ZIAUDDIN UNIVERSITY
EXAMINATION BOARD

Mathematics
XII
Teacher Resource

Contents

UNIT 1 MAPLE

1.1 Introduction

- Describe basic MAPLE commands

1.2 Polynomials

- Apply the MAPLE commands
- To factoring a polynomial.
- Expanding an expression.
- Simplifying an expression.
- Simplifying a rational expression.
- Substituting in to an expression.

A polynomial looks like this:

example of a polynomial
this one has 3 terms

Polynomial comes from *poly-* (meaning "many") and *-nomial* (in this case meaning "term") ... so it says "many terms"

A polynomial can have:

constants (like **3**, **−20**, or ½)

variables (like ***x*** and ***y***)

exponents (like the 2 in y^2), but only **0, 1, 2, 3, ...** etc are allowed

that can be combined using **addition, subtraction, multiplication and division** ...

... except ...

... **not** division by a variable (so something like **2/x** is right out)

So:

A polynomial can have constants, variables and exponents,
but never division by a variable.

Also they can have one or more terms, but not an infinite number of terms.

Polynomial or Not?

These **are** polynomials:

x

− 2

·$6y^2 - (79)x$

$xyz + 3xy^2z - 0.1xz - 200y + 0.5$

$12v^5 + 99w^5$

(Yes, "5" is a polynomial, **one term is allowed**, and it can be just a constant!)

These are **not** polynomials

xy^{-2} is not, because the exponent is "-2" (exponents can only be 0,1,2,...)

/(x+2) is not, because dividing by a variable is not allowed

/x is not either

'**x** is not, because the exponent is "½" (see fractional exponents)

But these **are** allowed:

/2 is allowed, because you can divide by a constant

lso **3x/8** for the same reason

'**2** is allowed, because it is a constant (= 1.4142...etc)

Monomial, Binomial, Trinomial

There are special names for polynomials with 1, 2 or 3 terms:

How do you remember the names? Think cycles!

There is also quadrinomial (4 terms) and quintinomial (5 terms),
but those names are not often used.

Variables

Polynomials can have no variable at all

Example: 21 is a polynomial. It has just one term, which is a constant.

Or one variable

Example: $x^4 - 2x^2 + x$ has three terms, but only one variable (x)

Or two or more variables

Example: $xy^4 - 5x^2z$ has two terms, and three variables (x, y and z)

What is Special About Polynomials?

Because of the strict definition, polynomials are **easy to work with**.

For example we know that:

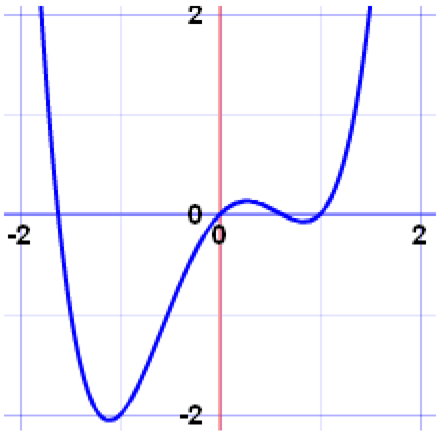
f you add polynomials you get a polynomial

f you multiply polynomials you get a polynomial

So you can do lots of additions and multiplications, and still have a polynomial as the result.

Also, polynomials of one variable are easy to graph, as they have smooth and continuous lines.

ample: x^4-2x^2+x



See how nice and smooth the curve is?

You can also divide polynomials (but the result may not be a polynomial).

Degree

The **degree** of a polynomial with only one variable is the **largest exponent** of that variable.

ample:

$4x^3 - x + 3$ The Degree is **3** (the largest exponent of ***x***)

For more complicated cases, read Degree (of an Expression).

Standard Form

The Standard Form for writing a polynomial is to put the terms with the highest degree first.

ample: Put this in Standard Form: $3x^2 - 7 + 4x^3 + x^6$

The highest degree is 6, so that goes first, then 3, 2 and then the constant last:

$x^6 + 4x^3 + 3x^2 - 7$

You **don't have to** use Standard Form, but it helps.

1.3 Graphics

- Plot a two – dimensional graph using MAPLE commands.
- Demonstrate domain and range of a plot.
- Sketch parametric equations.
- Describe plotting options.

1.4 Matrices

- Recognize matrix and vector entry arrangement through maple.
- Calculate matrix operations.
- Compute inverse and transpose of a matrix.

UNIT 2 FUNCTIONS AND LIMITS

2.1 Functions

- Identify through graph the domain and range of a function.
- Draw the graph of modulus function(i.e. y =$|x|$)

Learning Objectives

- 1.1.1. Use functional notation to evaluate a function.
- 1.1.2. Determine the domain and range of a function.
- 1.1.3. Draw the graph of a function.
- 1.1.4. Find the zeros of a function.
- 1.1.5. Recognize a function from a table of values.
- 1.1.6. Make new functions from two or more given functions.
- 1.1.7. Describe the symmetry properties of a function.

In this section, we provide a formal definition of a function and examine several ways in which functions are represented—namely, through tables, formulas, and graphs. We study formal notation and terms related to functions. We also define composition of functions and symmetry properties. Most of this material will be a review for you, but it serves as a handy reference to remind you of some of the algebraic techniques useful for working with functions.

Functions

Given two sets AA and B,B, a set with elements that are ordered pairs (x,y),(x,y), where xx is an element of AA and yy is an element of B,B, is a relation from AA to B.B. A relation from AA to BB defines a relationship between those two sets. A function is a special type of relation in which each element of the first set is related to exactly one element of the second set. The element of the first set is called the *input*; the element of the second set is called the *output*. Functions are used all the time in mathematics to describe relationships between two sets. For any function, when we know the input, the output is determined, so we say that the output is a function of the input. For example, the area of a square is determined by its side length, so we say that the area (the output) is a function of its side length (the input). The velocity of a ball thrown in the air can be described as a function of the amount of time the ball is in the air. The cost of mailing a package is a function of the weight of the package. Since functions have so many uses, it is important to have precise definitions and terminology to study them.

DEFINITION

A **function** f consists of a set of inputs, a set of outputs, and a rule for assigning each input to exactly one output. The set of inputs is called the **domain** of the function. The set of outputs is called the **range** of the function.

For example, consider the function $f,$ where the domain is the set of all real numbers and the rule is to square the input. Then, the input $x=3$ is assigned to the output $3^2=9.$ Since every nonnegative real number has a real-value square root, every nonnegative number is an element of the range of this function. Since there is no real number with a square that is negative, the negative real numbers are not elements of the range. We conclude that the range is the set of nonnegative real numbers.

For a general function f with domain $D,$ we often use x to denote the input and y to denote the output associated with $x.$ When doing so, we refer to x as the **independent variable** and y as the **dependent variable**, because it depends on $x.$ Using function notation, we write $y=f(x),$ and we read this equation as "y equals f of $x.$" For the squaring function described earlier, we write $f(x)=x^2.$

The concept of a function can be visualized using Figure 1.2, Figure 1.3, and Figure 1.4.

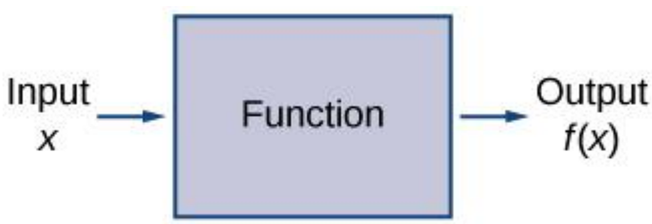


Figure 1.2 A function can be visualized as an input/output device.

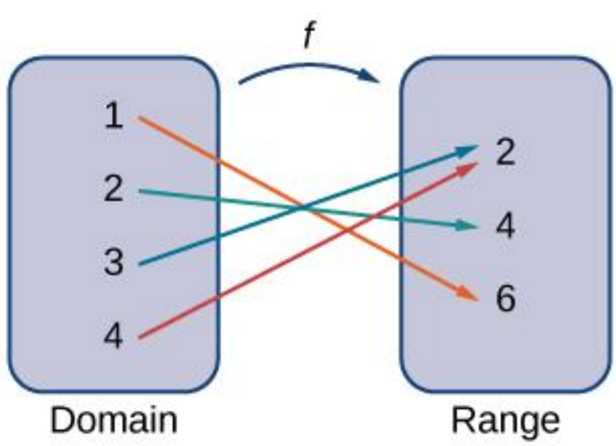


Figure 1.3 A function maps every element in the domain to exactly one element in the range. Although each input can be sent to only one output, two different inputs can be sent to the same output.

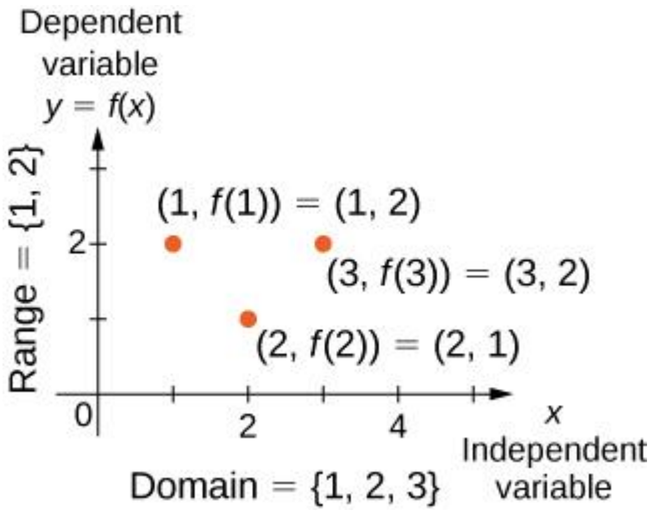


Figure 1.4 In this case, a graph of a function ff has a domain of {1,2,3}{1,2,3} and a range of {1,2}.{1,2}. The independent variable is xx and the dependent variable is y.y.

MEDIA

Visit this applet link to see more about graphs of functions.

We can also visualize a function by plotting points (x,y)(x,y) in the coordinate plane where y=f(x).y=f(x). The **graph of a function** is the set of all these points. For example, consider the function f,f, where the domain is the set D={1,2,3}D={1,2,3} and the rule is f(x)=3−x.f(x)=3−x. In Figure 1.5, we plot a graph of this function.

