+1} nx^{n-1} \tag{1182}$$

So far we have seen some examples of how a given function can be approximated by a power series. For example, $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$ if $x \in (-1,1)$ and $\frac{1}{(3+x)^2} = \sum_{n=1}^{\infty}\left(-\frac{1}{3}\right)^{n+1} nx^{n-1}$ if $x \in (-3,3)$. Now we would like to have a more systematic method to find such expansions of f as a power series, in particular, we would like to be able to find the coefficients a_n in the power series expansion $f(x) = \sum_{n=0}^{\infty} a_n x^n$.

Suppose for the time being that f can be differentiated as many times as we please. Assuming that we are able to write f as a power series, that is

$$\boxed{f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots +} \qquad (1183)$$

It is not hard to see that if we take $x = 0$ in the above equation then

$$\boxed{f(0) = a_0} \qquad (1184)$$

and so we have found the meaning of the first coefficient in the power series expansion, it is just the value of f at the origin! Since a power series can be differentiate term by term, if we differentiate 1183 we obtain

$$\boxed{f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + \cdots} \qquad (1185)$$

and again if we evaluate at 0 we obtain

$$\boxed{f'(0) = a_1} \qquad (1186)$$

If we continue in this way and differentiate the equation for $f''(x)$ we find

$$\boxed{f''(x) = \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = 2a_2 + 6a_3 x + \cdots +} \qquad (1187)$$

and again the same trick gives us

$$\boxed{f''(0) = 2a_2} \qquad (1188)$$

and at this point it is not difficult to see that if we continue this process then

$$\boxed{a_n = \frac{f^{(n)}(0)}{n!}} \qquad (1189)$$

Therefore, we see that the coefficients in the power series expansion for f are determined by the values of the derivatives of f at the origin. Now, there was nothing special about the origin, so we just found the following result [64]

•

[64] Strictly speaking we provided a heuristic argument. The full proof can be obtained using the Rolle's theorem and some ingenuity.

Taylor Series Representation of a Function: if a function f has derivatives of all orders in an open interval $I = (a-R, a+R)$ centered at $x = a$ then

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{1190}$$

if and only if

$$\lim_{N\to\infty} R_N(x) = 0 \tag{1191}$$

for all $x \in I$, where $R_N(x)$ is the remainder

$$R_N(x) = f(x) - P_N(x) \tag{1192}$$

and $P_N(x)$ the Nth **Taylor polynomial**

$$P_N(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{1193}$$

When the Taylor series of f is expanded about $a = 0$, we call the series a **Maclaurin series** of f. The following table has the Maclaurin series of the most used functions in Calculus: they can be computed using formula 1190.

Maclaurin Series	Interval
$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$	$(-1,1)$
$e^x = \sum_{n=0}^{\infty} x^n$	$(-\infty,\infty)$
$\cos(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{(2n)!}$	$(-\infty,\infty)$
$\sin(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{(2n+1)!}$	$(-\infty,\infty)$
$\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1} \frac{x^n}{n}$	$-1 < x \leq 1$
$(1+x)^k = \sum_{n=0}^{\infty} \left(\begin{array}{c} k \\ n \end{array} \right) x^n = 1 + kx + \frac{k(k-1)}{2!}x^2 + \cdots +$	$(-1,1)$

It is also important to notice that it is best to think of a function as an entity which exists independently of its Taylor expansion, and that for certain regions of its domain it agrees with such expansion.

For example, the function $f(x) = \frac{1}{1-x}$ has domain $\mathbb{R}\backslash\{1\}$. If we just consider the interval $(-1,1)$, then f agrees with the power series $\sum_{n=0}^{\infty} x^n$, however, outside this domain they don't agree, in fact, the power series diverges outside $(-1,1)$! Therefore, if we want to study f as a power series on a different interval, for example the interval $(1,5)$ we must expand f about a different point. As a matter of fact, if we take $a = 3$, then on $(1,5)$ we can say that

$$\frac{1}{1-x} = -\frac{1}{2} + \frac{x-3}{4} - \frac{1}{8}(x-3)^2 + \frac{1}{16}(x-3)^3 + \cdots + \tag{1194}$$

and we can easily see that this expansion looks very different from the one that works on the interval $(-1,1)$.

Example 210. Find the Taylor series around the origin (Maclaurin series) of $f(x) = e^{-x^2}$ and use it to evaluate $\int e^{-x^2}\,dx$

From the expansion for e^x

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{1195}$$

we find that

$$e^{-x^2} = \sum_{n=0}^{\infty} \frac{(-x^2)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{n!} \tag{1196}$$

The indefinite integral can be found integrating term by term

$$\begin{aligned} \int e^{-x^2}\,dx &= \int \textstyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{n!}\,dx \\ &= \textstyle\sum_{n=0}^{\infty} \int \frac{(-1)^n x^{2n}}{n!}\,dx \\ &= \textstyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!(n+1)} + C \\ &= \textstyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(n+1)!} \end{aligned} \tag{1197}$$

Example 211. Use the fact that $\frac{1}{1+x} = \sum_{n=0}^{\infty}(-1)^n x^n$, $-1 < x < 1$ in order to obtain the Taylor series of $\ln(1+x)$. Then find the Taylor series for $\ln(3+4x)$ about $x = 0$

Integrating $\frac{1}{1+x} = \sum_{n=0}^{\infty}(-1)^n x^n$ we find that

$$\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} + C \tag{1198}$$

to find the value of the constant C observe that when $x = 0$ we should obtain

$$\ln(1) = C \tag{1199}$$

so $C = 0$. To find the Taylor series for $\ln(3+4x)$ then

$$\begin{aligned} \ln(3+4x) &= \ln\left(3\left(1+\tfrac{4x}{3}\right)\right) \\ &= \ln 3 + \ln\left(1+\tfrac{4x}{3}\right) \\ &= \ln 3 + \textstyle\sum_{n=0}^{\infty} \frac{(-1)^n}{n+1}\left(\frac{4x}{3}\right)^{n+1} \end{aligned} \tag{1200}$$

Notice that this series converges whenever $-1 < \frac{4x}{3} < 1$, that is, $-\frac{3}{4} < x < \frac{3}{4}$.

Example 212. Find the Taylor series for $f(x) = \frac{7}{x^4}$ about $x = -3$

Write $f(x) = 7x^{-4}$ and we start finding the derivatives of f:

$$\begin{array}{lc} n=0 & f'(x) = 7x^{-4} \\ n=1 & f'(x) = -7(4)x^{-5} \\ n=2 & f''(x) = 7(4)(5)x^{-6} \\ n=3 & f^{(3)}(x) = -7(4)(5)(6)x^{-7} \\ n=4 & f^{(4)}(x) = 7(4)(5)(6)(7)x^{-8} \\ \vdots & \end{array} \tag{1201}$$

To see the pattern notice that we can rewrite the derivatives as

$$\begin{array}{lc} n=0 & f(x) = -\frac{7}{6}(3!)x^{-4} \\ n=1 & f'(x) = -\frac{7}{6}(4!)x^{-5} \\ n=2 & f''(x) = \frac{7}{6}(5!)x^{-6} \\ n=3 & f^{(3)}(x) = -\frac{7}{6}(6!)x^{-7} \\ n=4 & f^{(4)}(x) = \frac{7}{6}(7!)x^{-8} \\ \vdots & \end{array} \tag{1202}$$

It is now clear that

$$f^{(n)}(x) = \frac{7}{6}(-1)^n(n+3)!x^{-(n+4)} \tag{1203}$$

What we actually need are the values of the derivatives at $x = -3$:

$$f^{(n)}(-3) = \frac{7(-1)^n(n+3)!}{6(-3)^{n+4}} = \frac{7(n+3)!}{6(3)^{n+4}} \tag{1204}$$

Therefore the Taylor series for $\frac{7}{x^4}$ is

$$\frac{7}{x^4} = \sum_{n=0}^{\infty} \frac{f^{(n)}(-3)}{n!}(x+3)^n = \sum_{n=0}^{\infty} \frac{7(n+3)(n+2)(n+1)}{6(3)^{n+4}}(x+3)^n \tag{1205}$$

Example 213. Find the Taylor series of $f(x) = 7x^2 - 6x + 1$ about $x = 2$

We find the derivatives of $f(x)$:

$$\begin{array}{lc} n=0 & f(x) = 7x^2 - 6x + 1 \\ n=1 & f'(x) = 14x - 6 \\ n=2 & f''(x) = 14 \\ n\geq 3 & f^{(n)}(x) = 0 \end{array} \tag{1206}$$

In this case there is no need to find a general formula for the n th derivative since the Taylor series will actually be a polynomial (our function is a polynomial after all!). We find that

$$f(2) = 17 \qquad f'(2) = 22 \qquad f''(2) = 14 \qquad f^{(n)}(2) = 0 \tag{1207}$$

Therefore

$$\begin{aligned} 7x^2 - 6x + 1 &= \sum_{n=0}^{\infty} \frac{f^{(n)}(2)}{n!}(x-2)^n \\ &= f(2) + f'(2)(x-2) + \frac{1}{2}f''(2)(x-2)^2 \\ &= 17 + 22(x-2) + 7(x-2)^2 \end{aligned} \tag{1208}$$

Example 214. Find the value of the series $\sum_{n=0}^{\infty} \frac{1}{n!}$

Since $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ we see that

$$e = e^1 = \sum_{n=0}^{\infty} \frac{1}{n!} \tag{1209}$$

Example 215. The force exerted by the Earth on a point of mass m located at a distance x above it is given by $F(x) = \frac{GMm}{(R+x)^2}$ where R is the radius of the Earth. At $x = 0$ the force is mg. Show that is $x \ll R$ then $F(x) \simeq mg(1 - 2\frac{x}{R})$

In this case we use the binomial expansion

$$\begin{aligned} F(x) &= \frac{GMm}{(R+x)^2} \\ &= \frac{GMm}{\left(R\left(1+\frac{x}{R}\right)\right)^2} \\ &= \frac{GMm}{R^2\left(1+\frac{x}{R}\right)^2} \\ &= \frac{GMm}{R^2}\left(1+\frac{x}{R}\right)^{-2} \\ &\simeq \frac{GMm}{R^2}\left(1 - 2\frac{x}{R}\right) \\ &= mg(1 - 2\frac{x}{R}) \end{aligned} \tag{1210}$$

Part IX

Multivariable Calculus

Functions of Two Variables

Up to this point, we have considered functions of one variable, like $y(x) = x^2 + 1$, $y(x) = \sin x - 2$, etc. In this case we have seen that from the geometric point of view we can represent the graph of the function as a curve in the xy plane.

Now we want to extend the methods of calculus to functions of more than one variable. In particular, we will focus on functions of two variables, for example $z(x, y) = x^2 - 3xy + 1$, $z(x, y) = x\cos(x + y) + 1$, etc.

Our first objective is to give a geometric interpretation to the previous functions. In the case of functions of one variables, we needed an extra axis in order to be able to draw the graph of the function. Following the analogy with the function of one variable, we should use an extra axis to draw the graph of the function. Since we already need two axes for the domain of our function $z(x, y)$, one for the variable x and another for the variable y, we require a third axis which we naturally call the z axis. Therefore, we work with the **three dimensional Cartesian coordinate system,** that is, a coordinate system formed by three mutually perpendicular axes. The recipe to find the graph of the function $z(x, y)$ is the following.

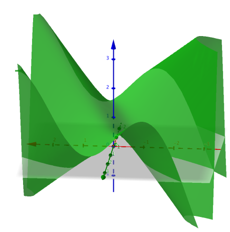


Figure 166: Graph of $z(x, y) = x\cos(x + y) - 1$

Finding the graph of a function $z(x, y)$:

1. Draw the cartesian coordinate system xyz: its points are triples (x, y, z)
2. For each point $(x, y, 0)$, that is, a point (x, y) on the xy plane at height 0, find the value $z(x, y)$ and draw the point $(x, y, z(x, y))$
3. The graph of $z(x, y)$ is the surface obtained by applying the procedure in step 2 to each point in the domain of $z(x, y)$

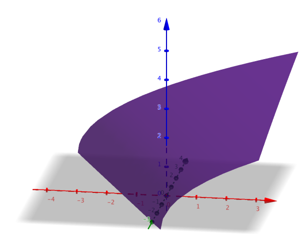


Figure 167: Graph of $z(x, y) = \sqrt{2x + y + 3}$

Example 216. Find the domain of the function $z(x, y) = \sqrt{2x + y + 3}$ and represent it on the xy plane.