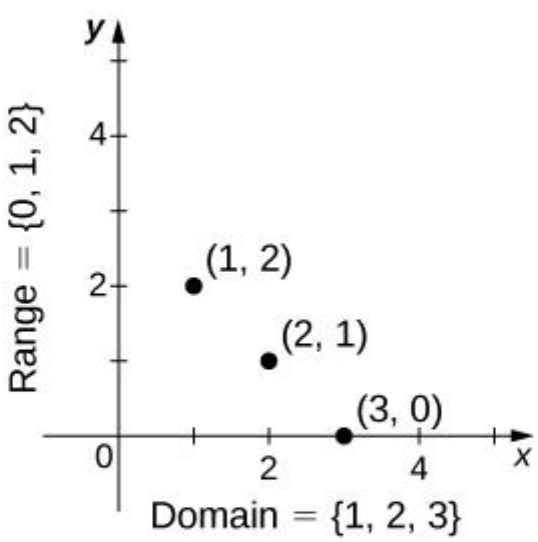


Figure 1.5 Here we see a graph of the function ff with domain {1,2,3}{1,2,3} and rule f(x)=3−x.f(x)=3−x. The graph consists of the points (x,f(x))(x,f(x)) for all xx in the domain.

Every function has a domain. However, sometimes a function is described by an equation, as in $f(x)=x^2,$ with no specific domain given. In this case, the domain is taken to be the set of all real numbers x for which $f(x)$ is a real number. For example, since any real number can be squared, if no other domain is specified, we consider the domain of $f(x)=x^2$ to be the set of all real numbers. On the other hand, the square root function $f(x)=\sqrt{x}$ only gives a real output if x is nonnegative. Therefore, the domain of the function $f(x)=\sqrt{x}$ is the set of nonnegative real numbers, sometimes called the *natural domain*.

For the functions $f(x)=x^2$ and $f(x)=\sqrt{x},$ the domains are sets with an infinite number of elements. Clearly we cannot list all these elements. When describing a set with an infinite number of elements, it is often helpful to use set-builder or interval notation. When using set-builder notation to describe a subset of all real numbers, denoted $\mathbb{R},$ we write

$$\{x \mid x \text{ has some property}\}.$$

We read this as the set of real numbers x such that x has some property. For example, if we were interested in the set of real numbers that are greater than one but less than five, we could denote this set using set-builder notation by writing

$$\{x \mid 1<x<5\}.$$

A set such as this, which contains all numbers greater than a and less than $b,$ can also be denoted using the interval notation $(a,b).$ Therefore,

$$(1,5)=\{x \mid 1<x<5\}.$$

The numbers 1 and 5 are called the *endpoints* of this set. If we want to consider the set that includes the endpoints, we would denote this set by writing

$$[1,5]=\{x \mid 1\le x\le 5\}.$$

We can use similar notation if we want to include one of the endpoints, but not the other. To denote the set of nonnegative real numbers, we would use the set-builder notation

$$\{x \mid 0\le x\}.$$

The smallest number in this set is zero, but this set does not have a largest number. Using interval notation, we would use the symbol $\infty,$ which refers to positive infinity, and we would write the set as

$$[0,\infty)=\{x \mid 0\le x\}.$$

It is important to note that ∞ is not a real number. It is used symbolically here to indicate that this set includes all real numbers greater than or equal to zero. Similarly, if we wanted to describe the set of all nonpositive numbers, we could write

(−∞,0]={x|x≤0}.(−∞,0]={x|x≤0}.

Here, the notation −∞−∞ refers to negative infinity, and it indicates that we are including all numbers less than or equal to zero, no matter how small. The set

(−∞,∞)={x|xis any real number}(−∞,∞)={x|xis any real number}

refers to the set of all real numbers.

Some functions are defined using different equations for different parts of their domain. These types of functions are known as *piecewise-defined functions*. For example, suppose we want to define a function ff with a domain that is the set of all real numbers such that f(x)=3x+1f(x)=3x+1 for x≥2x≥2 and f(x)=x2f(x)=x2 for x<2.x<2. We denote this function by writing

f(x)={3x+1x≥2x2x<2.f(x)={3x+1x≥2x2x<2.

When evaluating this function for an input x,x, the equation to use depends on whether x≥2x≥2 or x<2.x<2. For example, since 5>2,5>2, we use the fact that f(x)=3x+1f(x)=3x+1 for x≥2x≥2 and see that f(5)=3(5)+1=16.f(5)=3(5)+1=16. On the other hand, for x=−1,x=−1, we use the fact that f(x)=x2f(x)=x2 for x<2x<2 and see that f(−1)=1.f(−1)=1.

2.3 Inverse of Composition of Functions

- Compute the inverse of composition of two given functions.

This article includes a lot of function composition. If you need a review on this subject, we recommend that you go here before reading this article.

Inverse functions, in the most general sense, are functions that "reverse" each other. For example, if a function takes a*a*a to b*b*b, then the inverse must take b*b*b to a*a*a.

Let's take functions f*f*f and g*g*g for example: f(x)=\dfrac{x+1}{3}*f*(*x*)=3*x*+1f, left parenthesis, x, right parenthesis, equals, start fraction, x, plus, 1, divided by, 3, end fraction and g(x)=3x-1*g*(*x*)=3*x*−1g, left parenthesis, x, right parenthesis, equals, 3, x, minus, 1.

Notice how f(5)=2*f*(5)=2f, left parenthesis, 5, right parenthesis, equals, 2 and g(2)=5*g*(2)=5g, left parenthesis, 2, right parenthesis, equals, 5.

[Please show me the calculations]

f(5)f, left parenthesis, 5, right parenthesis55ff

\begin{aligned}f(x)&=\dfrac{x+1}{3}\\ \\ f(\blueD{5})&=\dfrac{\blueD5+1}{3}\\ \\ &=\goldD2 \end{aligned}

g(2)g, left parenthesis, 2, right parenthesis22gg

\begin{aligned}g(x)&=3x-1\\ \\ g(\goldD{2})&=3(\goldD2)-1\\ \\ &=\blueD5 \end{aligned}

\blueD{5}5ff\goldD{2}2gg\blueD{5}5

Here we see that when we apply f*f* followed by g*g*, we get the original input back. Written as a composition, this is g(f(5))=5*g*(*f*(5))=5g, left parenthesis, f, left parenthesis, 5, right parenthesis, right parenthesis, equals, 5.

But for two functions to be inverses, we have to show that this happens for *all possible inputs* regardless of the order in which f*f* and g*g* are applied. This gives rise to the inverse composition rule.

The inverse composition rule

These are the conditions for two functions f*f* and g*g* to be inverses:

f(g(x))=x*f*(*g*(*x*))=*x*f, left parenthesis, g, left parenthesis, x, right parenthesis, right parenthesis, equals, x for all x*x*x in the domain of g*g*
g(f(x))=x*g*(*f*(*x*))=*x*g, left parenthesis, f, left parenthesis, x, right parenthesis, right parenthesis, equals, x for all x*x*x in the domain of f*f*
This is because if f*f* and g*g* are inverses, composing f*f* and g*g* (in either order) creates the function that for every input returns that input. We call this function "the identity function".

Example 1: Functions f*f* and g*g* are inverses

Let's use the inverse composition rule to verify that f*f* and g*g* above are indeed inverse functions.

Recall that f(x)=\dfrac{x+1}{3}*f*(*x*)=3*x*+1f, left parenthesis, x, right parenthesis, equals, start fraction, x, plus, 1, divided by, 3, end fraction and g(x)=3x-1*g*(*x*)=3*x*−1g, left parenthesis, x, right parenthesis, equals, 3, x, minus, 1.

Let's find f(g(x))*f*(*g*(*x*))f, left parenthesis, g, left parenthesis, x, right parenthesis, right parenthesis and g(f(x))*g*(*f*(*x*))g, left parenthesis, f, left parenthesis, x, right parenthesis, right parenthesis.

f(g(x))*f*(*g*(*x*))f, left parenthesis, g, left parenthesis, x, right parenthesis, right parenthesis	**\qquad \qquad g(f(x))*g*(*f*(*x*))g, left parenthesis, f, left parenthesis, x, right parenthesis, right parenthesis**
\begin{aligned} f(\greenD{g(x)})&=\dfrac{\greenD{g(x)}+1}{3}\\\\&=\dfrac{\greenD{3x-1}+1}{3}\\\\&=\dfrac{3x}{3}\\\\&=x\\\end{aligned}*f*(*g*(*x*))=3*g*(*x*)+1=33*x*−1+1=33*x*=*x*	\qquad\qquad \begin{aligned}g(\purpleC{f(x)})&=3 \left(\purpleC{f(x)}\right)-1\\\\&=3\left(\purpleC{\dfrac{x+1}{3}}\right)-1\\\\&=x+1-1\\\\&=x\\\end{aligned}*g*(*f*(*x*)) =3(*f*(*x*))−1=3(3*x*+1)−1=*x*+1−1=*x*

So we see that functions f*f*f and g*g*g are inverses because f(g(x))=x*f*(*g*(*x*))=*x*f, left parenthesis, g, left parenthesis, x, right parenthesis, right parenthesis, equals, x and g(f(x))=x*g*(*f*(*x*))=*x*g, left parenthesis, f, left parenthesis, x, right parenthesis, right parenthesis, equals, x.

2.4 Transcendental functions

- Identify algebraic, trigonometric, inverse trigonometric, exponential, logarithmic, hyperbolic (and their identities), explicit and implicit functions, and parametric representation of functions

The hyperbolic functions appear with some frequency in applications, and are quite similar in many respects to the trigonometric functions. This is a bit surprising given our initial definitions.

Definition 4.11.1 The **hyperbolic cosine** is the function

coshx=ex+e−x2,cosh x=ex+e−x2,

and the **hyperbolic sine** is the function

sinhx=ex−e−x2.sinh x=ex−e−x2.

Notice that $\cosh$ is even (that is, $\cosh(-x)=\cosh(x)$) while $\sinh$ is odd ($\sinh(-x)=-\sinh(x)$), and $\cosh x+\sinh x=e^x$. Also, for all x, $\cosh x>0$, while $\sinh x=0$ if and only if $e^x-e^{-x}=0$, which is true precisely when $x=0$.

Lemma 4.11.2 *The range of* $\cosh x$ *is* $[1,\infty)$.

Proof.
Let $y=\cosh x$. We solve for x:

$$
\begin{aligned}
y &= \frac{e^x+e^{-x}}{2}\\
2y &= e^x+e^{-x}\\
2ye^x &= e^{2x}+1\\
0 &= e^{2x}-2ye^x+1\\
e^x &= \frac{2y\pm\sqrt{4y^2-4}}{2}\\
e^x &= y\pm\sqrt{y^2-1}
\end{aligned}
$$

From the last equation, we see $y^2\ge 1$, and since $y\ge 0$, it follows that $y\ge 1$.

Now suppose $y\ge 1$, so $y\pm\sqrt{y^2-1}>0$. Then $x=\ln(y\pm\sqrt{y^2-1})$ is a real number, and $y=\cosh x$, so y is in the range of $\cosh(x)$.