Since we can only take square roots of nonnegative numbers we require

$$2x + y + 3 \geq 0 \tag{1211}$$

To represent the points (x, y) that satisfy this inequality we represent first the equation

$$2x + y + 3 = 0 \tag{1212}$$

on the xy plane. This is the equation of the straight line

$$y = -2x - 3 \tag{1213}$$

and it separates the xy plane into two regions. All the points that satisfy the inequality correspond to only one of the regions. To determine which region satisfies the inequality choose a random point, for example $(0, 0)$. It is easy to see that $(0, 0)$ satisfies the inequality $2x + y + 3 \geq 0$ so the region that works must contain the origin as shown in the next figure.

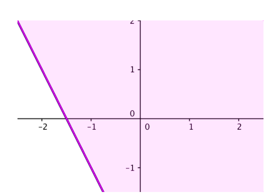


Figure 168: Points (x, y) that satisfy the inequality $2x + y + 3 \geq 0$

Example 217. Find the domain of $f(x,y) = \frac{\sqrt{4-x^2-y^2}}{x-y}$ and represent it on the xy plane.

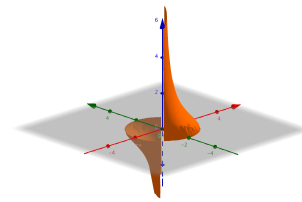


Figure 169: Graph of $f(x,y) = \sqrt{4-x^2-y^2}$

In this case we need the expression inside the square root to be non-negative, that is,

$$4 - x^2 - y^2 \geq 0 \tag{1214}$$

Again to plot the inequality on the xy plane we fist plot the equality which is

$$4 - x^2 - y^2 = 0 \implies x^2 + y^2 = 4 \tag{1215}$$

and this is the equation of a circle of radius 2 centered at the origin. Again, the circle divides the xy plane into two regions and all the points satisfying the inequality belong to one of the regions. It is easy to see that $(0,0)$ satisfies the inequality $4 - x^2 - y^2 \geq 0$, so the interior of the circle must correspond to the region which satisfies the inequality.

Also, we need the denominator to be different from zero, that is

$$x - y \neq 0 \tag{1216}$$

This means that the domain cannot include the equation

$$x = y \tag{1217}$$

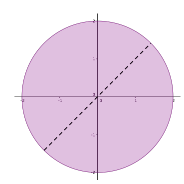


Figure 170: Domain of $f(x,y) = \frac{\sqrt{4-x^2-y^2}}{x-y}$

And so the domain must be all points (x,y) which belong to the disk of radius 2 centered at the origin but which do not lie on the straight line $x = y$.

In general, finding the graph of a function of two variables is not easy and before the age of computers, which has certainly facilitated representing functions, people had to come up with useful ways for studying functions of two variables.

One of such methods is to study the graph of the function by drawing its **level curves.** We will explain the idea behind it with an example. Consider the graph of the function $z(x,y) = x^2 + y^2$. We already know the graph will be represented by a surface, and to determine its shape we consider the intersection of the surface with planes parallel to the xy plane. The equation of a plane parallel to the xy plane is $z = c$, because it consists of all points (x,y,z) whose height z equals c. For example, the equation $z = 1$ represents the plane parallel to the xy plane of height 1 and the intersection of the graph of the surface $z = x^2 + y^2$ with the plane $z = 1$ consists of all the points satisfying the equations

$$\begin{cases} z = x^2 + y^2 \\ z = 1 \end{cases} \tag{1218}$$

Substituting the second equation into the first one we obtain the equation

$$x^2 + y^2 = 1 \tag{1219}$$

which is simply the equation of a circle centered at the point $(0,0,1)$ of radius 1. The corresponding level curve is the "shadow" this curve makes on the xy plane, which is the circle of radius 1 centered at the origin.

Similarly, the intersection of the graph of the surface $z = x^2 + y^2$ with the plane $z = 4$ consists of all the points satisfying the equations

$$\begin{cases} z = x^2 + y^2 \\ z = 4 \end{cases} \tag{1220}$$

Substituting the second equation into the first one we obtain the equation

$$x^2 + y^2 = 4 \tag{1221}$$

which is simply the equation of a circle centered at the point $(0,0,4)$ of radius 2. The corresponding level curve is the "shadow" this curve makes on the xy plane, which is the circle of radius 2 centered at the origin.

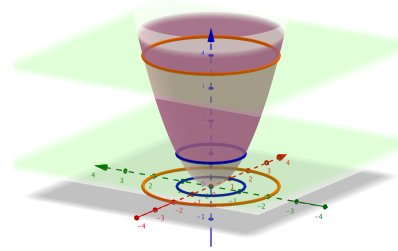


Figure 171: Level curves of $z(x,y) = x^2 + y^2$

Finding Level Curves: Suppose that $z(x,y)$ is a function of two variables x and y. If c is some value of the function z , then the equation

$$z(x,y) = c \tag{1222}$$

describes a curve lying on the plane $z = c$ called the **trace** of the graph of z in the plane $z = c$.
If this trace is projected onto the xy plane, the resulting curve in the xy plane is called a **level curve.** By drawing the level curves corresponding to several admissible values of c, we obtain a **contour map**.

Example 218. Sketch the level curves of $f(x,y) = \frac{y}{x^2+1}$ corresponding to $z = -1, 0, 1$.

The intersection of $z = f(x,y) = \frac{y}{x^2+1}$ with the plane $z = -1$ are the points satisfying the equation

$$-1 = \frac{y}{x^2+1} \implies -x^2 - 1 = y \tag{1223}$$

which is the equation of a parabola.

The intersection of $z = f(x,y) = \frac{y}{x^2+1}$ with the plane $z = 0$ are the points satisfying the equation

$$0 = \frac{y}{x^2+1} \implies 0 = y \tag{1224}$$

which is the equation of the x axis.

The intersection of $z = f(x,y) = \frac{y}{x^2+1}$ with the plane $z = 1$ are the points satisfying the equation

$$1 = \frac{y}{x^2+1} \implies x^2 + 1 = y \tag{1225}$$

which is the equation of a parabola.

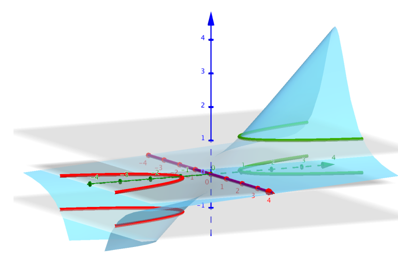


Figure 172: Graph of $f(x,y) = \frac{y}{x^2+1}$

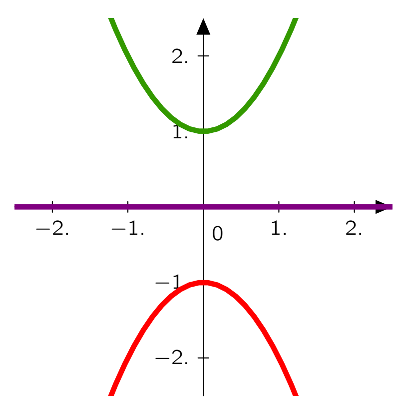


Figure 173: Level curves of $f(x,y) = \frac{y}{x^2+1}$ corresponding to $z = -1, 0, 1$

Differentiation

Partial Derivatives

When we studied a function of one variable $y = f(x)$, we saw that the problem of finding the tangent line led to the concept of the derivative. In a similar way, since the graph of a function of two variables $z = f(x, y)$ is a surface, we now want to find the **tangent plane** to a point on the graph of the surface. As we are about to see, this will lead us to the concept of the **partial derivative** .

Just as the equation of a line can be written as $y = mx + b$, the equation of a plane can be written as $z = ax + by + c$, where c represents here the intersection of the plane with the z axis and a, b specify how the plane is tilted with respect to the xy plane.

In the case of a line, it was possible to find the value of m once we knew two points (x_0, y_0) and (x_1, y_1) which belonged to the line: the formula was

$$m = \frac{y_1 - y_0}{x_1 - y_0} \tag{1226}$$

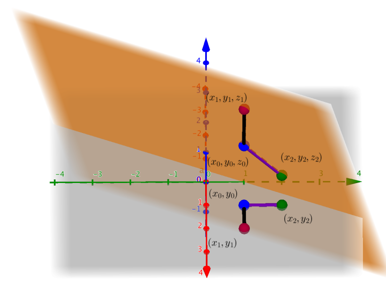


Figure 174: Equation of a plane $z = ax + by + c$

Following the analogy, it is reasonable to suspect that we can determine a and b once we know three points (x_0, y_0, z_0), (x_1, y_1, z_1) and (x_2, y_2, z_2) on the plane.

Because the equation of the plane is $z = ax + by + c$, each of the three points can be rewritten as $(x_0, y_0, ax_0 + by_0 + c)$, $(x_1, y_1, ax_1 + by_1 + c)$ and $(x_2, y_2, ax_2 + by_2 + c)$. To simplify our problem, we may suppose that

$$\text{assumption}: \begin{array}{cc} x_1 = x_0 + h & y_1 = y_0 \\ \\ x_2 = x_0 & y_2 = y_0 + h \end{array} \tag{1227}$$

where h is a positive constant: in this way, the segment connecting (x_0, y_0) and (x_1, y_1) is parallel to the x axis and the segment connecting (x_0, y_0) and (x_2, y_2) is parallel to the y axis as the image shows. In this case the three points on the plane are

$$\text{points on the plane: } \begin{array}{l} P_0 = (x_0, y_0, ax_0 + by_0 + c) \\ \\ P_1 = (x_0 + h, y_0, a(x_0 + h) + by_0 + c) \\ \\ P_2 = (x_0, y_0 + h, ax_0 + b(y_0 + h) + c) \end{array} \tag{1228}$$

From the image of the plane, we also see that the segment connecting (x_0, y_0) and (x_1, y_1) is the shadow of the segment connecting the points P_0 and P_1 on the plane. The slope of this line, which we conveniently denote as m_x , can be computed as [65]

$$m_x = \frac{a(x_0 + h) + by_0 + c - (ax_0 + by_0 + c)}{x_0 + h - x_0} = \frac{ah}{h} = a \tag{1229}$$

[65] The idea here is that since the value of the coordinate y is the same for both points the only real variable is x and z is playing the role of y in the usual formula for the slope

Similarly, the segment connecting (x_0, y_0) and (x_2, y_2) is the shadow of the segment connecting the points P_0 and P_2 on the plane. The slope of this line, which we conveniently denote as m_y , can be computed as [66]

$$m_y = \frac{ax_0 + b(y_0 + h) + c - (ax_0 + by_0 + c)}{y_0 + h - y_0} = \frac{bh}{h} = b \tag{1230}$$

[66] The idea here is that since the value of the coordinate x is the same for both points the only real variable is y and z is playing the role of y in the usual formula for the slope

Therefore, we just found the following:

Equation of a Plane: the equation of a plane can be written as

$$z = ax + by + c \tag{1231}$$

where c is the intersection of the plane with the z axis. To find a and b take three points $P_0 = (x_0, y_0, z_0)$, $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ such that:

- the segment connecting (x_0, y_0) and (x_1, y_1) is parallel to the x axis
- the segment connecting (x_0, y_0) and (x_2, y_2) is parallel to the y axis

Under these hypotheses:

- the number a is the slope of the straight segment (contained in the plane) connecting P_0 and P_1
- the number b is the slope of the straight segment (contained in the plane) connecting P_0 and P_2

We will now determine the equation of the tangent plane to a general surface $z = f(x, y)$ passing through a point (x_0, y_0, z_0) on the surface.

The equation of the tangent plane is going to be

$$z = ax + by + c \tag{1232}$$

and we need to determine a, b, c. Since the plane goes through the point (x_0, y_0, z_0), once we know a and b, the value of c can be found because thanks to equation 1232 we have that

$$\boxed{c = z_0 - ax_0 - by_0} \tag{1233}$$

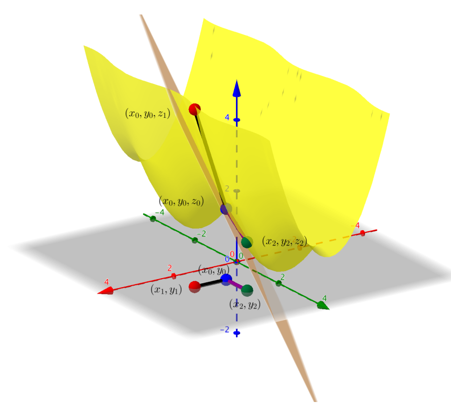


Figure 175: Finding the tangent plane

so the problem is reduced to determining a and b. We can use our previous description of a and b as slopes, however, in order to use such a description we need two additional points on the plane, which we don't have. Therefore, instead of using points on the plane we will use points on the graph of the surface and take a limit to find the values of a and b.

More precisely, choose a number h, different from zero, and consider the three points

$$\text{points on the surface:}\quad \begin{aligned} P_0 &= (x_0, y_0, f(x_0, y_0)) \\ P_1 &= (x_0 + h, y_0, f(x_0 + h, y_0)) \\ P_2 &= (x_0, y_0 + h, f(x_0, y_0 + h)) \end{aligned} \tag{1234}$$

We know that P_0 belongs to both the plane and the surface, but in principle P_1 and P_2 only belong to the surface, not the plane. However, as long as h is small enough these points should be close to the tangent plane and since a, b where the slopes m_x, m_y described above we have

$$\begin{aligned} m_x &\simeq \tfrac{f(x_0+h,y_0)-f(x_0,y_0)}{x_0+h-x_0} = \tfrac{f(x_0+h,y_0)-f(x_0,y_0)}{h} \\ m_y &\simeq \tfrac{f(x_0,y_0+h)-f(x_0,y_0)}{y_0+h-y_0} = \tfrac{f(x_0,y_0+h)-f(x_0,y_0)}{h} \end{aligned} \tag{1235}$$

and if we take the limit $h \to 0$ we should have

$$\begin{aligned} m_x &= \lim_{h\to 0} \tfrac{f(x_0+h,y_0)-f(x_0,y_0)}{h} \\ m_y &= \lim_{h\to 0} \tfrac{f(x_0,y_0+h)-f(x_0,y_0)}{h} \end{aligned} \tag{1236}$$

These formulas are strikingly similar to the definition we gave for the derivative $y'(x)$ of a function $y(x)$ and so we call these quantities the **partial derivatives of $f(x, y)$.** Therefore, we have found the following result

Partial Derivatives of $z = f(x, y)$: suppose $f(x, y)$ is a function of the two variables x and y. If (x_0, y_0) belongs to the domain of f, the partial derivative of f with respect to x at the point (x_0, y_0) is

$$f_x(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \tag{1237}$$

provided the limit exists. The partial partial derivative of f with respect to y at the point (x_0, y_0) is

$$f_y(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h} \tag{1238}$$

provided the limit exists.

It is possible to compute the partial derivative of a function using the limit definition just as we did for functions of one variable. However, in practice we won't need to do this since we can compute a partial derivative as an ordinary derivative by treating the variable we are not using for differentiating as a constant. For example, if $f(x, y) = x^2y +$

$3\sin x$ and we want to find $\frac{\partial f}{\partial x}$ then we consider y as a constant and differentiate the function as if it were a function only of x and so

$$\frac{\partial f}{\partial x} = (2x)y + 3\cos x \tag{1239}$$

Similarly, to find $\frac{\partial f}{\partial y}$ we consider x as a constant and differentiate the function as if it were a function only of y and so

$$\frac{\partial f}{\partial y} = x^2 \tag{1240}$$

Example 219. Find $f_x(1,2)$ and $f_y(1,2)$ if $f(x,y) = e^{xy}$.

The partial derivatives of f are

$$\begin{aligned} \tfrac{\partial f}{\partial x} &= ye^{xy} \\ \tfrac{\partial f}{\partial y} &= xe^{xy} \end{aligned} \tag{1241}$$

Therefore

$$\begin{aligned} f_x(1,2) &= 2e^2 \\ f_y(1,2) &= e^2 \end{aligned} \tag{1242}$$

We can also find higher derivatives of functions of two variables. In this case, there are four possible "second-derivatives": differentiating with respect to x two times, differentiating with respect to y two times, differentiate with respect to x first and then with respect to y or differentiate with respect to y first and then with respect to x. These two last cases may seem the case but in principle they could be different. Fortunately, for the functions we will be working with there is really no difference between them. As an example, if

$$f(x,y) = x\sin(x^2 + y) \tag{1243}$$

Then

$$\frac{\partial f}{\partial x} = f_x = \sin(x^2+y) + 2x^2\cos(x^2+y) \tag{1244}$$

We can now consider f_x as a new function of two variables and take its derivatives with respect to x or y. For example

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = f_{xx} = 2x\cos(x^2+y) + 4x\cos(x^2+y) - 4x^3\sin(x^2+y) \tag{1245}$$

We can now differentiate with respect to y twice

$$\begin{aligned} \tfrac{\partial f}{\partial y} &= f_y = x\cos(x^2+y) \\ \tfrac{\partial^2 f}{\partial y^2} &= f_{yy} = -x\sin(x^2+y) \end{aligned} \tag{1246}$$

Now we differentiate f with respect to x first and then with respect to y:

$$\begin{aligned} \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) &= \frac{\partial}{\partial y}\left(\sin(x^2+y)+2x^2\cos(x^2+y)\right) \\ &= \cos(x^2+y)-2x^2\sin(x^2+y \end{aligned} \tag{1247}$$

Similarly, we can differentiate with respect to y first and then with respect to x:

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) &= \frac{\partial}{\partial x}\left(x\cos(x^2+y)\right) \\ &= \cos(x^2+y)-2x^2\sin(x^2+y) \end{aligned} \tag{1248}$$

We see from this example that

$$f_{xy}=\frac{\partial^2 f}{\partial y \partial x}=\frac{\partial^2 f}{\partial x \partial y}=f_{yx} \tag{1249}$$

Again, for the functions we will be working with, it will always be the case that one can differentiate in whichever order one prefers. The mathematical way to say this is that the **mixed partial derivatives commute.**

Maxima and Minima

One of the most important applications of the derivative of a single variable function was to determine the relative maxima and minima of a function. These points had to be critical numbers, that is, the derivative evaluated at these numbers was zero or undefined. However, not every number with a vanishing (or undefined) derivative corresponded to a relative maximum or a relative minimum. That was the case with $x = 0$ for the function $f(x) = x^3$. For functions of two variables, a similar situation occurs, that is, we will see that the relative maxima and relative minima must have vanishing partial derivatives. This is because at a relative maximum and/or minimum, the graph of f has a horizontal tangent plane [67] and so it must have an equation of the form $z = c$, which implies (using formula 1231) that the partial derivatives must vanish. However, not every point to with vanishing partial derivatives corresponds to a relative maximum or relative minimum. These latter points are called **saddle points.**

For example, the origin $(0,0)$ is a saddle point for $f(x,y) = xy+1$ because we can find curves on the graph of f which approach the point $(0,0,1)$ and "see it" as a minimum (the pink curve) as well as curves on the graph of f which approach the point $(0,0,1)$ and "see it" as a maximum (the green curve).

Similarly, we will soon show that for $f(x,y) = x^2y+\frac{1}{3}y^3-x^2-y^2+2$ the point $(0,0)$ is a relative maximum, $(0,2)$ is a relative minimum and $(1,1)$, $(-1,1)$ saddle points.

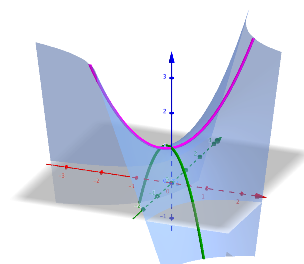


Figure 176: The origin $(0,0)$ is a saddle point for $f(x,y) = xy+1$

[67] Just as for functions of one variable it had a horizontal tangent line

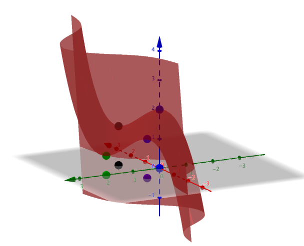


Figure 177: Critical points of $f(x,y) = x^2y+\frac{1}{3}y^3-x^2-y^2+2$

Critical points and Relative Extrema of a function $f(x, y)$:

- A **critical point** of f is a point (a, b) in the domain of f such that

$$\frac{\partial f}{\partial x}(a, b) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(a, b) = 0 \tag{1250}$$

  or at least one of the partial derivatives does not exist.

- f has a **relative maximum** at (a, b) if $f(x, y) \leq f(a, b)$ for all points (x, y) that are sufficiently close to (a, b). The number $f(a, b)$ is called a **relative maximum value**.

- f has a **relative minimum** at (a, b) with **relative minimum value** $f(a, b)$ if $f(x, y) \geq f(a, b)$ for all points (x, y) that are sufficiently close to (a, b)

- (a, b) is called a **saddle point** if it is a critical point but it is neither a relative minimum nor a relative maximum.

For functions of one variable we had two ways to classify a critical point as a relative maximum, a relative minimum or neither. The first method was to use the first derivative test: one makes a table of signs for the first derivative and studies how the sign of the first derivative changes. However, this method is harder to use directly for a function of two variables because now a point can be approached from more than two sides (its left or its right) so making a table of signs is not practical.

However, the second method, which has the second derivative test, does generalize nicely to functions of two variables (and more variables for that matter). The justification of this method is beyond the tools we possess at this point, so we will just state the theorem.

The Second Derivative Test: to classify the relative extrema of a function $f(x, y)$

- Find the critical points of $f(x, y)$ by solving the system of simultaneous equations

$$\begin{aligned} \frac{\partial f}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} &= 0 \end{aligned} \tag{1251}$$

- Define the **discriminant**

$$D(x, y) = f_{xx} f_{yy} - f_{xy}^2 \tag{1252}$$

and let (a, b) be a critical point of f.

1. If $D(a, b) > 0$ and $f_{xx}(a, b) < 0$ then $f(x, y)$ has a relative maximum at the point (a, b).
2. If $D(a, b) > 0$ and $f_{xx}(a, b) > 0$ then $f(x, y)$ has a relative minimum at the point (a, b).
3. If $D(a, b) < 0$ then (a, b) is a saddle point.
4. If $D(a, b) = 0$ the test is inconclusive.

Example 220. Classify the critical points of $f(x, y) = x^2 y + \frac{1}{3} y^3 - x^2 - y^2 + 2$ using the second derivative test.

We find the critical points of f. The partial derivatives of f are

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2xy - 2x = 2x(y-1) \\ \frac{\partial f}{\partial y} &= x^2 + y^2 - 2y = x^2 + y(y-2) \end{aligned} \tag{1253}$$

The critical points must solve the equations $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$, that is,

$$\begin{cases} 2x(y-1) = 0 \\ x^2 + y(y-2) = 0 \end{cases} \tag{1254}$$

The first equation has solution $x = 0$ or $y = 1$. If we substitute $x = 0$ into the second equation we have

$$y(y-2) = 0 \implies y = 0 \text{ or } y = 2 \tag{1255}$$

and so the two critical points corresponding to $x = 0$ are

$$(0, 0), \quad (0, 2) \tag{1256}$$

If $y = 1$ we substitute it into the second equation to obtain

$$x^2 - 1 = 0 \implies x = \pm 1 \tag{1257}$$

and so the corresponding critical points are

$$(1,1), \quad (-1,1) \tag{1258}$$

Now we proceed to classify these critical points. To compute the discriminant we compute the second order derivatives

$$\begin{aligned} f_{xx} &= \frac{\partial}{\partial x}\left(2x(y-1)\right) = 2(y-1) \\ f_{xy} &= \frac{\partial}{\partial y}\left(2x(y-1)\right) = 2x \\ f_{yy} &= \frac{\partial}{\partial y}\left(x^2+y^2-2y\right) = 2y-2 = 2(y-1) \end{aligned} \tag{1259}$$

The discriminant therefore is

$$\begin{aligned} D(x,y) &= f_{xx}f_{yy} - f_{xy}^2 \\ &= 4(y-1)^2 - 4x^2 \\ &= 4\left((y-1)^2 - x^2\right) \end{aligned} \tag{1260}$$

We evaluate the discriminant at each critical point and apply the second derivative test:

Critical Point	$f_{xx}(x,y) = 2(y-1)$	$D(x,y) = 4\left((y-1)^2 - x^2\right)$	Classification
$(0,0)$	-2	4	relative maximum
$(0,2)$	2	4	relative minimum
$(1,1)$	0	-4	saddle point
$(-1,1)$	0	-4	saddle point

and this is what we claimed before.

Example 221. Show that the surface $z = xy$ has neither a maximum nor a minimum point.

The partial derivatives are

$$\begin{aligned} \frac{\partial z}{\partial x} &= y \\ \frac{\partial z}{\partial y} &= x \end{aligned} \tag{1261}$$

and so the only critical point is the origin, that is, $(0,0)$. To show that it is a saddle point we compute $D(0,0)$. The second order partial derivatives are

$$\begin{aligned} \frac{\partial^2 z}{\partial x^2} &= 0 \\ \frac{\partial^2 z}{\partial x \partial y} &= 1 \\ \frac{\partial^2 z}{\partial y^2} &= 0 \end{aligned} \tag{1262}$$

and so

$$D(x,y) = f_{xx}f_{yy} - f_{xy}^2 = -1 \tag{1263}$$

In particular $D(0,0) = -1$ which implies by the second derivative test that $(0,0)$ is a saddle point.

Example 222. If the product of the sines of the angles of a triangle is a maximum, show that the triangle is equilateral.