□

Definition 4.11.3 The other hyperbolic functions are

$$
\begin{aligned}
\tanh x &= \frac{\sinh x}{\cosh x}\\
\coth x &= \frac{\cosh x}{\sinh x}\\
\operatorname{sech} x &= \frac{1}{\cosh x}\\
\operatorname{csch} x &= \frac{1}{\sinh x}
\end{aligned}
$$

The domain of $\coth$ and csch is $x\ne 0$ while the domain of the other hyperbolic functions is all real numbers. Graphs are shown in figure 4.11.1

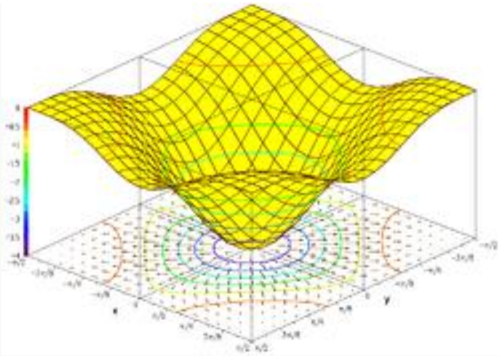
<="" div="" style="zoom: 0; position: relative; overflow: hidden; background-color: rgb(255, 255, 255); border-style: solid; border-width: 1px; border-color: rgb(53, 106, 160); border-radius: 10px; margin-left: auto; margin-right: auto; height: 207px;"> -1 1 1 2 3	-1 1 1 -1 2 -2	-1 1 1 -1 2 -2

Coshcosh	sinhsinh	tanhtanh
−1−1 11 11 −1−1 22 −2−2	−1−1 11 11 −1−1 22 −2−2	−1−1 11 11 −1−1 22 −2−2
Sechsech	cschcsch	cothcoth

Figure 4.11.1. *The hyperbolic functions.*

Certainly the hyperbolic functions do not closely resemble the trigonometric functions graphically. But they do have analogous properties, beginning with the following identity.

Functions of two variables[edit]



Plot of the graph of $f(x, y) = -(\cos(x^2) + \cos(y^2))^2$, also showing its gradient projected on the bottom plane.

The graph of the trigonometric function

is

If this set is plotted on a three dimensional Cartesian coordinate system, the result is a surface (see figure).

Oftentimes it is helpful to show with the graph, the gradient of the function and several level curves. The level curves can be mapped on the function surface or can be projected on the bottom plane. The second figure shows such a drawing of the graph of the function:

Generalizations[edit]

The graph of a function is contained in a Cartesian product of sets. An X–Y plane is a cartesian product of two lines, called X and Y, while a cylinder is a cartesian product of a line and a circle, whose height, radius, and angle assign precise locations of the points. Fibre bundles are not Cartesian products, but appear to be up close. There is a corresponding notion of a graph on a fibre bundle called a section.

2.6 Limit of a function

- Identify a real number by a point on the number line
- Define and represent
  I. Open interval,
  II. Closed interval,
  III. Half open and half closed intervals on the number line.

- Explain the meaning of phrase:
  I. x tends to zero ($x \to 0$),
  II. X tends to a zero ($x \to \alpha$),
  III. X tends to infinity ($x \to \infty$)
- Define limit of a sequence.
- Find the limit of a sequence whose nth term is given.
- Define limit of a function.
- State the theorems on limits of sum, difference, product and quotient of functions and demonstrate through examples.
- Apply the theorems on limits of sum, difference, product and quotient of functions and demonstrate through examples.

The Definition of a Limit

- Return to the *Limits and l'Hôpital's Rule* starting page

Choose one of the following topics:

- A list of basic limit laws
- The statement of l'Hôpital's Rule

While a table of numbers can certainly suggest that a limit has a certain value, it cannot definitely prove that the limit has that value. For instance, look at the function we looked at earlier.

$$f(x) = \frac{x^3 - 4x^2 + x + 6}{x^2 - 5x + 6}$$

The table of values we found earlier for *x* nearing 2 was:

f(1.5)	=	2.5	*f*(2.5)	=	3.5
f(1.7)	=	2.7	*f*(2.3)	=	3.3
f(1.8)	=	2.8	*f*(2.2)	=	3.2
f(1.9)	=	2.9	*f*(2.1)	=	3.1
f(1.99)	=	2.99	*f*(2.01)	=	3.01
f(1.999)	=	2.999	*f*(2.001)	=	3.001

The values certainly look like they are approaching 3, but how do we know for certain? Perhaps they are approaching 3.000075 or 2.999996. What we need is a precise definition of a limit, which will tell us when we are exactly correct.

Definition of Limit

Let *f* be a function defined on some open interval that contains the number *a*, except possibly at *a* itself. We say *the limit of f(x) as x approaches a is L*, and we write

$$\lim_{x \to a} f(x) = L$$

if for every number $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that

$$|f(x) - L| < \epsilon$$

whenever

$$0 < |x - a| < \delta.$$

Using our example function, *f*(*x*) is as above, *a*=2, and we think that *L*=3. Let ϵ be any positive number. Then we can choose δ to be equal to ϵ. In which case, if $0 < |x-2| < \delta$,

$$|f(x) - L| = \left| \frac{x^3 - 4x^2 + x + 6}{x^2 - 5x + 6} - 3 \right| =$$

$$\left| \frac{x^3 - 4x^2 + x + 6}{x^2 - 5x + 6} - \frac{3x^2 - 15x + 18}{x^2 - 5x + 6} \right| = \left| \frac{x^3 - 7x^2 + 16x - 12}{x^2 - 5x + 6} \right| =$$

$$\left|\frac{(x-3)(x-2)(x-2)}{(x-3)(x-2)}\right|.$$

We can cancel the (x-3) term from the top and bottom of the fraction, as well as one of the (x-2) terms from the top with the same on the bottom, to get

$$|f(x)-L| = |x-2| < \delta = \epsilon.$$

So, since the definition fits exactly, we can state with certainty that the limit as x approaches 2 of $f(x)$ is 3. Also, we can notice that if we used any value other than 3 for L, we wouldn't get the cancellation we did in the last step, so we would not have been able to fit the definition.

Below are the definitions of several related limit concepts.

Left-Hand [or Right-Hand] Limit

Let f be a function defined on some open interval (b,a) [or (a,b)]. We say *the left-hand [or right-hand] limit of f(x) as x approaches a is L*, (or *the limit of f(x) as x approaches a from the left [or right] is L*) and we write

$$\lim_{x\to a^-} f(x) = L \left[\lim_{x\to a^+} f(x) = L\right]$$

if for every number $\epsilon > 0$ there is a corresponding number $\delta > 0$ such that

$$|f(x)-L| < \epsilon$$

whenever

$$a-\delta < x < a\,[a < x < a+\delta].$$

[Negative] Infinite Limit

Let f be a function defined on some open interval that contains the number a, except possibly at a itself. We say *the limit of f(x) as x approaches a is [negative] infinity*, and we write

$$\lim_{x\to a} f(x) = [-]\infty$$

if for every number $N > 0$ [$N < 0$] there is a corresponding number $\delta > 0$ such that

$$f(x) > N\ [f(x) < N]$$

whenever

$$0 < |x - a| < \delta.$$

Limit at [Negative] Infinity

Let *f* be a function defined on some open interval from *a* to infinity [from negative infinity to *a*]. Then we say *the limit of f(x) as x approaches [negative] infinity is L*, and we write

$$\lim_{x \to [-]\infty} f(x) = L$$

if for every number $\epsilon > 0$ there is a corresponding number *N* such that

$$|f(x) - L| < \epsilon$$

whenever

$$x > N \; [x < N].$$

[Negative] Infinite Limit at {Negative} Infinity

Let *f* be a function defined on some open interval from *a* to infinity {from negative infinity to *a*}. We say *the limit of f(x) as x approaches {negative} infinity is [negative] infinity*, and we write

$$\lim_{x \to \{-\}\infty} f(x) = [-]\infty$$

if for every number $M > 0$ there is a corresponding number *N* such that

$$f(x) > M \; [f(x) < M]$$

whenever

$$x > N \; \{x < N\}.$$

Similar definitions can be made for infinite-valued left- or right-hand limits.

2.7 Important Limits Evaluate the limits of functions of the following types:

- $\frac{x^n - a^n}{x - a}, \frac{x - a}{\sqrt{x} - \sqrt{a}}$ *when* $x \to a,$

- $(1+\frac{1}{x})^x \ when\ x \to \infty,$
- $(1+x)^{\frac{1}{z}}, \frac{\sqrt{x+a}-\sqrt{a}}{x}, \frac{a^z-1}{x},$
- $\frac{(1+x)^n-1}{x}, and\ \frac{\sin x}{x}\ when\ x \to 0$

- Evaluate limits of different algebraic, exponential and trigonometric functions
- Apply MAPLE command limit to evaluate limit of a functions.

Formulas of Useful Limits

1) If limx→af(x)=llimx→af(x)=l and limx→ag(x)=mlimx→ag(x)=m, then

- limx→a[f(x)±g(x)]=l±mlimx→a[f(x)±g(x)]=l±m
- limx→af(x)·g(x)=l·mlimx→af(x)·g(x)=l·m
- limx→af(x)g(x)=lmlimx→af(x)g(x)=lm, where m≠0m≠0
- limx→ac f(x)=c llimx→ac f(x)=c l
- limx→a1f(x)=1llimx→a1f(x)=1l, where l≠0l≠0

2) limn→∞(1+1n)n=elimn→∞(1+1n)n=e, where nn is a real number.
3) limn→0(1+n)1n=elimn→0(1+n)1n=e, where nn is a real number.
4) limx→0Sinxx=1limx→0Sinxx=1, where xx is measured in radians.
5) limx→0Tanxx=1limx→0Tanxx=1
6) limx→0Cosx–1x=0limx→0Cosx–1x=0

7) limx→axn–anx–a=nan–1limx→axn–anx–a=nan–1

8) limx→0an–1x=lnalimx→0an–1x=lna

Basic Derivative ResultsFormulas and Results of Complex Numbers ⇒

2.8 Continuous Discontinuous Functions

- Describe left hand and right hand limits through examples.
- Define continuity of function at a point and in an interval.
- Evaluate test continuity and discontinuity of a function at a point and in an interval
- Apply MAPLE command is cont to test continuity of a function at a point and in a given interval.

What is a left-hand limit?

Precalculus_ Limits_ One-Sided Limits

1 Answer

seph

Oct 19, 2014

A left-hand limit means the limit of a function as it approaches from the left-hand side.
On the other hand, A right-hand limit means the limit of a function as it approaches from the right-hand side.
When getting the limit of a function as it approaches a number, the idea is to check the behavior of the function as it approaches the number. We substitute values as close as possible to the number being approached.
The closest number is the number being approached itself. Hence, one usually just substitutes the number being approached to get the limit.
However, we cannot do this if the resulting value is undefined.
But we can still check its behavior as it approaches from one side.
One good example is limx→01x.
When we substitute x=0 into the function, the resulting value is undefined.
Let's check its limit as it approaches from the left-hand side
f(x)=1x
f(−1)=1−1=−1
f(−12)=1−12=−2
f(−110)=1−110=−10
f(−11000)=1−11000=−1000
f(−11000000)=1−11000000=−1000000
Notice that as we get closer and closer to x=0 from the left-hand side, the resulting value we gets larger and larger (though negative). We can conclude that the limit as x→0 from the left-hand side is −∞

Now let's check the limit from the right-hand side
f(x)=1x
f(1)=11=1
f(12)=112=2
f(110)=1110=10
f(11000)=111000=1000
f(11000000)=111000000=1000000
The limit as x→0 from the right-hand side is ∞

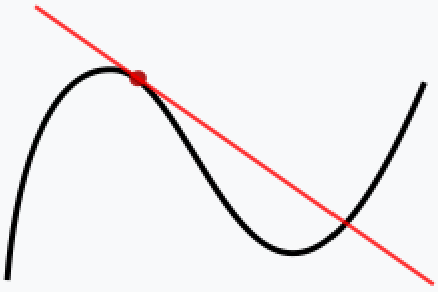
When the left-hand side limit of a function is different from the right-hand side limit, we can conclude that the function is discontinuous at the number being approached.

UNIT 3 DIFFERENTIATION

3.1 Derivative of a Functions

- Differentiate between independent and dependent variables.
- Evaluate corresponding change in the dependent variable when independent variable is incremented (or decremented).
- Describe the concept of rate of change.
- Define derivative of a function as an instantaneous rate of change of a variable with respect to another variable.
- Explain derivative or differential coefficient of a function.
- Find the derivative $y=x^n$, where neZ (the set of integers), from first principles.
- Find the derivative by first principles of $y=(ax+b)^n$, where $n=\frac{p}{q}$ and p,q are integers such that $q\neq0$

This article is about the term as used in calculus. For a less technical overview of the subject, see differential calculus. For other uses, see Derivative (disambiguation).



The graph of a function, drawn in black, and a tangent line to that function, drawn in red. The slope of the tangent line is equal to the derivative of the function at the marked point.

Part of a series of articles about **Calculus**
• Fundamental theorem • Limits of functions • Continuity • Mean value theorem • Rolle's theorem
Differential[hide] **Definitions** • Derivative (generalizations)

- Differential
  - infinitesimal
  - of a function
    - total

Concepts

- Differentiation notation
- Second derivative
- Third derivative
- Change of variables
- Implicit differentiation
- Related rates
- Taylor's theorem

Rules and identities

- Sum
- Product
- Chain
- Power
- Quotient
- Inverse
- General Leibniz
- Faà di Bruno's formula

Integral[show]

Series[show]

Vector[show]

Multivariable[show]

Specialized[show]

Glossary of calculus[show]

- v
- t
- e

The **derivative** of a function of a real variable measures the sensitivity to change of the function value (output value) with respect to a change in its argument (input value). Derivatives are a fundamental tool of calculus. For example, the derivative of the position of a moving object with respect to time is the object's velocity: this measures how quickly the position of the object changes when time advances.

The derivative of a function of a single variable at a chosen input value, when it exists, is the slope of the tangent line to the graph of the function at that point. The tangent line is the best linear approximation of the function near that input value. For this reason, the derivative is often described as the "instantaneous rate of change", the ratio of the instantaneous change in the dependent variable to that of the independent variable.

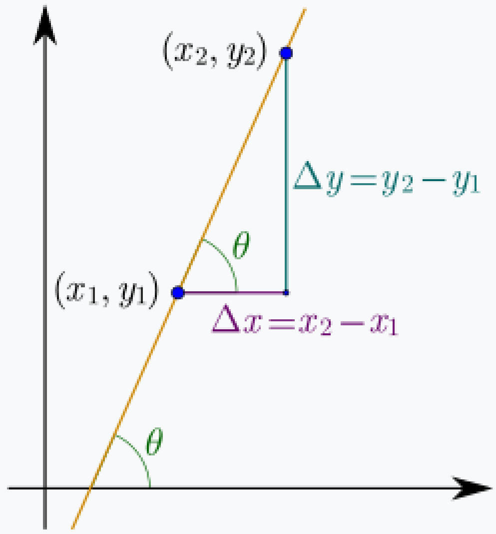
Derivatives may be generalized to functions of several real variables. In this generalization, the derivative is reinterpreted as a linear transformation whose graph is (after an appropriate translation) the best linear approximation to the graph of the original function. The Jacobian matrix is the matrix that represents this linear transformation with respect to the basis given by the choice of independent and dependent variables. It can be calculated in terms of the partial derivatives with respect to the independent variables. For a real-valued function of several variables, the Jacobian matrix reduces to the gradient vector.

The process of finding a derivative is called **differentiation**. The reverse process is called *antidifferentiation*. The fundamental theorem of calculus relates antidifferentiation with integration. Differentiation and integration constitute the two fundamental operations in single-variable calculus.[Note 1]

Contents

Differentiation[edit]

Differentiation is the action of computing a derivative. The derivative of a function *y* = *f*(*x*) of a variable *x* is a measure of the rate at which the value *y* of the function changes with respect to the change of the variable *x*. It is called the *derivative* of *f* with respect to *x*. If *x* and *y* are real numbers, and if the graph of *f* is plotted against *x*, the derivative is the slope of this graph at each point.



Slope of a linear function:

The simplest case, apart from the trivial case of a constant function, is when *y* is a linear function of *x*, meaning that the graph of *y* is a line. In this case, *y* = *f*(*x*) = *mx* + *b*, for real numbers *m* and *b*, and the slope *m* is given by

where the symbol Δ (Delta) is an abbreviation for "change in", and the combinations and refer to corresponding changes, i.e.: . The above formula holds because

Thus

This gives the value for the slope of a line.

If the function *f* is not linear (i.e. its graph is not a straight line), then the change in *y* divided by the change in *x* varies over the considered range: differentiation is a method to find a unique value for this rate of change, not

across a certain range but at any given value of *x*.

Rate of change as a limit value

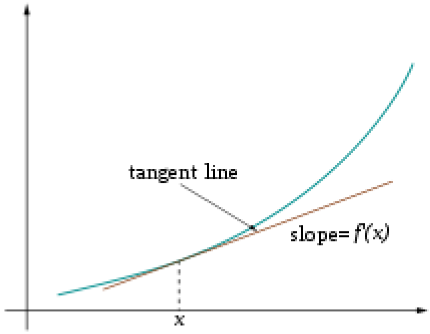


Figure 1. The tangent line at (*x*, *f*(*x*))

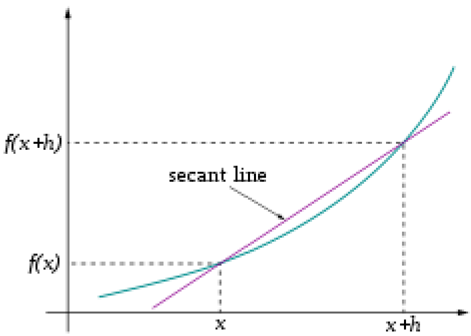


Figure 2. The secant to curve *y*= *f*(*x*) determined by points (*x*, *f*(*x*)) and (*x* + *h*, *f*(*x* + *h*))

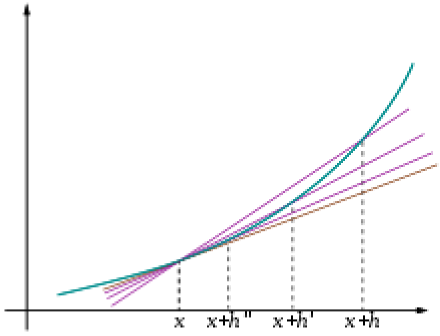


Figure 3. The tangent line as limit of secants

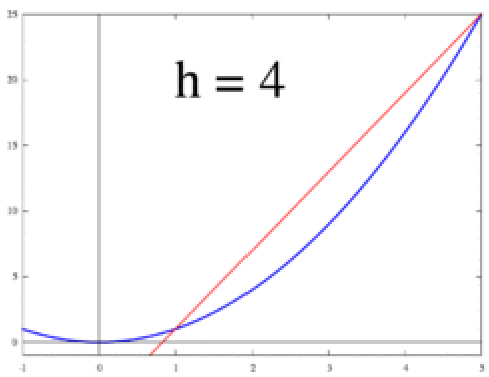


Figure 4. Animated illustration: the tangent line (derivative) as the limit of secants

The idea, illustrated by Figures 1 to 3, is to compute the rate of change as the limit value of the ratio of the differences $\Delta y / \Delta x$ as Δx becomes infinitely small.

Notation[edit]

Main article: Notation for differentiation

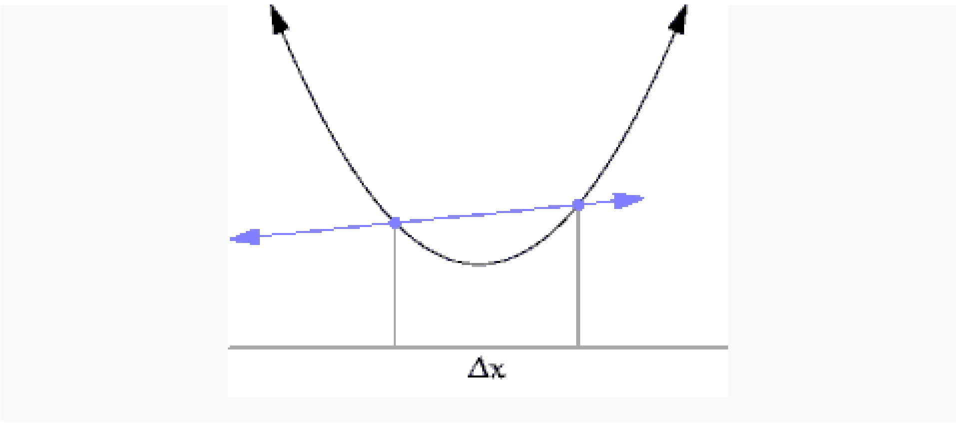
Two distinct notations are commonly used for the derivative, one deriving from Leibniz and the other from Joseph Louis Lagrange.

In Leibniz's notation, an infinitesimal change in *x* is denoted by *dx*, and the derivative of *y* with respect to *x* is written

suggesting the ratio of two infinitesimal quantities. (The above expression is read as "the derivative of *y* with respect to *x*", "*dy* by *dx*", or "*dy* over *dx*". The oral form "*dy dx*" is often used conversationally, although it may lead to confusion.)