Call α, β, γ the three angles of the triangle. The product of the sines of the angles is

$$\sin\alpha\sin\beta\sin\gamma \tag{1264}$$

and since these are the angles of a triangle

$$\alpha+\beta+\gamma=\pi \tag{1265}$$

which means that we can solve for one the angles in terms of the other two

$$\gamma=\pi-\alpha-\beta \tag{1266}$$

and so the function that we are trying to maximize is

$$\boxed{f(\alpha,\beta)=\sin\alpha\sin\beta\sin(\pi-\alpha-\beta)} \tag{1267}$$

Since α, β are the angles of a triangle, we can assume that $0<\alpha<\pi$ and $0<\beta<\pi$. The partial derivatives of f are

$$\begin{aligned}
\frac{\partial f}{\partial \alpha} &= \cos\alpha\sin\beta\sin(\pi-\alpha-\beta)-\sin\alpha\sin\beta\cos(\pi-\alpha-\beta)\\
&= \sin\beta\left(\cos\alpha\sin(\pi-(\alpha+\beta))-\sin\alpha\cos(\pi-(\alpha+\beta))\right)\\
\frac{\partial f}{\partial \beta} &= \sin\alpha\cos\beta\sin(\pi-\alpha-\beta)-\sin\alpha\sin\beta\cos(\pi-\alpha-\beta)\\
&= \sin\alpha\left(\cos\beta\sin(\pi-(\alpha+\beta))-\sin\beta\cos(\pi-(\alpha+\beta))\right)
\end{aligned} \tag{1268}$$

Before finding the critical points, we also compute the higher order derivatives

$$\begin{aligned}
\frac{\partial^2 f}{\partial \alpha^2} &= -2\sin\beta\left(\sin\alpha\sin(\pi-(\alpha+\beta))+\cos\alpha\cos(\pi-(\alpha+\beta))\right)\\
\frac{\partial}{\partial \alpha}\left(\frac{\partial f}{\partial \beta}\right) &= \cos\alpha\left(\cos\beta\sin(\pi-(\alpha+\beta))-\sin\beta\cos(\pi-(\alpha+\beta))\right)\\
&\quad -\sin\alpha\left(\cos\beta\cos(\pi-(\alpha+\beta))+\sin\beta\sin(\pi-(\alpha+\beta))\right)\\
\frac{\partial^2 f}{\partial \beta^2} &= -2\sin\alpha(\sin\beta\sin(\pi-(\alpha+\beta))+\cos\beta\cos(\pi-(\alpha+\beta)))
\end{aligned} \tag{1269}$$

The critical points must solve the equations $\frac{\partial f}{\partial \alpha}=\frac{\partial f}{\partial \beta}=0$, that is,

$$\begin{aligned}
&\sin\beta\left(\cos\alpha\sin(\pi-(\alpha+\beta))-\sin\alpha\cos(\pi-(\alpha+\beta))\right)=0\\
&\sin\alpha\left(\cos\beta\sin(\pi-(\alpha+\beta))-\sin\beta\cos(\pi-(\alpha+\beta))\right)=0
\end{aligned} \tag{1270}$$

Because we are assuming that $0<\alpha<\pi$ and $0<\beta<\pi$, $\sin\alpha$ and $\sin\beta$ are never 0, so we must solve the equations

$$\begin{aligned}
&\cos\alpha\sin(\pi-(\alpha+\beta))-\sin\alpha\cos(\pi-(\alpha+\beta))=0\\
&\cos\beta\sin(\pi-(\alpha+\beta))-\sin\beta\cos(\pi-(\alpha+\beta))=0
\end{aligned} \tag{1271}$$

If we use the identities $\sin(\theta_1-\theta_2) = \sin\theta_1\cos\theta_2 - \cos\theta_1\sin\theta_2$ and $\cos(\theta_1-\theta_2) = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2$ and so the equations are the same as

$$\boxed{\begin{array}{c}\cos\alpha\sin(\alpha+\beta)+\sin\alpha\cos(\alpha+\beta)=0 \quad (\bullet)\\[2ex] \cos\beta\sin(\alpha+\beta)+\sin\beta\cos(\alpha+\beta)=0 \quad (\bullet\bullet)\end{array}} \tag{1272}$$

Multiply the first equation by $\cos\beta$ and the second equation by $\cos\alpha$ to get

$$\begin{array}{c}\cos\beta\cos\alpha\sin(\alpha+\beta)+\cos\beta\sin\alpha\cos(\alpha+\beta)=0 \quad (\star)\\[2ex] \cos\alpha\cos\beta\sin(\alpha+\beta)+\cos\alpha\sin\beta\cos(\alpha+\beta)=0 \quad (\star\star)\end{array} \tag{1273}$$

If we subtract both equations, that is , $(\star)-(\star\star)$ to obtain

$$\begin{array}{rc} & \cos\beta\sin\alpha\cos(\alpha+\beta)-\cos\alpha\sin\beta\cos(\alpha+\beta)=0\\[2ex] \Longrightarrow & (\cos\beta\sin\alpha-\cos\alpha\sin\beta)\cos(\alpha+\beta)=0\\[2ex] \Longrightarrow & \sin(\alpha-\beta)\cos(\alpha+\beta)=0\end{array} \tag{1274}$$

Therefore, either $\sin(\alpha-\beta)=0$ or $\cos(\alpha+\beta)=0$. If $\cos(\alpha+\beta)=0$ then $\alpha+\beta=\frac{\pi}{2}$ and equations $(\bullet)$ and $(\bullet\bullet)$ become

$$\begin{array}{c}\cos\alpha\sin(\alpha+\beta)=0 \implies \cos\alpha=0 \implies \alpha=\frac{\pi}{2}\\[2ex] \cos\beta\sin(\alpha+\beta)=0 \implies \cos\beta=0 \implies \beta=\frac{\pi}{2}\end{array} \tag{1275}$$

Clearly each these equations can't be satisfied simultaneously so the case $\cos(\alpha+\beta)=0$ does not occur. The case $\sin(\alpha-\beta)=0$ implies that $\alpha=\beta$ and so $(\bullet)$ and $(\bullet\bullet)$ become the same equal to

$$\cos\alpha\sin(2\alpha)+\sin\alpha\cos(2\alpha)=0 \tag{1276}$$

Because $\sin(2\alpha)=2\sin\alpha\cos\alpha$ and $\cos(2\alpha)=\cos^2\alpha-\sin^2\alpha$ we have the equation

$$\begin{array}{rc} & 2\sin\alpha\cos^2\alpha+\sin\alpha\cos^2\alpha-\sin^3\alpha=0\\[2ex] \Longrightarrow & 3\cos^2\alpha-\sin^2\alpha=0\\[2ex] \Longrightarrow & 3-3\sin^2\alpha-\sin^2\alpha=0\\[2ex] \Longrightarrow & 3=4\sin^2\alpha\\[2ex] \Longrightarrow & \frac{3}{4}=\sin^2\alpha\\[2ex] \Longrightarrow & \sin\alpha=\pm\frac{\sqrt{3}}{2}\end{array} \tag{1277}$$

and since $0<\alpha<\beta$ we must have $\sin\alpha=\frac{\sqrt{3}}{2}$ which implies that $\alpha=\frac{\pi}{3}$. Therefore we found that

$$\alpha=\beta=\gamma=\frac{\pi}{3} \tag{1278}$$

and so the triangle must be equilateral. To show that it maximizes $f(\alpha,\beta)$ we compute

$$f_{xx}\left(\frac{\pi}{3},\frac{\pi}{3}\right)=-2\left(\frac{\sqrt{3}}{2}\right)\left(\left(\frac{\sqrt{3}}{2}\right)^2+\left(\frac{1}{2}\right)^2\right) \tag{1279}$$

which is negative. Similarly

$$\begin{aligned} D\left(\tfrac{\pi}{3},\tfrac{\pi}{3}\right) &= f_{xx}\left(\tfrac{\pi}{3},\tfrac{\pi}{3}\right)f_{yy}\left(\tfrac{\pi}{3},\tfrac{\pi}{3}\right)-\left(f_{xy}\left(\tfrac{\pi}{3},\tfrac{\pi}{3}\right)\right)^2 \\ &= 4\left(\tfrac{\sqrt{3}}{2}\right)^2\left(\left(\tfrac{\sqrt{3}}{2}\right)^2+\left(\tfrac{1}{2}\right)^2\right)^2-\left(-\left(\tfrac{\sqrt{3}}{2}\right)\left(\tfrac{1}{2}\right)^2-\left(\tfrac{\sqrt{3}}{2}\right)^3\right) \end{aligned} \tag{1280}$$

which is positive. Therefore, the second derivative test shows that we get a maximum, which is what we wanted to show.

Example 223. Find the dimensions of a box (top included) which contains a given volume V and uses minimum material (i.e, has minimum surface area)

Let x,y,z be the sides of the box. The volume is

$$V=xyz \tag{1281}$$

and the surface area is

$$S=2xy+2xz+2yz \tag{1282}$$

Since V is fixed we can solve for z and find

$$\boxed{z=\frac{V}{xy}} \tag{1283}$$

Substituting in the formula for S we find the function we want to minimize

$$\boxed{S(x,y)=2xy+2\frac{V}{y}+2\frac{V}{x}} \tag{1284}$$

The partial derivatives are

$$\begin{aligned} \tfrac{\partial S}{\partial x} &= 2y-2\tfrac{V}{x^2} \\ \tfrac{\partial S}{\partial y} &= 2x-2\tfrac{V}{y^2} \end{aligned} \tag{1285}$$

and the second order derivatives are

$$\begin{aligned} \tfrac{\partial^2 S}{\partial x^2} &= 4\tfrac{V}{x^3} \\ \tfrac{\partial}{\partial y}\left(\tfrac{\partial S}{\partial x}\right) &= 2 \\ \tfrac{\partial^2 S}{\partial y^2} &= 4\tfrac{V}{y^3} \end{aligned} \tag{1286}$$

The critical points must satisfy the equations

$$
\begin{aligned}
2y - 2\frac{V}{x^2} = 0 \implies y = \frac{V}{x^2} \implies V = yx^2(\bullet) \\
2x - 2\frac{V}{y^2} = 0 \implies x = \frac{V}{y^2} \implies V = xy^2(\bullet\bullet)
\end{aligned} \tag{1287}
$$

Setting $(\bullet) = (\bullet\bullet)$ we get $x = y$ and substituting in $(\bullet)$ we $V = x^3$ and so we have the box must be a cube of sides $x = \sqrt[3]{V}$. To show that it corresponds to a minimum observe that

$$S_{xx}\left(\sqrt[3]{V}, \sqrt[3]{V}\right) = 4 \tag{1288}$$

and that

$$
\begin{aligned}
D(x, y) = 16\frac{V^2}{x^3y^3} - 2 \\
\implies \quad D\left(\sqrt[3]{V}, \sqrt[3]{V}\right) = 14
\end{aligned} \tag{1289}
$$

and so by the second derivative test we obtain a relative minimum.

Another use of optimizing functions of two variables is the **least squares method.** Suppose that you want to determine experimentally the relationship between two quantities x and y, for example, you are trying to write y as a function of x. In order to determine this relationship you perform n measurements and you call the results $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $\cdots$, $P_n = (x_n, y_n)$.

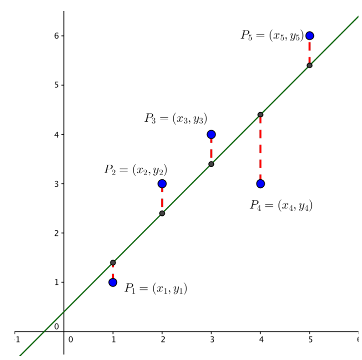


Figure 178: Least Squares Method with $n = 5$

Problem 224. When you plot the previous points they appear to lie on a line so you propose that

$$\boxed{y_{reg}(x) = mx + b} \tag{1290}$$

where m, b are two parameters that we need to determine. This line is called the least square line or the **regression line**. The least squares method consists in finding the values of m, b such that the total error

$$\boxed{E(m, b) = \sum_{i=1}^{n} \left(y_{reg}(x_i) - y_i\right)^2 = \sum_{i=1}^{n} \left(mx_i + b - y_i\right)^2} \tag{1291}$$

between the experimental value and the theoretical approximation is as small as possible. We are interested in finding the values of m and b that minimize E. Therefore these values of m and b correspond to a critical point of E and so it must solve the equations

$$\boxed{\begin{aligned} \frac{\partial E}{\partial m} = 0 \\ \frac{\partial E}{\partial b} = 0 \end{aligned}} \tag{1292}$$

For convenience, we use the following notation

$$\boxed{\begin{aligned} \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \\ \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \end{aligned}} \tag{1293}$$

To describe the critical point of E [68], the partial derivative $\frac{\partial E}{\partial m}$ equals

[68] As we are about to see, there is only one critical point of E.

$$\begin{aligned}\frac{\partial E}{\partial m} &= 2\left(\sum_{i=1}^{n} mx_i^2 + bx_i - x_iy_i\right) \\ &= 2\left(m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_iy_i\right) \\ &= 2\left(m\sum_{i=1}^{n} x_i^2 + bn\bar{x} - \sum_{i=1}^{n} x_iy_i\right)\end{aligned} \tag{1294}$$

The partial derivative $\frac{\partial E}{\partial b}$ equals

$$\begin{aligned}\frac{\partial E}{\partial b} &= 2\sum_{i=1}^{n}\left(mx_i + b - y_i\right) \\ &= 2\left(m\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} b - \sum_{i=1}^{n} y_i\right) \\ &= 2\left(mn\bar{x} + bn - n\bar{y}\right)\end{aligned} \tag{1295}$$

Therefore the equations $\frac{\partial E}{\partial m} = \frac{\partial E}{\partial b} = 0$ are equivalent to

$$\boxed{\begin{aligned} m\sum_{i=1}^{n} x_i^2 + bn\bar{x} &= \sum_{i=1}^{n} x_iy_i \\ mn\bar{x} + bn &= n\bar{y}\end{aligned}} \tag{1296}$$

These are the **normal equations** for m and b. It is possible to show that the solution of the normal equations [69] are

[69] To solve the equations one can use Cramer's rule

$$\boxed{\begin{aligned} m &= \frac{\sum_{i=1}^{n} x_iy_i - n(\bar{x})(\bar{y})}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \\ b &= \frac{\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_iy_i}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}\end{aligned}} \tag{1297}$$

It can be shows that these values of m and b correspond to a minimum of E so we found the following:

Least Squares Method: suppose we are given n data points

$$P_1 = (x_1, y_1), \quad , P_2 = (x_2, y_2), \quad , P_n = (x_n, y_n) \tag{1298}$$

Then the **least squares (regression) line** for the data is given by the linear equation

$$y = f(x) = mx + b \tag{1299}$$

where the constant m and b satisfy the **normal equations**

$$\boxed{\begin{array}{c} m\sum_{i=1}^{n} x_i^2 + bn\bar{x} = \sum_{i=1}^{n} x_i y_i \\ \\ mn\bar{x} + bn = n\bar{y} \end{array}} \tag{1300}$$

where

$$\boxed{\begin{array}{c} \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \\ \\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \end{array}} \tag{1301}$$

The solution of these equations is

$$\boxed{\begin{array}{c} m = \frac{\sum_{i=1}^{n} x_i y_i - n(\bar{x})(\bar{y})}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \\ \\ b = \frac{\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \end{array}} \tag{1302}$$

Example 225. Find the regression line for $P_1 = (1, 1)$, $P_2 = (2, 3)$, $P_3 = (3, 4)$, $P_4 = (4, 3)$, $P_5 = (5, 6)$

The values for $\bar{x}$ and $\bar{y}$ are

$$\begin{array}{c} \bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{1+2+3+4+5}{5} = 3 \\ \\ \bar{y} = \frac{\sum_{i=1}^{5} y_i}{5} = \frac{1+3+4+3+6}{5} = \frac{17}{5} \end{array} \tag{1303}$$

To compute m and b we also need

$$\begin{array}{c} \sum_{i=1}^{5} x_i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55 \\ \\ \sum_{i=1}^{5} x_i y_i = 1 \cdot 1 + 2 \cdot 3 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 6 = 61 \end{array} \tag{1304}$$

Therefore using formulas 1302 for m, b we find that

$$\begin{array}{c} m = \frac{\sum_{i=1}^{n} x_i y_i - n(\bar{x})(\bar{y})}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} = \frac{61-5(3)(\frac{17}{5})}{55-5(3)^2} = \frac{10}{10} = 1 \\ \\ b = \frac{\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} = \frac{\frac{17}{5}(55)-3(61)}{55-5(3)^2} = \frac{4}{10} = 0.4 \end{array} \tag{1305}$$

And so the equation for the regression line (which corresponds to the image shown before) is

$$y = x + 0.4 \tag{1306}$$

Integration

Double Integrals

Now we will consider the problem of defining the integral of a function of two variables $z(x, y)$. First of all, we should point out that we won't try to find an analogue for the *indefinite* integral of a function $\int y(x)dx$. A pragmatic reason [70] as to why we won't do this is that $\int y(x)dx$ was useful to compute $\int_a^b y(x)dx$, but we will have a different way to compute the (definite) double integral of a function $z(x, y)$.

[70] There are deeper theoretical reasons which are too technical to try to explain here

In analogy with $\int_a^b y(x)dx$, which could be interpreted as the area under the curve, the double integral we will define computes the "volume under the surface" and it will be denoted as

$$\int \int_R z(x, y)dA \tag{1307}$$

where R is the domain of integration and dA stands for an "infinitesimal" of area. This domain of integration R is a subset of the xy plane, and we will focus first in the case that R is a rectangle

$$R = [a, b] \times [c, d] \tag{1308}$$

To have an example in mind, suppose our function is $z(x, y) = \sqrt{2 + x}\sqrt{-y}$ and we want to find the volume between the graph and the region of the xy plane $R = [-2, 4] \times [-4, 0]$.

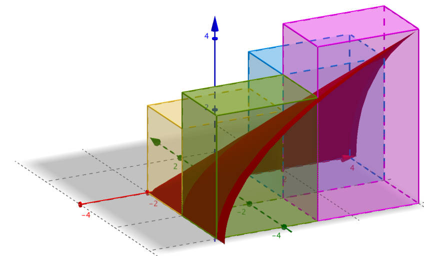


Figure 179: Approximating $\int \int_{[-2,4]\times[-4,0]} \sqrt{2+x}\sqrt{-y}dA$

When we tried to compute $\int_a^b f(x)dy$, we approximated the area by rectangles of very small width. We will use a similar strategy and approximate $\int \int_R z(x, y)dA$ by the volume of parallelepipeds with a very small base. Suppose that we divide R into a rectangular grid composed of mn rectangles, each of length $\triangle x$ and width $\triangle y$, as a result of partitioning the side of the rectangle R of length $b - a$ into m segments and the side of length $d - c$ into n segments. Therefore

$$\triangle x = \frac{b-a}{m} \qquad \triangle y = \frac{d-c}{n} \tag{1309}$$

and each rectangle has area

$$\triangle A = \triangle x \triangle y = \left(\frac{b-a}{m}\right)\left(\frac{d-c}{n}\right) \tag{1310}$$

In our example, we are taking $m = 2$ and $n = 2$ so that we have 4 rectangles with $\triangle x = \frac{4-(-2)}{2} = 3$ and $\triangle y = \frac{0-(-4)}{2} = 2$ as can be seen from the figure. The area of each rectangle is $\triangle A = 6$.

Label the mn rectangles $R_1, \cdots, R_{mn}$. If (x_i, y_i) is any point in R_i for $1 \le i \le mn$, then **the Riemann sum of** $z(x, y)$ **over the region** R is defined as

$$z(x_1, y_1)\triangle A + z(x_2, y_2)\triangle A + \cdots + f(x_{mn}, y_{mn})\triangle A \tag{1311}$$

If the limit exists as both m and n tend to infinity [71] we call this limit the **double integral of** $z(x, y)$ **over the region** R and denote it as $\int\int_R z(x, y)dA$, as we mentioned before. The only question that remains is how to compute $\int\int_R z(x, y)dA$. Fortunately, the following theorem says that the computation of $\int\int_R z(x, y)dA$ can be reduced to integrating functions of one variable and finding ordinary definite integrals.

[71] If we wanted to be more precise we would need to specify what we mean by the limit as *both* m and n tend to infinity. For example, are we taking first $m \to \infty$ and then $n \to \infty$ or vice versa or something different? We won't try to answer this question and it will be enough to have an intuitive idea of what we are doing.

Fubini's Theorem: suppose that $z(x, y)$ is continuous on the rectangle

$$R = [a, b] \times [c, d] = \{(x, y) \mid a \le x \le b, \quad c \le y \le d\} \tag{1312}$$

Then the double integral $\int\int_R z(x, y)dA$ can be computed by integrating the **iterated integrals** $\int_a^b \left(\int_c^d z(x, y)dy\right) dx$ or $\int_c^d \left(\int_a^b z(x, y)dx\right) dy$, that is

$$\int\int_R z(x, y)dA = \int_a^b \left(\int_c^d z(x, y)dy\right) dx = \int_c^d \left(\int_a^b z(x, y)dx\right) dy \tag{1313}$$

Let's explain why we mean by iterated integrals using our example with $z(x, y) = \sqrt{2 + x}\sqrt{-y}$ and $R = [-2, 4] \times [-4, 0]$. According to Fubini's Theorem

$$\begin{aligned} & \int\int_R \sqrt{2+x}\sqrt{-y}dA \\ = \quad & \int_{-2}^{4} \left(\int_{-4}^{0} \sqrt{2+x}\sqrt{-y}dy\right) dx \\ = \quad & \int_{-4}^{0} \left(\int_{-2}^{4} \sqrt{2+x}\sqrt{-y}dx\right) dy \end{aligned} \tag{1314}$$

The idea is that an iterated integral is similar to the idea of partial differentiation in that we are we integrate with respect to a variable pretending that the remaining variable is constant, just as in partial differentiation we differentiate with respect to one variable pretending that the remaining variable is constant.

Therefore, $\int_{-2}^{4} \left(\int_{-4}^{0} \sqrt{2+x}\sqrt{-y}dy\right) dx$ means the following: integrate first with respect to y pretending that x is a constant and then integrate the result of the first integration with respect to x. If we follow the instructions then we can start by saying that

$$\int_{-2}^{4} \left(\int_{-4}^{0} \sqrt{2+x}\sqrt{-y}dy\right) dx = \int_{-2}^{4} \sqrt{2+x} \left(\int_{-4}^{0} \sqrt{-y}dy\right) dx \tag{1315}$$

Because $\sqrt{2+x}$ is being treated as a constant with respect to the innermost integral and constants can be taken out of the integral. Since

$$\int_{-4}^{0} \sqrt{-y}dy = -\frac{2}{3}(-y)^{3/2}\Big|_{y=-4}^{y=0} = \frac{2}{3}(4)^{3/2} = \frac{16}{3} \tag{1316}$$

Then

$$\begin{aligned}
\int_{-2}^{4} \sqrt{2+x}\left(\int_{-4}^{0} \sqrt{-y}dy\right) dx &= \int_{-2}^{4} \sqrt{2+x}\left(\tfrac{16}{3}\right) dx \\
&= \tfrac{16}{3}\int_{-2}^{4} \sqrt{2+x}dx \\
&= \tfrac{16}{3}\left(\tfrac{2}{3}(x+2)^{3/2}\right)\Big|_{x=-2}^{x=4} \\
&= \tfrac{32}{9}(6)^{3/2}
\end{aligned} \tag{1317}$$

Therefore $\int\int_R \sqrt{2+x}\sqrt{-y}dA = \frac{32}{9}(6)^{3/2}$. Notice that in this iterated integral we integrated first with respect to y. Thanks to Fubini's theorem we are free to compute the integration in the other order, that is, integrate first with respect to x.

The way in which we find $\int_{-4}^{0}\left(\int_{-2}^{4} \sqrt{2+x}\sqrt{-y}dx\right) dy$ is entirely analogous to the previous calculation. Now we are integrating first with respect to x and so we treat y as a constant. Therefore, we can say that

$$\int_{-4}^{0}\left(\int_{-2}^{4} \sqrt{2+x}\sqrt{-y}dx\right) dy = \int_{-4}^{0} \sqrt{-y}\left(\int_{-2}^{4} \sqrt{2+x}dx\right) dy \tag{1318}$$

and now we have to find

$$\int_{-2}^{4} \sqrt{2+x}dx = \left(\frac{2}{3}(x+2)^{3/2}\right)\Big|_{x=-2}^{x=4} = \frac{2}{3}(6)^{3/2} \tag{1319}$$

and in this way

$$\begin{aligned}
\int_{-4}^{0} \sqrt{-y}\left(\int_{-2}^{4} \sqrt{2+x}dx\right) dy &= \int_{-4}^{0} \sqrt{-y}\left(\tfrac{2}{3}(6)^{3/2}\right) dy \\
&= \tfrac{2}{3}(6)^{3/2}\int_{-4}^{0} \sqrt{-y}dy \\
&= \tfrac{2}{3}(6)^{3/2}\left(\tfrac{16}{3}\right) \\
&= \tfrac{32}{9}(6)^{3/2}
\end{aligned} \tag{1320}$$

and we see that we obtain the same answer.