In Lagrange's notation, the derivative with respect to *x* of a function $f(x)$ is denoted $f'(x)$ (read as "*f* prime of *x*") or $f_x'(x)$ (read as "*f* prime *x* of *x*"), in case of ambiguity of the variable implied by the differentiation. Lagrange's notation is sometimes incorrectly attributed to Newton.

Rigorous definition[edit]



A secant approaches a tangent when .

The most common approach to turn this intuitive idea into a precise definition is to define the derivative as a limit of difference quotients of real numbers.[1] This is the approach described below.

Let *f* be a real valued function defined in an open neighborhood of a real number *a*. In classical geometry, the tangent line to the graph of the function *f* at *a* was the unique line through the point (*a*, *f*(*a*)) that did *not* meet the graph of *f* transversally, meaning that the line did not pass straight through the graph. The derivative of *y* with respect to *x* at *a* is, geometrically, the slope of the tangent line to the graph of *f* at (*a*, *f*(*a*)). The slope of the tangent line is very close to the slope of the line through (*a*, *f*(*a*)) and a nearby point on the graph, for example (*a* + *h*, *f*(*a* + *h*)). These lines are called secant lines. A value of *h* close to zero gives a good approximation to the slope of the tangent line, and smaller values (in absolute value) of *h* will, in general, give better approximations. The slope *m* of the secant line is the difference between the *y* values of these points divided by the difference between the *x* values, that is,

This expression is Newton's difference quotient. Passing from an approximation to an exact answer is done using a limit. Geometrically, the limit of the secant lines is the tangent line. Therefore, the limit of the difference quotient as *h* approaches zero, if it exists, should represent the slope of the tangent line to (*a*, *f*(*a*)). This limit is defined to be the derivative of the function *f* at *a*:

When the limit exists, *f* is said to be ***differentiable*** at *a*. Here *f*′(*a*) is one of several common notations for the derivative (see below). From this definition it is obvious that a differentiable function *f* is increasing if and only if its derivative is positive, and is decreasing iff its derivative is negative. This fact is used extensively when analyzing function behavior, e.g. when finding local extrema.

Equivalently, the derivative satisfies the property that

which has the intuitive interpretation (see Figure 1) that the tangent line to *f* at *a* gives the *best linear approximation*

to *f* near *a* (i.e., for small *h*). This interpretation is the easiest to generalize to other settings (see below).

Substituting 0 for h in the difference quotient causes division by zero, so the slope of the tangent line cannot be found directly using this method. Instead, define $Q(h)$ to be the difference quotient as a function of h:

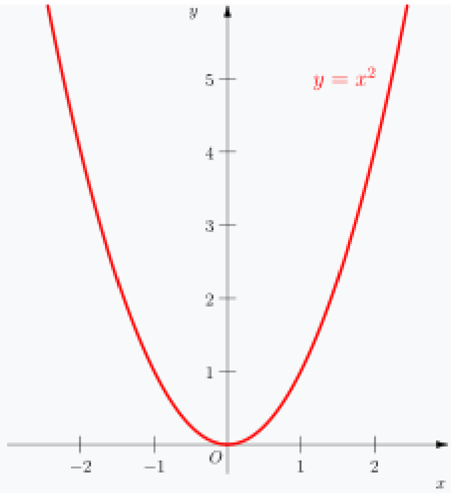
$Q(h)$ is the slope of the secant line between $(a, f(a))$ and $(a + h, f(a + h))$. If f is a continuous function, meaning that its graph is an unbroken curve with no gaps, then Q is a continuous function away from $h = 0$. If the limit $\lim_{h\to 0} Q(h)$ exists, meaning that there is a way of choosing a value for $Q(0)$ that makes Q a continuous function, then the function f is differentiable at a, and its derivative at a equals $Q(0)$.

In practice, the existence of a continuous extension of the difference quotient $Q(h)$ to $h = 0$ is shown by modifying the numerator to cancel h in the denominator. Such manipulations can make the limit value of Q for small h clear even though Q is still not defined at $h = 0$. This process can be long and tedious for complicated functions, and many shortcuts are commonly used to simplify the process.

Definition over the hyperreals[edit]

Relative to a hyperreal extension $\mathbf{R} \subset {}^*\mathbf{R}$ of the real numbers, the derivative of a real function $y = f(x)$ at a real point x can be defined as the shadow of the quotient $\Delta y/\Delta x$ for infinitesimal Δx, where $\Delta y = f(x + \Delta x) - f(x)$. Here the natural extension of f to the hyperreals is still denoted f. Here the derivative is said to exist if the shadow is independent of the infinitesimal chosen.

Example[edit]



The square function

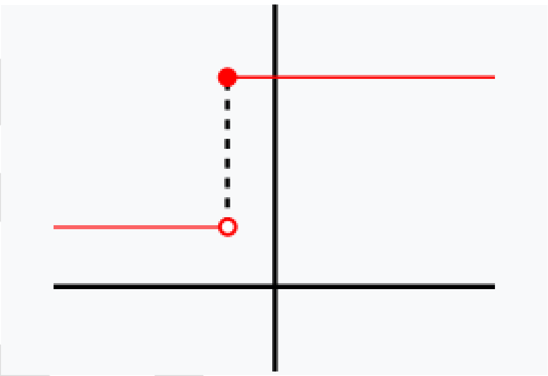
The square function given by $f(x) = x^2$ is differentiable at $x = 3$, and its derivative there is 6. This result is established by calculating the limit as h approaches zero of the difference quotient of $f(3)$:

The last expression shows that the difference quotient equals $6 + h$ when $h \neq 0$ and is undefined when $h = 0$, because of the definition of the difference quotient. However, the definition of the limit says the difference quotient does not need to be defined when $h = 0$. The limit is the result of letting h go to zero, meaning it is the value that $6 + h$ tends to as h becomes very small:

Hence the slope of the graph of the square function at the point (3, 9) is 6, and so its derivative at $x = 3$ is $f'(3) = 6$.

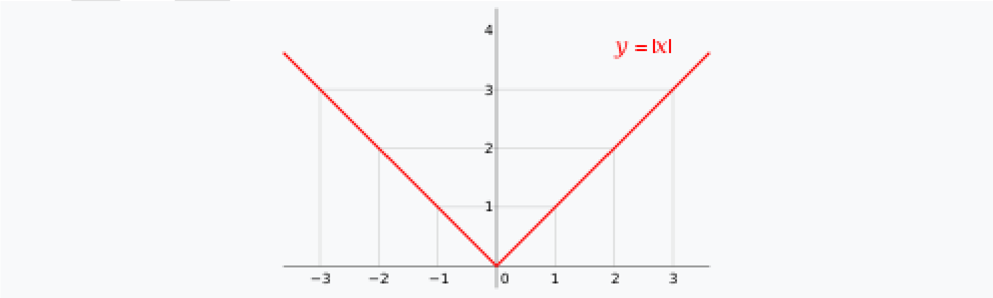
More generally, a similar computation shows that the derivative of the square function at $x = a$ is $f'(a) = 2a$:

Continuity and differentiability[edit]



This function does not have a derivative at the marked point, as the function is not continuous there (specifically, it has a jump discontinuity).

If f is differentiable at a, then f must also be continuous at a. As an example, choose a point a and let f be the step function that returns the value 1 for all x less than a, and returns a different value 10 for all x greater than or equal to a. f cannot have a derivative at a. If h is negative, then $a + h$ is on the low part of the step, so the secant line from a to $a + h$ is very steep, and as h tends to zero the slope tends to infinity. If h is positive, then $a + h$ is on the high part of the step, so the secant line from a to $a + h$ has slope zero. Consequently, the secant lines do not approach any single slope, so the limit of the difference quotient does not exist.[Note 2]



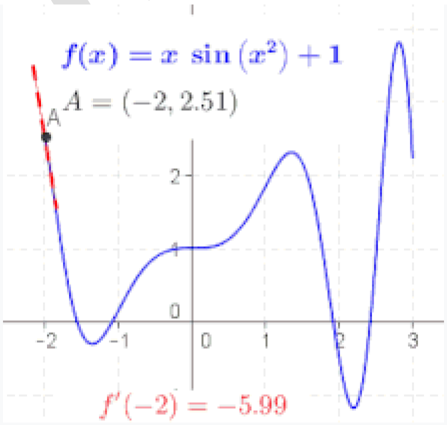
The absolute value function is continuous, but fails to be differentiable at $x = 0$ since the tangent slopes do not approach the same value from the left as they do from the right.

However, even if a function is continuous at a point, it may not be differentiable there. For example, the absolute value function given by $f(x) = |x|$ is continuous at $x = 0$, but it is not differentiable there. If h is positive, then the slope of the secant line from 0 to h is one, whereas if h is negative, then the slope of the secant line from 0 to h is negative one. This can be seen graphically as a "kink" or a "cusp" in the graph at $x = 0$. Even a function with a smooth graph is not differentiable at a point where its tangent is vertical: For instance, the function given by $f(x) = x^{1/3}$ is not differentiable at $x = 0$.

In summary, a function that has a derivative is continuous, but there are continuous functions that do not have a derivative.

Most functions that occur in practice have derivatives at all points or at almost every point. Early in the history of calculus, many mathematicians assumed that a continuous function was differentiable at most points. Under mild conditions, for example if the function is a monotone function or a Lipschitz function, this is true. However, in 1872 Weierstrass found the first example of a function that is continuous everywhere but differentiable nowhere. This example is now known as the Weierstrass function. In 1931, Stefan Banach proved that the set of functions that have a derivative at some point is a meager set in the space of all continuous functions.[2] Informally, this means that hardly do any random continuous functions have a derivative at even one point.

The derivative as a function[edit]



The derivative at different points of a differentiable function. In this case, the derivative is equal to:

Let f be a function that has a derivative at every point in its domain. We can then define a function that maps every point to the value of the derivative of at . This function is written f' and is called the *derivative function* or the *derivative of f*.

Sometimes f has a derivative at most, but not all, points of its domain. The function whose value at a equals $f'(a)$ whenever $f'(a)$ is defined and elsewhere is undefined is also called the derivative of f. It is still a function, but its domain is strictly smaller than the domain of f.

Using this idea, differentiation becomes a function of functions: The derivative is an operator whose domain is the set of all functions that have derivatives at every point of their domain and whose range is a set of functions. If we denote this operator by *D*, then *D*(*f*) is the function *f*′. Since *D*(*f*) is a function, it can be evaluated at a point *a*. By the definition of the derivative function, *D*(*f*)(*a*) = *f*′(*a*).