Example 226. Find $\int\int_R (4-x-y)dA$ where $R = [0,2]\times[0,1]$

Again we will compute both iterated integrals $\int_0^2\int_0^1 (4-x-y)dydx$ and $\int_0^1\int_0^2 (4-x-y)dxdy$ to show that both integrals agree. In practice, however, it is only necessary to compute only of the iterated integrals.

$$
\begin{aligned}
\int_0^2 \int_0^1 (4-x-y)\,dy\,dx &= \int_0^2 \left(4y - xy - \frac{y^2}{2}\right)\Big|_{y=0}^{y=1} dx \\
&= \int_0^2 \left(\frac{7}{2} - x\right) dx \\
&= \left(\frac{7}{2}x - \frac{x^2}{2}\right)\Big|_{x=0}^{x=2} \\
&= 5
\end{aligned}
\tag{1321}
$$

The second iterated integral is

$$
\begin{aligned}
\int_0^1 \int_0^2 (4-x-y)\,dx\,dy &= \int_0^1 \left(4x - \frac{x^2}{2} - yx\right)\Big|_{x=0}^{x=2} dy \\
&= \int_0^1 (6-2y)\,dy \\
&= \int_0^1 \left(6y - y^2\right)\Big|_{y=0}^{1} \\
&= 5
\end{aligned}
\tag{1322}
$$

Now we discuss the problem of computing the double integral $\int\int_R z(x,y)dA$ in the case that R is not a rectangle, but rather the region

$$R = \{(x,y) \mid g_1(x) \le y \le g_2(x), a \le x \le b\} \tag{1323}$$

between two vertical segments and two curves $g_1(x)$, $g_2(x)$ as shown in the figure. This region is sometimes called a **type I region.**

We would like to compute the integral using Fubini's Theorem. In order to use the theorem, we need the integral to take place over a rectangle. What we can do is find a rectangle $[a,b] \times [c,d]$ that contains R and consider instead

$$\int\int_{[a,b]\times[c,d]} Z(x,y)dA \tag{1324}$$

where

$$Z(x,y) = \begin{cases} z(x,y) & \text{if } (x,y) \in R \\ 0 & \text{if } (x,y) \notin R \end{cases} \tag{1325}$$

Because of Fubini's Theorem this is the same as

$$\int_a^b \int_c^d Z(x,y)dydx \tag{1326}$$

Observe that we need to compute first $\int_c^d Z(x,y)dy$. In this case x is a constant and from the figure we see that it makes sense to break the interval $[c,d]$ into the subintervals $[c, g_1(x)]$, $[g_1(x), g_2(x)]$ and

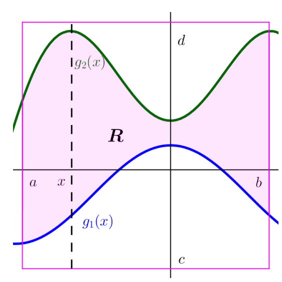


Figure 180: Integral over a type I region

$[g_2(x), d]$ so that

$$\begin{aligned}
\int_c^d Z(x,y)dy &= \int_c^{g_1(x)} Z(x,y)dy + \int_{g_1(x)}^{g_2(x)} Z(x,y)dy + \int_{g_2(x)}^{d} Z(x,y)dy \\
&= \int_c^{g_1(x)} 0dy + \int_{g_1(x)}^{g_2(x)} z(x,y)dy + \int_{g_2(x)}^{d} 0dy \\
&= \int_{g_1(x)}^{g_2(x)} z(x,y)dy
\end{aligned} \tag{1327}$$

Therefore

$$\int_a^b \int_c^d Z(x,y)dydx = \int_a^b \int_{g_1(x)}^{g_2(x)} z(x,y)dy \tag{1328}$$

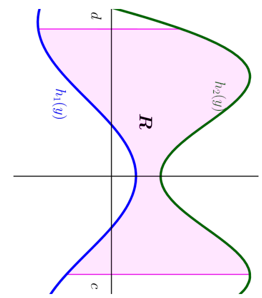


Figure 181: Integral over a type II region

but clearly the left hand side equals $\int\int_R z(x,y)dydx$ since Z was defined to be 0 outside of R so there is no contribution to the integral coming from the region outside R. Therefore, just found the following:

Integration over non-rectangular regions:

- Type I Region: suppose $g_1(x)$ and $g_2(x)$ are continuous functions on $[a,b]$ and the region R is defined by

$$R = \{(x,y) \mid g_1(x) \le y \le g_2(x), a \le x \le b\} \tag{1329}$$

  Then

$$\int\int_R f(x,y)dA = \int_a^b \left[\int_{g_1(x)}^{g_2(x)} f(x,y)dy\right] dx \tag{1330}$$

- Type II Region: suppose $h_1(y)$ and $h_2(y)$ are continuous functions on $[c,d]$ and the region R is defined by

$$R = \{(x,y) \mid h_1(y) \le x \le h_2(y); c \le y \le d\} \tag{1331}$$

  Then

$$\int\int_R f(x,y)dA = \int_c^d \left[\int_{h_1(y)}^{h_2(y)} f(x,y)dx\right] dy \tag{1332}$$

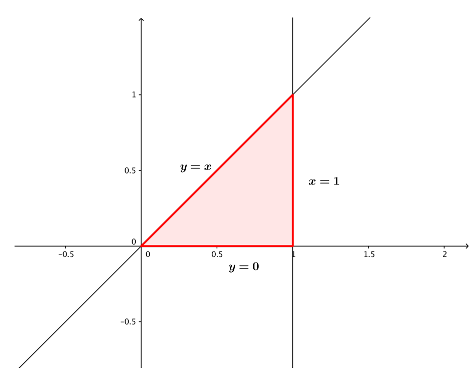


Figure 182: Triangular Region of Integration

Example 227. Find $\int\int_R \frac{\sin x}{x} dA$ were R is triangle whose sides lie down on the x axis and the lines $y = x$, $x = 1$.

Observe that this region is simultaneously type I and type II. We start by treating it as a type I region. In this case x takes its values on the interval $[0,1]$. If we fix an x inside this interval, then the values that y takes start at the horizontal line $y = 0$ and end at the straight line

$y = x$. Therefore the integral becomes

$$\begin{aligned}\int\int_R \frac{\sin x}{x} dA &= \int_0^1 \int_0^x \frac{\sin x}{x} dy dx \\ &= \int_0^1 \left(\frac{\sin x}{x}\right) \int_0^x dy dx \\ &= \int_0^1 \left(\frac{\sin x}{x}\right) x dx \\ &= \int_0^1 \sin x dx \\ &= -\cos x|_{x=0}^{x=1} \\ &= 1 - \cos 1\end{aligned} \tag{1333}$$

If we treat the region as type II then we use that y takes its values on the interval $[0,1]$. For a fixed y, the values x take start at $x = y$ and end at $x = 1$ and so the integral can be written as

$$\int\int_R \frac{\sin x}{x} dA = \int_0^1 \int_y^1 \frac{\sin x}{x} dx dy \tag{1334}$$

Notice that in this case we would need to find an antiderivative of $\frac{\sin x}{x}$ which is no easy task and so from a practical point of view this integral can't be computed. This example shows that despite the fact that the region is both of type I and type II, it may not be equally easy to find the corresponding iterated integrals.

Example 228. Find $\int\int_R \frac{x+y}{x+y+2} dA$ where R is the region shown in the following figure.

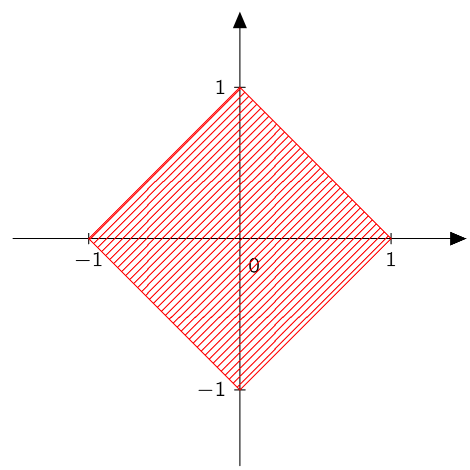


Figure 183: Rhomboidal Region

We will treat this region as a type I region. In this case x takes values on the interval $[-1,1]$. For a fixed x, y takes values between either the lower left - upper left sides of the rhombus or the lower right - upper right sides of the rhombus. This depends on whether x is positive or negative. Therefore, the first thing to do is find the equations of the four lines. These are

$$\begin{array}{cc} \text{lower left} & y = -x-1 \\ \text{upper left} & y = x+1 \\ \text{lower right} & y = x-1 \\ \text{upper right} & y = -x+1 \end{array} \tag{1335}$$

Therefore, the integral is equal to

$$\int\int_R \frac{x+y}{x+y+2} dA = \int_{-1}^0 \int_{-x-1}^{x+1} \frac{x+y}{x+y+2} dy dx + \int_0^1 \int_{x-1}^{-x+1} \frac{x+y}{x+y+2} dy dx \tag{1336}$$

To find the integral we make the substitution (remember that x is constant)

$$\begin{aligned} u &= x+y+2 \\ du &= dy \end{aligned} \tag{1337}$$

and so the integrals become

$$\begin{aligned} & \int_{-1}^{0}\int_{1}^{2x+3} \frac{u-2}{u} du dx + \int_0^1 \int_{2x+1}^{3} \frac{u-2}{u} du dx \\ = & \int_{-1}^{0}\int_{1}^{2x+3} \left(1-\frac{2}{u}\right) du dx + \int_0^1 \int_{2x+1}^{3} \left(1-\frac{2}{u}\right) du dx \\ = & \int_{-1}^{0} (2x+2-2\ln(2x+3))\, dx \\ & + \int_0^1 (2-2x-2(\ln 3 - \ln(2x+1))\, dx \end{aligned} \tag{1338}$$

Using an substitution and integration by parts we can show that

$$\begin{aligned} \int \ln(2x+3)dx &= \tfrac{1}{2}(2x+3)(-1+\ln(2x+3)) + C \\ \int \ln(2x+1)dx &= \tfrac{1}{2}(2x+1)(-1+\ln(2x+1)) + C \end{aligned} \tag{1339}$$

and so the last integrals equal

$$\begin{aligned} & \left(x^2+2x-(2x+3)(-1+\ln(2x+3))\right|_{x=-1}^{x=0} \\ & + \left(2x-x^2-2x\ln 3+(2x+1)(-1+\ln(2x+1))\right|_{x=0}^{x=1} \\ = & -3(-1+\ln 3) - [1-2-(1)(-1+\ln 1)] \\ & +1-2\ln 3+3(-1+\ln 3)-1(-1+\ln 1) \\ = & \; 3-3\ln 3 \\ & +2-2\ln 3-3+3\ln 3 \\ = & \; 2-2\ln 3 \end{aligned} \tag{1340}$$

Example 229. Write the double integral $\int_1^2 \int_0^{\ln x} (x-1)\sqrt{1+e^{2y}}\, dy dx$ in the order $dxdy$. Do not find the value.

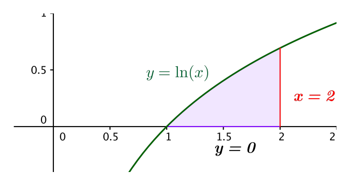


Figure 184: Changing order of integration

The way the integral is written, we need to find an antiderivative of $\sqrt{1+e^{2y}}$, which is not impossible but it is somewhat tedious. Therefore, we will invert the order of integration.

To do this notice that the largest value of y in the region of integration is $\ln 2$ and so the interval of integration for y is $[0, \ln 2]$. For a fixed value of y, x starts at the curve $y = \ln x$, which we write as $x = e^y$, and ends at the curve $x = 2$. Therefore the integral is the same as

$$\int_0^{\ln 2} \int_{e^y}^{2} (x-1)\sqrt{1+e^{2y}}\, dx dy \tag{1341}$$

We finish this section on integration stating the interpretation of $\int\int_R z(x,y)dA$ as a volume and how to compute the average value of a function.

Let R be a region in the xy plane and let f be continuous and nonnegative on R. Then the volume of the solid under a surface bounded above by $z = f(x, y)$ and below by R is given by

$$V = \int_R \int f(x, y) dA \tag{1342}$$

Example 230. Find the volume of the solid that lies under $z = x^2 + y^2$ and above the square $0 \leq x \leq 2$, $-1 \leq y \leq 1$.

The volume is

$$\begin{aligned} V &= \int_0^2 \int_{-1}^1 (x^2 + y^2) dy dx \\ &= \int_0^2 \left(x^2 y + \frac{y^3}{3}\right)\Big|_{y=-1}^{y=1} dx \\ &= \int_0^2 \left(2x^2 + \frac{2}{3}\right) dx \\ &= \left(\frac{2x^3}{3} + \frac{2}{3}x\right)\Big|_{x=0}^{x=2} \\ &= \frac{20}{3} \end{aligned} \tag{1343}$$

If f is integrable over the plane region R, then its average value over R is given by

$$\frac{\int_R \int f(x, y) dA}{\text{area of } R} = \frac{\int_R \int f(x, y) dA}{\int_R \int dA} \tag{1344}$$

Example 231. Find the average value of the function $f(x, y) = e^{-x^2}$ over the plane region R: the triangle with vertices $(0, 0)$, $(1, 0)$ and $(1, 1)$.

The area of the triangle is $\frac{1}{2}$ and so the average value is

$$\begin{aligned} & \frac{\int_0^1 \int_0^x e^{-x^2} dy dx}{\frac{1}{2}} \\ &= 2\int_0^1 x e^{-x^2} dx \\ &= \left(-e^{-x^2}\right)\Big|_{x=0}^{x=1} \\ &= (1 - e^{-1}) \end{aligned} \tag{1345}$$

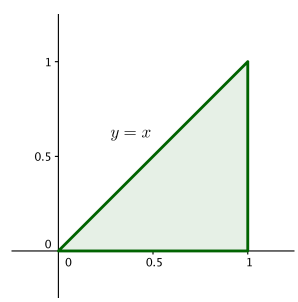


Figure 185: Average Value

Part X

Probability and Calculus

Random Variables and Probability Distributions

We will now see how calculus can be useful to Probability Theory. There are many different uses of the word "probability", but we will think of it as an assignment of numbers between 0 and 1 to the possible outcomes of an experiment, satisfying certain conditions.

The totality of all outcomes of the experiment is called the **sample space** and an element of this sample space is called an **event.** Given an event E, the **probability of the event,** called $P(E)$, is the likelihood of occurrence of this event and we take it to be as a number between 0 and 1. For example, if our experiment is tossing a coin and the outcomes are whether it lands "heads" H or "tails" T then if the coin is unbiased [72] we have that

$$P(H) = \frac{1}{2}, \quad P(T) = \frac{1}{2} \tag{1346}$$

[72] This may seem somewhat circular. After all, the only way we can tell if a coin is unbiased or not is by knowing if the probability of landing tails or heads are the same or not. Therefore, we can take this as the definition of what we mean by an "unbiased coin".

A **probability function (**or **probability distribution)** is the function which assigns to each event its probability. This probability function must satisfy certain conditions. In the case our experiment only has a finite number of outcomes (like throwing a dice or tossing a coin) the probability function is known as a **discrete probability function** and the criteria it must satisfy are the following:

A **Discrete Probability Function** P with domain $\{x_1, \cdots, x_n\}$ (these are the outcomes of an experiment) is a function satisfying the conditions:

- $0 \leq P(x_i) \leq 1$ for $1 \leq i \leq n$, that is, the probability of each event is between 0 and 1.
- $P(x_1) + \cdots + P(x_n) = 1$, that is, something is always measured.

There are different ways to organize the information given by a discrete probability function. One way is to organize it graphically by means of a **histogram.** To write a histogram, think of the numbers $\{x_1, \cdots, x_n\}$ as the values that a variable can take. Such a variable is called a **random variable**. For example, suppose that we are measuring the different ages of a population and you determine (measure) that they are

Age	15	20	21	27	30	32	35	37

These eight numbers are the outcomes of our experiment and so we can consider the collection $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ which represent the ages of the population where $x_1 = 15$, $x_2 = 20$, $x_3 = 21$, $x_4 = 27$, $x_5 = 30$, $x_6 = 32$, $x_7 = 35$, $x_8 = 37$. Therefore our random variable X is a variable whose values are the different ages we measured in our experiment [73], that is, a variable whose values lie in the set $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$. In order to have complete information about our experiment we also need to know the number of times we observed each age, that is, the number of people with a given age. Therefore the table we will use to make the histogram is

Age	15	20	21	27	30	32	35	37
Frequency	2	1	4	1	5	1	2	4

[73] Strictly speaking a random variable is a function so it has a domain and a range. However, in practice we tend to ignore the domain of X and focus only on the values it takes.

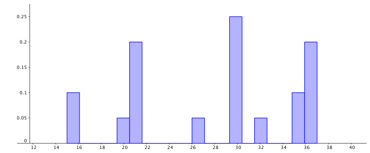


Figure 186: Histogram

Observe from the table that 20 people had their ages measured. Therefore, we can create a new row in which we write the relative frequency of each age measured, that is, the frequency of each age divided by the total number of people. We will interpret this relative frequency as the probability of each outcome.

Age x_i	15	20	21	27	30	32	35	37
Frequency	2	1	4	1	5	1	2	4
Probability $P(X = x_i)$	0.1	0.05	0.2	0.05	0.25	0.05	0.1	0.2

The histogram of the random variable X is constructed by locating the values of the random variable on a horizontal line. Then above each number, we draw a rectangle whose height is equal to the probability associated with that value of the random variable.

In order to use the techniques of Calculus we will focus in the case that our random variable X assumes any value in an interval. In such a case we call our random variable a **continuous random variable** . Some continuous random variable include the temperature inside a room, the life span of a light bulb and the velocity of a particle of air inside a room.

We interpret $P(a \leq X \leq b)$ as the probability that the random variable X assumes a value in the interval $a \leq X \leq b$. In the case of a discrete random variable, we required that the sum of the probabilities of all the events to add up to one. The continuous version of a sum is an integral so we impose the following requirement on our probability distribution function, which we call the **probability density function.**

A **probability density function** of a random variable X in an interval I, where I may be bounded or unbounded, is a nonnegative function f having the property that

$$\int_I f(x)dx = 1 \tag{1347}$$

The probability that an observed value of the random variable X lies in the interval $[a, b]$ is given by

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{1348}$$

Example 232. Suppose X is a random variable on $[0,3]$ with probability density function $f(x) = \frac{x}{\sqrt{x^2+16}}$.

a) Show that f is indeed a probability density function.

We need to verify that f is positive and that it has integral 1. Because the domain of f is $[0,3]$, it is clear that f is positive. To show that $\int_0^3 \frac{x}{\sqrt{x^2+16}} dx$ we can use the substitution $u = x^2+16$, $du = 2xdx$ to obtain

$$\begin{aligned}
\int_0^3 \frac{x}{\sqrt{x^2+16}} dx &= \int_{16}^{25} \frac{du}{2\sqrt{u}} \\
&= (\sqrt{u})\Big|_{u=16}^{u=25} \\
&= (\sqrt{25} - \sqrt{16}) \\
&= 5-4 \\
&= 1
\end{aligned} \tag{1349}$$

b) What is $P(X \geq 1)$?

We integrate the density function from 1 to 3:

$$\begin{aligned}
P(X \geq 1) &= \int_1^3 \frac{x}{\sqrt{x^2+16}} dx \\
&= \int_{17}^{25} \frac{du}{2\sqrt{u}} \\
&= (\sqrt{u})\Big|_{u=17}^{u=25} \\
&= 5 - \sqrt{17}
\end{aligned} \tag{1350}$$

c) What is $P(X < 1)$?

We can either compute $\int_0^1 \frac{x}{\sqrt{x^2+16}} dx$ or use the fact that

$$P(X < 1) = 1 - P(X \geq 1) = 1 - \left(5 - \sqrt{17}\right) = \sqrt{17} - 4 \tag{1351}$$

Sometimes we may be interested in the outcomes of an experiment that depend on more that one random variable. For example, we might be interested in studying if there is any relationship between the weight X and the height Y of an individual. In such a case our measurements would consist in measuring those variables in a population to determine if the variables are correlated or not. From a statistical point of view, the information one can acquire from X and Y is contained in the **joint probability density function**.

A **joint probability density function** of the random variables X and Y on a region D is a nonnegative function $f(x,y)$ having the property

$$\int\int_D f(x,y)dA = 1 \tag{1352}$$

The probability that the observed values of the random variables X and Y lie in a region $R \subset D$ is given by

$$P[(X,Y) \text{ in } R] = \int\int_R f(x,y)dA \tag{1353}$$

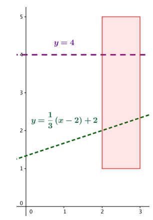


Figure 187: Joint probability function

Example 233. Consider the function of two variables $f(x,y) = \frac{4c(x+y)}{105}$ defined on the domain $D = \{(x,y) \mid 2 \le x \le 3, 2 \le y \le 5\}$ where c is a constant.

a) Find the value of c so that $f(x,y)$ is a joint probability density function of the random variables X and Y that take values on the domain D.

We need to find c so that $\int\int_D f(x,y) = 1$, that is, we need

$$\begin{aligned} 1 &= \int_2^3\int_2^5 \tfrac{4c(x+y)}{105}dydx \\ &= \tfrac{4c}{105}\int_2^3 \left(xy+\tfrac{y^2}{2}\right)\Big|_{y=2}^{y=5} dx \\ &= \tfrac{4c}{105}\int_2^3 \left(3x+\tfrac{21}{2}\right) dx \\ &= \tfrac{4c}{105}\left(\tfrac{3x^2}{2}+\tfrac{21x}{2}\right)\Big|_{x=2}^{x=3} \\ &= \tfrac{4c}{105}(18) \end{aligned} \tag{1354}$$

Therefore the value for c is

$$c = \frac{105}{72} \tag{1355}$$

and so the density function is

$$f(x,y) = \frac{1}{18}(x+y) \tag{1356}$$

b) Find $P(Y \ge 4)$.

We need to find the double integral

$$\begin{aligned} P(Y \geq 4) &= \int_2^3 \int_4^5 \tfrac{1}{18}(x+y) \\ &= \tfrac{1}{18}\int_2^3 (xy + \tfrac{y^2}{2})\Big|_{y=4}^{y=5} dx \\ &= \tfrac{1}{18}\int_2^3 (x + \tfrac{9}{2})\, dx \\ &= \tfrac{1}{18}\left(\tfrac{x^2}{2} + \tfrac{9}{2}x\right)\Big|_{x=2}^{x=3} \\ &= \tfrac{1}{18}(7) \end{aligned} \tag{1357}$$

c) find $P(Y \leq \frac{1}{3}(X-2)+2)$

First we plot the line $y = \frac{1}{3}(x-2)+2$ on the xy plane. We must integrate the region below this line which is inside the rectangle D and so

$$\begin{aligned} P\left(Y \leq \tfrac{1}{3}(X-2)+2\right) &= \int_2^3 \int_2^{\frac{1}{3}(x-2)+2} \tfrac{1}{18}(x+y)dydx \\ &= \tfrac{1}{18}\int_2^3 \left(xy + \tfrac{y^2}{2}\right)\Big|_{y=2}^{y=\frac{x+4}{3}} dx \\ &= \tfrac{1}{18}\int_2^3 \left(x\left(\tfrac{x+4}{3}\right) + \tfrac{1}{2}\left(\tfrac{x^2+8x+16}{9}\right) - 2x - 2\right) dx \\ &= \tfrac{1}{18}\int_2^3 \tfrac{7x^2-4x-20}{18} dx \\ &= \tfrac{1}{324}\left(7\tfrac{x^3}{3} - 2x^2 - 20x\right)\Big|_{x=2}^{x=3} \\ &= \tfrac{1}{324}\left(\tfrac{43}{3}\right) \end{aligned} \tag{1358}$$

Example 234. X and Y are random variables that take values in the region $R = \{0 \leq x \leq 3, 0 \leq y \leq 1\}$ and have a joint probability density function $f(x,y) = cye^x$, where c is a positive constant. Find the value of c that makes c a probability density function and set up an iterated integral that gives the probability that $X \geq 4Y$. Do not evaluate.

We need

$$\begin{aligned} 1 &= \int_0^3 \int_0^1 cye^x dydx \\ &= c\int_0^3 e^x \left(\tfrac{y^2}{2}\right)\Big|_{y=0}^{1} dx \\ &= \tfrac{c}{2}\int_0^3 e^x dx \\ &= \tfrac{c}{2}(e^3-1) \end{aligned} \tag{1359}$$

Therefore the value of c is

$$c = \frac{2}{e^3 - 1} \tag{1360}$$

The probability that $X \geq 4Y$ can be found by plotting the line $x = 4y$ in the xy plane and so

$$P(X \geq 4Y) = \int_0^3 \int_0^{\frac{x}{4}} f(x,y)dydx = \frac{2}{e^3 - 1}\int_0^3 \int_0^{\frac{x}{4}} ye^x dydx \tag{1361}$$

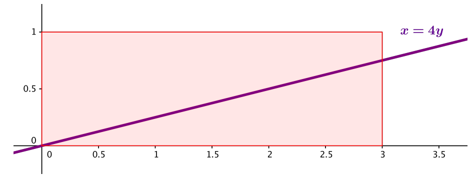


Figure 188: Joint Probability Distribution 2

Expected Value and Standard Deviation

Once we know the probability distribution of a random variable, we can use it to extract statistical information about it. The simplest (and perhaps most important) information we can compute with it is the **expected value** of the random variable. We define it separately for discrete and continuous random variables:

Expected Value of a Discrete Random Variable: Let X be a random discrete variable that assumes the values $x_1, \cdots, x_n$ with probabilities $p_1, \cdots, p_n$. The **expected value** of X is denoted by $E(X)$ and we define it as

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \tag{1362}$$

In particular, when each outcome is equiprobable, that is, it has the same probability of occurring, then the expected value is the **average** (or **mean**) of the n numbers $x_1, \cdots, x_n$, that is

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n} \tag{1363}$$

Expected Value of a Continuous Random Variable: suppose that f is the probability density function associated with a continuous random variable X. If f is defined on $[a, b]$, then the expected value of X is

$$E(X) = \int_a^b x f(x) dx \tag{1364}$$

We can think of the expected value of a random variable X as a measure of the central tendency of the probability distribution associated with X. In repeated trials of an experiment with a random variable X, the average of the observed values of X gets closer and closer to the expected value of X as the number of trials gets larger and larger.