For comparison, consider the doubling function given by *f*(*x*) = 2*x*; *f* is a real-valued function of a real number, meaning that it takes numbers as inputs and has numbers as outputs:

The operator *D*, however, is not defined on individual numbers. It is only defined on functions:

Because the output of *D* is a function, the output of *D* can be evaluated at a point. For instance, when *D* is applied to the square function, $x \mapsto x^2$, *D* outputs the doubling function $x \mapsto 2x$, which we named *f*(*x*). This output function can then be evaluated to get *f*(1) = 2, *f*(2) = 4, and so on.

Higher derivatives[edit]

Let *f* be a differentiable function, and let *f* ′ be its derivative. The derivative of *f* ′ (if it has one) is written *f* ′′ and is called the ***second derivative*** *of f*. Similarly, the derivative of the second derivative, if it exists, is written *f* ′′′ and is called the ***third derivative*** *of f*. Continuing this process, one can define, if it exists, the *n*th derivative as the derivative of the (*n*-1)th derivative. These repeated derivatives are called *higher-order derivatives*. The *n*th derivative is also called the **derivative of order *n***.

If *x*(*t*) represents the position of an object at time *t*, then the higher-order derivatives of *x* have specific interpretations in physics. The first derivative of *x* is the object's velocity. The second derivative of *x* is the acceleration. The third derivative of *x* is the jerk. And finally, the fourth derivative of *x* is the jounce.

A function *f* need not have a derivative (for example, if it is not continuous). Similarly, even if *f* does have a derivative, it may not have a second derivative. For example, let

Calculation shows that *f* is a differentiable function whose derivative

at is given by

f′(*x*) is twice the absolute value function at , and it does not have a derivative at zero. Similar examples show that a function can have a *k*th derivative for each non-negative integer *k* but not a (*k* + 1)th derivative. A function that has *k* successive derivatives is called *k times differentiable*. If in addition the *k*th derivative is continuous, then the function is said to be of differentiability class C^k. (This is a stronger condition than having *k* derivatives, as shown by the second example of Smoothness § Examples.) A function that has infinitely many derivatives is called *infinitely differentiable* or *smooth*.

On the real line, every polynomial function is infinitely differentiable. By standard differentiation rules, if a polynomial of degree *n* is differentiated *n* times, then it becomes a constant function. All of its subsequent derivatives are identically zero. In particular, they exist, so polynomials are smooth functions.

The derivatives of a function *f* at a point *x* provide polynomial approximations to that function near *x*. For example, if *f* is twice differentiable, then

in the sense that

If *f* is infinitely differentiable, then this is the beginning of the Taylor series for *f* evaluated at *x* + *h* around *x*.

Inflection point[edit]

Main article: Inflection point

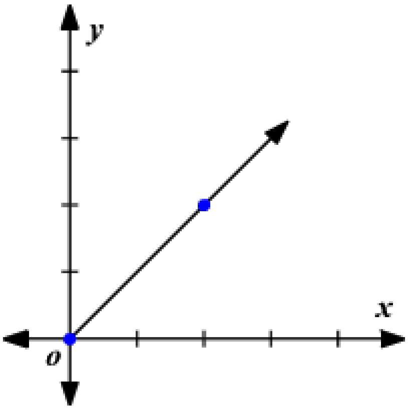
A point where the second derivative of a function changes sign is called an *inflection point*.[3] At an inflection point, the second derivative may be zero, as in the case of the inflection point *x* = 0 of the function given by , or it may fail to exist, as in the case of the inflection point *x* = 0 of the function given by . At an inflection point, a function switches from being a convex function to being a concave function or vice versa.

Rate of Change

A rate of change is a rate that describes how one quantity changes in relation to another quantity. If xx is the independent variable and yy is the dependent variable, then
rate of change=change in ychange in xrate of change=change in ychange in x
Rates of change can be positive or negative. This corresponds to an increase or decrease in the yy -value between the two data points. When a quantity does not change over time, it is called zero rate of change.

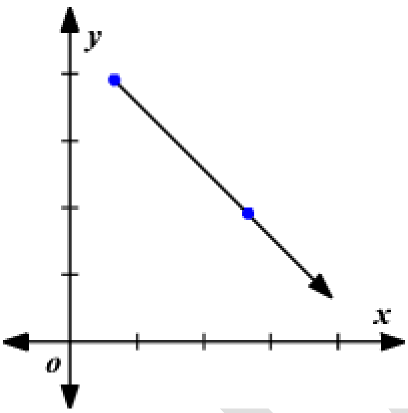
Positive rate of change

When the value of xx increases, the value of yy increases and the graph slants upward.



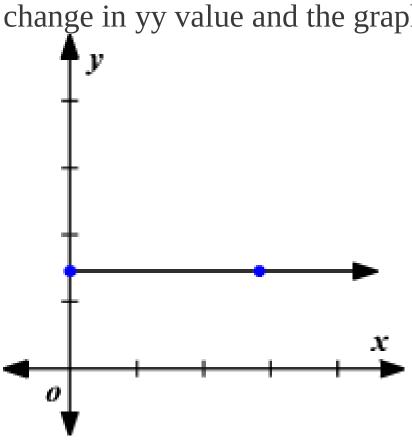
Negative rate of change

When the value of xx increases, the value of yy decreases and the graph slants downward.



Zero rate of change

When the value of xx increases, the value of yy remains constant. That is, there is no change in yy value and the graph is a horizontal line .



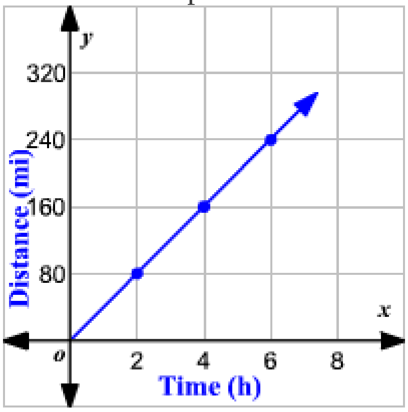
Example:

Use the table to find the rate of change. Then graph it.

Time driving (h) x246Distance Travelled (mi) y80160240Time Driving (h) xDistance Travelled (mi) y28041606240
A rate of change is a rate that describes how one quantity changes in relation to another

quantity.

rate of change=change in ychange in x=change in distancechange in time=160−804−2=802=401rate of change=change in ychange in x=change in distancechange in time=160−804−2=802=401
The rate of change is 401401 or 4040. This means a vehicle is traveling at a rate of 4040 miles per hour.



3.2 Prove the following theorems for differentiation.

I. The derivative of a constant is zero.
II. The derivative of any constant multiple of a function is equal to the product of that constant and the derivative of the function.
III. The derivative of a sum (or difference) of two functions is equal to the sum (or difference) of their derivatives.
IV. Derivative of two product functions
V. $(u\frac{d}{dx}v+v\frac{d}{dx}u)$
VI. Derivative of two quotient functions
VII. $\frac{(v\frac{d}{dx}u-u\frac{d}{dx}v)}{v2}$

6. Derivatives of Products and Quotients

by M. Bourne

PRODUCT RULE

If *u* and *v* are two functions of *x*, then the derivative of the product *uv* is given by...

Don't miss...

Later in this section:
Quotient Rule

\displaystyle\frac{{{d}{\left({u}{v}\right)}}}{{{\left.{d}{x}\right.}}}={u}\frac{{{d}{v}}}{{{\left.{d}{x}\right.}}}+{v}\frac{{{d}{u}}}{{\left.{d}{x}\right.}}*dxd*(*uv*)=*udxdv*+*vdxdu*

In words, this can be remembered as:

"The derivative of a product of two functions is the first times the derivative of the second, plus the second times the derivative of the first."

Where does this formula come from? Like all the differentiation formulas we meet, it is based on derivative from first principles.

Example 1

If we have a product like

$y = (2x^2 + 6x)(2x^3 + 5x^2)$

we can find the derivative without multiplying out the expression on the right.

Answer

Example 2

Find the derivative of

$y = (x^3 - 6x)(2 - 4x^3)$

Answer

Why not just multiply them out?

In Example 1 and Example 2, it appears that it would be easier to just expand out the brackets first, then differentiate the result.

However, later we'll need to differentiate functions such as \displaystyle{y}=\sqrt{{{x}^{2}-{3}{x}}}{\left(\sin{{4}}{x}^{2}\right)}*y*=*x*2−3*x*(sin4*x*2) (in the chapter Differentiation of Transcendental Functions.) It is not possible to multiply this expression term-by-term, so we need a method to differentiate products of such functions.

Note

We can write the product rule in many different ways:

\displaystyle\frac{{{d}{\left({u}{v}\right)}}}{{{\left.{d}{x}\right.}}}={u}{v}\text{'}+{v}{u}\text{'}*dxd*(*uv*)=*uv*’+*vu*’

OR

\displaystyle\frac{{{d}{\left({f}{g}\right)}}}{{{\left.{d}{x}\right.}}}*dxd*(*fg*) \displaystyle= f{{\left({x}\right)}}\frac{d}{{{\left.{d}{x}\right.}}} g{{\left({x}\right)}}+ g{{\left({x}\right)}}\frac{d}{{{\left.{d}{x}\right.}}} f{{\left({x}\right)}}=*f*(*x*)*dxdg*(*x*)+*g*(*x*)*dxdf*(*x*)… etc.

Continues below ⇩

QUOTIENT RULE

(A **quotient** is just a fraction.)

If *u* and *v* are two functions of *x*, then the derivative of the quotient \displaystyle\frac{u}{{v}}*vu* is given by...

\displaystyle\frac{d}{{{\left.{d}{x}\right.}}}{\left(\frac{u}{{v}}\right)}=\frac{{{v}\frac{{{d}{u}}}{{{\left.{d}{x}\right.}}}-{u}\frac{{{d}{v}}}{{{\left.{d}{x}\right.}}}}}{{{v}^{2}}}*dxd*(*vu*)=*v*2*vdxdu*−*udxdv*

In words, this can be remembered as:

"The derivative of a quotient equals bottom times derivative of top minus top times derivative of the bottom, divided by bottom squared."

Example 3

We wish to find the derivative of the expression:

\displaystyle{y}=\frac{{{2}{x}^{3}}}{{{4}-{x}}}*y*=4−*x*2*x*3

Answer

Example 4

Find \displaystyle\frac{{{\left.{d}{y}\right.}}}{{{\left.{d}{x}\right.}}}*dxdy*
if \displaystyle{y}=\frac{{{4}{x}^{2}}}{{{x}^{3}+{3}}}*y*=*x*3+34*x*2

Answer

Interactives

Go to the differentiation applet to explore Examples 3 and 4 and see what we've found.

Other ways of Writing Quotient Rule

You can also write **quotient rule** as:

\displaystyle\frac{d}{{{\left.{d}{x}\right.}}}{\left(\frac{f}{{g}}\right)}=\frac{{{g}\
\frac{{{d}{f}}}{{{\left.{d}{x}\right.}}}-{f}\
\frac{{{d}{g}}}{{{\left.{d}{x}\right.}}}}}{{{{g}^{2}}}}*dxd*(*gf*)=*g*2*g dxdf*−*f dxdg*

OR

\displaystyle\frac{d}{{{\left.{d}{x}\right.}}}{\left(\frac{u}{{v}}\right)}=\frac{{{v}{u}'-{u}{v}'}}{{{v}^{2}}}*dxd*(*vu*)=*v*2*vu*′−*uv*′

3.3 Application of theorem on differentiation

- Differentiate:
  - I. Constant multiple of x^n,
  - II. Sum(or difference) of functions,
  - III. Polynomials,
  - IV. Product of functions
- Quotient of two functions.

Many important applied problems involve finding the best way to accomplish some task. Often this involves finding the maximum or minimum value of some function: the minimum time to make a certain journey, the minimum cost for doing a task, the maximum power that can be generated by a device, and so on. Many of these problems can be solved by finding the appropriate function and then using techniques of calculus to find the maximum or the minimum value required.

Generally such a problem will have the following mathematical form: Find the largest (or smallest) value of f(x)f(x) when a≤x≤ba≤x≤b. Sometimes aa or bb are infinite, but frequently the real world imposes some constraint on the values that xx may have.

Such a problem differs in two ways from the local maximum and minimum problems we encountered when graphing functions: We are interested only in the function between a and b, and we want to know the largest or smallest value that $f(x)$ takes on, not merely values that are the largest or smallest in a small interval. That is, we seek not a local maximum or minimum but a **global** maximum or minimum, sometimes also called an **absolute** maximum or minimum.

Any global maximum or minimum must of course be a local maximum or minimum. If we find all possible local extrema, then the global maximum, *if it exists*, must be the largest of the local maxima and the global minimum, *if it exists*, must be the smallest of the local minima. We already know where local extrema can occur: only at those points at which $f'(x)$ is zero or undefined. Actually, there are two additional points at which a maximum or minimum can occur if the endpoints a and b are not infinite, namely, at a and b. We have not previously considered such points because we have not been interested in limiting a function to a small interval. An example should make this clear.

-2

1

Figure 6.1.1. *The function* $f(x)=x^2$ *restricted to* $[-2,1]$

Example 6.1.1 Find the maximum and minimum values of $f(x)=x^2$ on the interval $[-2,1]$, shown in figure 6.1.1. We compute $f'(x)=2x$, which is zero at $x=0$ and is always defined.

Since $f'(1)=2$ we would not normally flag $x=1$ as a point of interest, but it is clear from the graph that *when* $f(x)$ *is restricted to* $[-2,1]$ *there is a local maximum at* $x=1$. Likewise we would not normally pay attention to $x=-2$, but since we have truncated f at -2 we have introduced a new local maximum there as well. In a technical sense nothing new is going on here: When we truncate f we actually create a new function, let's call it g, that is defined only on the interval $[-2,1]$. If we try to compute the derivative of this new function we actually find that it does not have a derivative at -2 or 1. Why? Because to compute the derivative at 1 we must compute the limit

$$\lim_{\Delta x\to 0}\frac{g(1+\Delta x)-g(1)}{\Delta x}.$$

This limit does not exist because when $\Delta x>0$, $g(1+\Delta x)$ is not defined. It is simpler, however, simply to remember that we must always check the endpoints.

So the function g, that is, f restricted to $[-2,1]$, has one critical value and two finite endpoints, any of which might be the global maximum or minimum. We could first determine which of these are local maximum or minimum points (or neither); then the largest local maximum must be the global maximum and the smallest local minimum must be the global minimum. It is usually easier, however, to compute the value of f at every point at which the global maximum or minimum might occur; the largest of these is the global maximum, the smallest is the global minimum.

So we compute $f(-2)=4$, $f(0)=0$, $f(1)=1$. The global maximum is 4 at $x=-2$ and the global minimum is 0 at $x=0$.

It is possible that there is no global maximum or minimum. It is difficult, and not particularly useful, to express a complete procedure for determining whether this is the case. Generally, the best approach is to gain enough understanding of the shape of the graph to decide. Fortunately, only a rough idea of the shape is usually needed.

There are some particularly nice cases that are easy. A continuous function on a closed interval $[a,b]$ *always* has both a global maximum and a global minimum, so examining the critical values and the endpoints is enough:

Theorem 6.1.2 *(Extreme value theorem) If* f *is continuous on a closed interval* $[a,b]$, *then it has both a minimum and a maximum point. That is, there are real numbers* c *and* d *in* $[a,b]$ *so that for every* x *in* $[a,b]$, $f(x)\le f(c)$ *and* $f(x)\ge f(d)$.

Another easy case: If a function is continuous and has a single critical value, then if there is a local maximum at the critical value it is a global maximum, and if it is a local minimum it is a global minimum. There may also be a global minimum in the first case, or a global maximum in the second case, but that will generally require more effort to determine.

Example 6.1.3 Let $f(x)=-x^2+4x-3$. Find the maximum value of $f(x)$ on the interval $[0,4]$. First note that $f'(x)=-2x+4=0$ when $x=2$, and $f(2)=1$. Next observe that $f'(x)$ is defined for all x, so there are no other critical values. Finally, $f(0)=-3$ and $f(4)=-3$. The largest value of $f(x)$ on the interval $[0,4]$ is $f(2)=1$.

Example 6.1.4 Let $f(x)=-x^2+4x-3$. Find the maximum value of $f(x)$ on the interval $[-1,1]$.

First note that $f'(x)=-2x+4=0$ when $x=2$. But $x=2$ is not in the interval, so we don't use it. Thus the only two points to be checked are the endpoints; $f(-1)=-8$ and $f(1)=0$. So the largest value of $f(x)$ on $[-1,1]$ is $f(1)=0$.

Example 6.1.5 Find the maximum and minimum values of the function $f(x)=7+|x-2|$ for x between 1 and 4 inclusive. The derivative $f'(x)$ is never zero, but $f'(x)$ is undefined at $x=2$, so we compute $f(2)=7$. Checking the end points we get $f(1)=8$ and $f(4)=9$. The smallest of these numbers is $f(2)=7$, which is, therefore, the minimum value of $f(x)$ on the interval $1\le x\le 4$, and the maximum is $f(4)=9$.

-2

2

Figure 6.1.2. $f(x)=x^3-x$

Example 6.1.6 Find all local maxima and minima for $f(x)=x^3-x$, and determine whether there is a global maximum or minimum on the open interval $(-2,2)$. In example 5.1.2 we found a local maximum at (−3–√/3,23–

√/9)(−3/3,23/9) and a local minimum at (3–√/3,−23–√/9)(3/3,−23/9). Since the endpoints are not in the interval (−2,2)(−2,2) they cannot be considered. Is the lone local maximum a global maximum? Here we must look more closely at the graph. We know that on the closed interval [−3–√/3,3–√/3][−3/3,3/3] there is a global maximum at x=−3–√/3x=−3/3 and a global minimum at x=3–√/3x=3/3. So the question becomes: what happens between −2−2 and −3–√/3−3/3, and between 3–√/33/3 and 22? Since there is a local minimum at x=3–√/3x=3/3, the graph must continue up to the right, since there are no more critical values. This means no value of ff will be less than −23–√/9−23/9 between 3–√/33/3 and 22, but it says nothing about whether we might find a value larger than the local maximum 23–√/923/9. How can we tell? Since the function increases to the right of 3–√/33/3, we need to know what the function values do "close to" 22. Here the easiest test is to pick a number and do a computation to get some idea of what's going on. Since f(1.9)=4.959>23–√/9f(1.9)=4.959>23/9, there is no global maximum at −3–√/3−3/3, and hence no global maximum at all. (How can we tell that 4.959>23–√/94.959>23/9? We can use a calculator to approximate the right hand side; if it is not even close to 4.9594.959 we can take this as decisive. Since 23–√/9≈0.384923/9≈0.3849, there's really no question. Funny things can happen in the rounding done by computers and calculators, however, so we might be a little more careful, especially if the values come out quite close. In this case we can convert the relation 4.959>23–√/94.959>23/9 into (9/2)4.959>3–√(9/2)4.959>3 and ask whether this is true. Since the left side is clearly larger than 4·44·4 which is clearly larger than 3–√3, this settles the question.)

A similar analysis shows that there is also no global minimum. The graph of f(x)f(x) on (−2,2)(−2,2) is shown in figure 6.1.2.

Example 6.1.7 Of all rectangles of area 100, which has the smallest perimeter?

First we must translate this into a purely mathematical problem in which we want to find the minimum value of a function. If xx denotes one of the sides of the rectangle, then the adjacent side must be 100/x100/x (in order that the area be 100). So the function we want to minimize is

f(x)=2x+2100xf(x)=2x+2100x

since the perimeter is twice the length plus twice the width of the rectangle. Not all values of xx make sense in this problem: lengths of sides of rectangles must be positive, so x>0x>0. If x>0x>0 then so is 100/x100/x, so we need no second condition on xx.

We next find f′(x)f′(x) and set it equal to zero: 0=f′(x)=2−200/x20=f′(x)=2−200/x2. Solving f′(x)=0f′(x)=0 for xx gives us x=±10x=±10. We are interested only in x>0x>0, so only the value x=10x=10 is of interest. Since f′(x)f′(x) is defined everywhere on the interval (0,∞)(0,∞), there are no more critical values, and there are no endpoints. Is there a local maximum, minimum, or neither at x=10x=10? The second derivative is f′′(x)=400/x3f′′(x)=400/x3, and f′′(10)>0f′′(10)>0, so there is a local minimum. Since there is only one critical value, this is also the global minimum, so the rectangle with smallest perimeter is the 10×1010×10 square.

Example 6.1.8 You want to sell a certain number nn of items in order to maximize your profit. Market research tells you that if you set the price at $1.50, you will be able to sell

5000 items, and for every 10 cents you lower the price below \$1.50 you will be able to sell another 1000 items. Suppose that your fixed costs ("start-up costs") total \$2000, and the per item cost of production ("marginal cost") is \$0.50. Find the price to set per item and the number of items sold in order to maximize profit, and also determine the maximum profit you can get.

The first step is to convert the problem into a function maximization problem. Since we want to maximize profit by setting the price per item, we should look for a function $P(x)$ representing the profit when the price per item is x. Profit is revenue minus costs, and revenue is number of items sold times the price per item, so we get $P=nx-2000-0.50n$. The number of items sold is itself a function of x, $n=5000+1000(1.5-x)/0.10$, because $(1.5-x)/0.10$ is the number of multiples of 10 cents that the price is below \$1.50. Now we substitute for n in the profit function:

$$P(x)=(5000+1000(1.5-x)/0.10)x-2000-0.5(5000+1000(1.5-x)/0.10)=-10000x^2+25000x-12000$$

We want to know the maximum value of this function when x is between 0 and 1.5. The derivative is $P'(x)=-20000x+25000$, which is zero when $x=1.25$. Since $P''(x)=-20000<0$, there must be a local maximum at $x=1.25$, and since this is the only critical value it must be a global maximum as well. (Alternately, we could compute $P(0)=-12000$, $P(1.25)=3625$, and $P(1.5)=3000$ and note that $P(1.25)$ is the maximum of these.) Thus the maximum profit is \$3625, attained when we set the price at \$1.25 and sell 7500 items.