For discrete random variables the interpretation is more complicated. For example, suppose that X represents the electric charge of individual particles inside a gas. If it were somehow possible to measure the individual charges of the electrons (which have charge $-e$) and protons (which have charge e), then the expected charge of a particle inside the gas would be

$$E(X) = -ep_1 + ep_2 \tag{1365}$$

where p_1 is the probability of measuring the charge of an electron and p_2

the probability of measuring the charge of a proton. If we assume that $p_1 = p_2 = \frac{1}{2}$ then the expected value is zero, although any particular experiment we will make is never 0.

Example 235. If the probability density function of a random variable X has the form

$$f(x) = ke^{-kx} \tag{1366}$$

where $x \geq 0$ and k is a positive constant, we say that X has an exponential density function and that X is an exponential random variable. Exponential random variables are used to represent the life span of an electronic component, the duration of telephone call, the waiting time in a doctor's office, and the time between successive flight arrivals and departures in an airport, to mention but a few applications.

a) Show that f is a probability distribution.

Since k is positive and the exponential is never negative, f is a positive function. Now we need to show that the integral of f is 1. Since the only condition on x is that it is not negative, we integrate over the interval $[0, \infty)$. In this way

$$\begin{aligned} \int_0^\infty ke^{-kx}dx &= \lim_{b\to\infty} -e^{-kx}\Big|_{x=0}^{x=b} \\ &= \lim_{b\to\infty}(-e^{-kb}+1) \\ &= 1 \end{aligned} \tag{1367}$$

b) Find the expected value of X.

To find the expected value we use integration by parts

$$\begin{aligned} E(X) &= \int_0^\infty xke^{-kx}dx \\ &= \lim_{b\to\infty}\int_0^b xke^{-kx}dx \\ &= \lim_{b\to\infty}\frac{1}{k}\left(-kxe^{-kx}-e^{-kx}\right)\Big|_{x=0}^{x=b} \\ &= \lim_{b\to\infty}\frac{1}{k}(-kbe^{-kb}-e^{-kb}+1) \\ &= \frac{1}{k} \end{aligned} \tag{1368}$$

No matter how useful, the expected value does not provide complete information about our random variable. In fact, very different random variables can have the same expected value. For example, if X is the random variable with density function $f_X(x) = 1$ on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$ then its expected value is

$$E(X) = \int_{-1/2}^{1/2} xf_X(x)dx = \int_{-1/2}^{1/2} xdx = 0 \tag{1369}$$

Similarly, if Y is the random variable with density function $f_Y(y) = 12y^2$ then

$$E(Y) = \int_{-1/2}^{1/2} yf_Y(y)dy = \int_{-1/2}^{1/2} 12y^3dy = 0 \tag{1370}$$

A way to distinguish these two random variables is by computing their **variance**. The variance is a measure of the degree of dispersion, or spread, of a probability distribution about its mean. The idea is that a probability distribution with a small spread about its mean will have a small variance, whereas one with a larger spread will have a larger variance.

Variance of a Discrete Random Variable: suppose that a random variable has probability distribution

x	x_1	x_2	$\cdots$	x_n
$P(X=x)$	p_1	p_2	$\cdots$	p_n

and expected value

$$\mu = E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \tag{1371}$$

Then the **variance** of X is

$$\mathrm{var}(X) = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \cdots + p_n(x_n-\mu)^2 \tag{1372}$$

Variance of a Continuous Random Variable: let X be a continuous random variable with probability density function $f(x)$ on $[a,b]$. If X has expected value

$$\mu = E(X) = \int_a^b x f(x) dx \tag{1373}$$

then its variance is

$$\mathrm{Var}(X) = \int_a^b (x-\mu)^2 f(x) dx \tag{1374}$$

For example, the variance of X is

$$\begin{aligned} \mathrm{Var}(X) &= \textstyle\int_{-1/2}^{1/2}(x-0)^2 f_X(x) dx \\ &= \textstyle\int_{-1/2}^{1/2} x^2 dx \\ &= \tfrac{1}{12} \end{aligned} \tag{1375}$$

while the variance of Y is

$$\begin{aligned} \mathrm{Var}(Y) &= \textstyle\int_{-1/2}^{1/2}(y-0)^2 f_Y(y) dy \\ &= \textstyle\int_{-1/2}^{1/2} y^2(12y^2) dy \\ &= \tfrac{12}{80} \end{aligned} \tag{1376}$$

Therefore, the variances are different so it distinguishes between X and Y.

Notice that if X is a random variable with physical units, for example, if X represents the temperature of a room then X is measured in

Fahrenheit. The expected value $E(X)$ is measured has the same units as X so it will also have units of Fahrenheit. On the other hand, the variance $\text{Var}(X)$ has units of temperature squared, so it would be useful to define a quantity that measures the spread of the distribution, but with the same units as the random variable. This quantity is known as the **standard deviation**, and we define it as follows:

Standard Deviation of a Random Variable: if X is a random variable with variance $\text{Var}(X)$, its standard deviation σ is the square root of the variance, that is,

$$\sigma = \sqrt{\text{Var}(X)} \tag{1377}$$

Example 236. Let X be a continuous random variable taking values on $0 \le x \le 2$ with associated density function $f(x) = \frac{1}{6}(x+2)$.

a) Show that $f(x)$ is in fact a probability density function for X

Observe that f is positive on the interval $[0,2]$. We just need to show that it has integral 1

$$\begin{aligned} \int_0^2 f(x)dx &= \int_0^2 \tfrac{1}{6}(x+2)dx \\ &= \tfrac{1}{6}\left(\tfrac{x^2}{2}+2x\right)\Big|_{x=0}^{x=2} \\ &= \tfrac{1}{6}(6) \\ &= 1 \end{aligned} \tag{1378}$$

b) Find $P(X \ge 3/2)$

$$\begin{aligned} P(X \ge 3/2) &= \int_{3/2}^2 f(x)dx \\ &= \tfrac{1}{6}\int_{3/2}^2 (x+2)dx \\ &= \left(\tfrac{1}{6}\right)\left(\tfrac{15}{8}\right) \\ &= \tfrac{16}{48} \end{aligned} \tag{1379}$$

c) Find $E(X)$

$$\begin{aligned} E(X) &= \int_0^2 \tfrac{x}{6}(x+2)dx \\ &= \tfrac{1}{6}\left(\tfrac{x^3}{3}+x^2\right)\Big|_{x=0}^{x=2} \\ &= \tfrac{1}{6}\left(\tfrac{8}{3}+4\right) \\ &= \tfrac{10}{9} \end{aligned} \tag{1380}$$

d) Find the variance $\text{Var}(X)$

Since the expected value is $\mu = E(X) = \frac{10}{9}$ the formula for the variance becomes

$$\begin{aligned}
\text{Var}(X) &= \int_0^2 (x-\mu)^2 f(x)dx \\
&= \int_0^2 (x-\tfrac{10}{9})^2 \tfrac{1}{6}(x+2)dx \\
&= \tfrac{1}{6}\int_0^2 \left(x^2 - \tfrac{20}{9}x + \tfrac{100}{81}\right)(x+2)dx \\
&= \tfrac{1}{6}\int_0^2 \left(x^3 - \tfrac{2x^2}{9} - \tfrac{260x}{81} + \tfrac{200}{81}\right) dx \\
&= \tfrac{1}{6}\left(\tfrac{52}{27}\right) \\
&= \tfrac{26}{81}
\end{aligned} \tag{1381}$$

e) Find the standard deviation σ of X

The standard deviation is the square root of the variance, that is

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\frac{26}{81}} = \frac{\sqrt{26}}{9} \tag{1382}$$

We finish this section deriving an alternative formula for the standard deviation. Using the definition we have that

$$\begin{aligned}
\sigma^2 &= \text{Var}(X) \\
&= \int_a^b (x-E(X))^2 f(x)dx \\
&= \int_a^b (x^2 - 2xE(X) + (E(X))^2) f(x)dx \\
&= \int_a^b x^2 f(x) - 2E(X)\int_a^b xf(x) + (E(X))^2 \int_a^b f(x)dx \\
&= E(X^2) - 2(E(X))^2 + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned} \tag{1383}$$

Therefore we have found that

$$\boxed{\sigma^2 = E(X^2) - (E(X))^2} \tag{1384}$$

Normal Distribution

We finish this review of probability with the most important probability density function of all, this is the so called **normal distribution.** The normal probability density function [74] with mean μ and standard deviation σ is defined to be

[74] It is clear that f is a positive function so the only thing that has to be verified is that $\int_{-\infty}^{\infty} f(x)dx = 1$ but we won't do this here.

$$f(x) = \frac{e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}}{\sigma\sqrt{2\pi}} \tag{1385}$$

for $-\infty < x < \infty$.

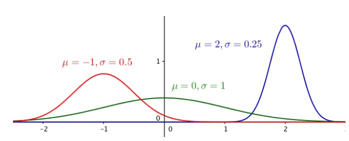


Figure 189: Example Normal Distributions

Many phenomena, including the heights of people in a given population, the weights of newborn infants and the IQs of college students have probability distributions that are approximately normal. The reason why the normal distribution appears so frequently is because of the central limit theorem, which lies at the foundations of probability and statistics. Sometimes the normal distribution is known as the Bell curve. In the examples shown in the image, the normal curve with mean $\mu = 0$ and standard deviation $\sigma = 1$ is called the **standard normal curve.** The corresponding distribution is called the **standard normal distribution**, and it is commonly denoted by Z.

Here are some properties of the normal distribution:

Normal Distribution: the normal probability density function with mean μ and standard deviation σ is defined to be

$$f(x) = \frac{e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}}{\sigma\sqrt{2\pi}} \tag{1386}$$

- The curve has a peak at $x = \mu$, that is, the maximum occurs at $x = \mu$ /
- The curve is symmetric with respect to the vertical line $x = \mu$
- About 68.27% of the area under the curve lies within 1 standard deviation of the mean (that is, between $\mu - \sigma$ and $\mu + \sigma$), about 95.45% of the area lies within 2 standard deviations of the mean , and about 99.73% of the area lies within 3 standard deviations of the mean
- If X is a normal random variable with mean μ and standard deviation σ, then it can be transformed into the standard normal random variable Z by means of the substitution $Z = \frac{X-\mu}{\sigma}$. In this case

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \tag{1387}$$

and

$$\begin{aligned} P(X < b) &= P\left(Z < \tfrac{b-\mu}{\sigma}\right) \\ P(X > a) &= P\left(Z > \tfrac{a-\mu}{\sigma}\right) \end{aligned} \tag{1388}$$

The idea of relating the values of a generic normal distribution to the standard normal distribution is that we can use a table that computes different values of $P(Z \leq z)$ and in that way we avoid finding any integrals, since these can only be found approximately most of the time.

Example 237. Let X be a normal variable with mean 10 and variance 4 , and let Z be the standard normal variable.

a) Find $P(X < 10)$.

Since the variance is 4, the standard deviation σ is $\sigma = 2$ so

$$\begin{aligned} P(X < 10) &= P\left(Z < \tfrac{10-10}{2}\right) \\ &= P(Z < 0) \\ &= \tfrac{1}{2} \end{aligned} \tag{1389}$$

b) Find the number a such that $P(X < 22) = P(Z < a)$.

Again we compare it to the standard normal distribution Z:

$$\begin{aligned} P(X < 22) &= P\left(Z < \tfrac{22-10}{2}\right) \\ &= P(Z < 6) \end{aligned} \tag{1390}$$

Therefore we must take $a = 6$.

c) Which is larger, $P(X > 13)$ or $P(Z < -1)$?

From the table for Z we have

$$P(Z < -1) = 0.1587 \qquad (1391)$$

and by the rescaling properties then

$$\begin{aligned} P(X > 13) &= P\left(Z > \tfrac{13-10}{2}\right) \\ &= P\left(Z > 1.5\right) \\ &= 1 - P(Z \leq 1.5) \qquad (1392) \\ &= 1 - 0.9332 \\ &= 0.0668 \end{aligned}$$

Therefore $P(Z < -1)$ is bigger.

List of Figures

Applications Integral

Index