0

0.5

-0.5

1

-1

0.5

0

1

$\text{Area}=0.7500$

Figure 6.1.3. *Rectangle in a parabola: drag the blue point, trying to maximize the area.*

Example 6.1.9 Find the largest rectangle (that is, the rectangle with largest area) that fits inside the graph of the parabola $y=x^2$ below the line $y=a$ (a is an unspecified constant value), with the top side of the rectangle on the horizontal line $y=a$; see figure 6.1.3.)

We want to find the maximum value of some function A(x)A(x) representing area.

The Chain Rule

Perhaps the hardest part of this problem is deciding what xx should represent. The lower right corner of the rectangle is at (x,x2)(x,x2), and once this is chosen the rectangle is completely determined. So we can let the xx in A(x)A(x) be the xx of the parabola f(x)=x2f(x)=x2. Then the area is A(x)=(2x)(a−x2)=−2x3+2axA(x)=(2x)(a−x2)=−2x3+2ax. We want the maximum value of A(x)A(x) when xx is in [0,a−−√][0,a]. (You might object to allowing x=0x=0 or x=a−−√x=a, since then the "rectangle" has either no width or no height, so is not "really" a rectangle. But the problem is somewhat easier if we simply allow such rectangles, which have zero area.)

Setting 0=A′(x)=−6x2+2a0=A′(x)=−6x2+2a we get x=a/3−−−√x=a/3 as the only critical value. Testing this and the two endpoints, we have A(0)=A(a−−√)=0A(0)=A(a)=0 and A(a/3−−−√)=(4/9)3–√a3/2A(a/3)=(4/9)3a3/2. The maximum area thus occurs when the rectangle has dimensions 2a/3−−−√×(2/3)a2a/3×(2/3)a.

(h−R,r)(h−R,r)

After reading this text, and/or viewing the video tutorial on this topic, you should be able to:

- explain what is meant by a function of a function
- state the chain rule
- differentiate a function of a function

Contents

5. A simple technique for differentiating directly

1. Introduction

In this unit we learn how to differentiate a 'function of a function'. We first explain what is meant by this term and then learn about the Chain Rule which is the technique used to perform the differentiation.

2. A function of a function

Consider the expression $\cos x^2$. Immediately we note that this is different from the straightforward cosine function, $\cos x$. We are finding the cosine of x^2, not simply the cosine of x. We call such an expression a 'function of a function'.

Suppose, in general, that we have two functions, $f(x)$ and $g(x)$. Then

$$y = f(g(x))$$

is a function of a function. In our case, the function f is the cosine function and the function g is the square function. We could identify them more mathematically by saying that

$$f(x) = \cos x \qquad g(x) = x^2$$

so that $f(g(x)) = f(x^2) = \cos x^2$

Now let's have a look at another example. Suppose this time that f is the square function and g is the cosine function. That is,

$$f(x) = x^2 \qquad g(x) = \cos x$$

then $f(g(x)) = f(\cos x) = (\cos x)^2$

We often write $(\cos x)^2$ as $\cos^2 x$. So $\cos^2 x$ is also a function of a function.

In the following section we learn how to differentiate such a function.

3. The chain rule

In order to differentiate a function of a function, $y = f(g(x))$, that is to find $\frac{dy}{dx}$, we need to do two things:

1. Substitute $u = g(x)$. This gives us

$$y = f(u)$$

Next we need to use a formula that is known as the Chain Rule.

2. Chain Rule

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

Key Point

Chain rule:

To differentiate $y = f(g(x))$, let $u = g(x)$. Then $y = f(u)$ and

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

Example

Suppose we want to differentiate $y = \cos x^2$.

Let $u = x^2$ so that $y = \cos u$.

It follows immediately that

$$\frac{du}{dx} \qquad \frac{dy}{du} -$$

The chain rule says

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

and so

$$\frac{dy}{dx} = -\sin u \times 2x$$
$$= -2x\sin x^2$$

Example

Suppose we want to differentiate $y = \cos^2 x = (\cos x)^2$.

Let $u = \cos x$ so that $y = u^2$

It follows that

$$\frac{du}{dx} = -\sin x \qquad \frac{dy}{du} = 2u$$

Then

$$dx \qquad du \times dx$$

$$= 2u \times -\sin x$$

$$= -2\cos x \sin x$$

d y $\frac{dy}{}$ $\frac{du}{}$

Example

Suppose we wish to differentiate $y = (2x - 5)^{10}$.

Now it might be tempting to say ‘surely we could just multiply out the brackets’. To multiply out the brackets would take a long time and there are lots of opportunities for making mistakes.

So let us treat this as a function of a function.

Let $u = 2x - 5$ so that $y = u^{10}$. It follows that

$$\frac{du}{dx} = 2 \qquad \frac{dy}{du} = 10u^9$$

Then

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

$$= 10u^9 \times 2$$

$$= 20(2x - 5)^9$$

4. Some examples involving trigonometric functions

In this section we consider a trigonometric example and develop it further to a more general case.

Example

Suppose we wish to differentiate $y = \sin 5x$.

Let $u = 5x$ so that $y = \sin u$. Differentiating

$$\frac{du}{dx} = 5 \qquad \frac{dy}{du} = \cos u$$

From the chain rule

$$\frac{dy}{} \quad \frac{dy}{} \quad \frac{du}{}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

$$= \cos u \times 5$$

$$= 5\cos 5x$$

Notice how the 5 has appeared at the front, - and it does so because the derivative of 5x was 5. So the question is, could we do this with any number that appeared in front of the x, be it 5 or 6 or $\frac{1}{2}$, 0.5 or for that matter n ?

So let's have a look at another example.

Example

Suppose we want to differentiate $y = \sin nx$.

Let $u = nx$ so that $y = \sin u$. Differentiating

$$\frac{du}{dx} = n \qquad \frac{dy}{du} = \cos u$$

Quoting the formula again:

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

So

$$\frac{dy}{dx} = \cos u \times n$$

$$= n\cos nx$$

So the n's have behaved in exactly the same way that the 5's behaved in the previous example.

Key Point

if $y = \sin nx$ then $\frac{dy}{dx} = n\cos nx$

For example, suppose $y = \sin 6x$ then $\frac{dy}{dx} = 6\cos 6x$ just by using the standard result.

Similar results follow by differentiating the cosine function:

Key Point

if $y = \cos nx$ then $\dfrac{dy}{dx} = -n\sin nx$

So, for example, if $y = \cos\frac{1}{2}x$ then $\dfrac{dy}{dx} = -\frac{1}{2}\sin\frac{1}{2}x$.

5. A simple technique for differentiating directly

In this section we develop, through examples, a further result.

Example

Suppose we want to differentiate $y = e^{x^3}$. Let $u = x^3$ so that $y = e^u$.

Differentiating

$$\frac{du}{dx} = 3x^2 \qquad \frac{dy}{du} = e^u$$

Quoting the formula again:

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

So

$$\frac{dy}{dx} = e^u \times 3x^2 = 3x^2 e^{x^3}$$

We will now explore how this relates to a general case, that of differentiating $y = f(g(x))$.

To differentiate $y = f(g(x))$, we let $u = g(x)$ so that $y = f(u)$.

The chain rule states

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

In what follows it will be convenient to reverse the order of the terms on the right:

$$\frac{dy}{dx} = \frac{du}{dx} \times \frac{dy}{du}$$

which, in terms of f and g we can write as

$$\frac{dy}{dx} = \frac{d}{dx}(g(x)) \times \frac{d}{du}(f(g((x)))$$

This gives us a simple technique which, with some practice, enables us to apply the chain rule

Directly

Key Point

(i) given $y = f(g(x))$, identify the functions $f(u)$ and $g(x)$ where $u = g(x)$.

(ii) differentiate g and multiply by the derivative of f where it is understood that the argument of f is $u = g(x)$.

Example

To differentiate $y = \tan x^2$ we apply these two stages:

(i) first identify $f(u)$ and $g(x)$: $f(u) = \tan u$ and $g(x) = x^2$.

(ii) differentiate $g(x)$: $\frac{dg}{dx} = 2x$. Multiply by the derivative of $f(u)$, which is $\sec^2 u$ to give

$$\frac{dy}{dx} = 2x\sec^2 x^2$$

Example

To differentiate $y = e^{1+x^2}$.

(i) first identify $f(u)$ and $g(x)$: $f(u) = e^u$ and $g(x) = 1 + x^2$.

(ii) differentiate $g(x)$: $\frac{dg}{dx} = 2x$. Multiply by the derivative of $f(u)$, which is e^u to give

$$\frac{dy}{dx} =$$

$$2xe^{1+x^2}\,dx$$

You should be able to verify the remaining examples purely by inspection. Try it!

Example

$$y = \sin(x + e^x)$$

$$\frac{dy}{dx} = (1 + e^x)\cos(x + e^x)$$

Example

$$y = \tan(x^2 + \sin x)$$

$$\frac{dy}{dx} = (2x + \cos x).\sec^2(x^2 + \sin x)$$

Example

$$y = (2 - x^5)^9$$

$$\frac{dy}{dx} = -5x^4 . 9(2 - x^5)^8$$

$$= -45x^4(2 - x^5)^8$$

Example

$$y = \ln(x + \sin x)$$

$$\frac{dy}{dx} = (1 + \cos x) . \frac{1}{x + \sin x}$$

$$= \frac{1 + \cos x}{x + \sin x}$$

Exercises

1. Find the derivative of each of the following:

a) $(3x - 7)^{12}$ b) $\sin(5x + 2)$ c) $\ln(2x - 1)$ d) e^{2-3x}

$\sqrt{5x-3}$

e)f) (6*x* + 5) $^{5/3}$ g) $\frac{}{(3-1x)^4}$ h) cos(1 − 4*x*)

2. Find the derivative of each of the following:

a) ln(sin*x*) b) sin(ln*x*) c) $e^{-\cos x}$ d) $\cos(e^{-x})$

e) $(\sin x + \cos x)^3$ f) $\sqrt{1+x^2}$ g) $\frac{1}{\cos x}$ h) $\frac{1}{(x^2 + 2x + 1)^2}$

3. Find the derivative of each of the following:

a) $\ln\left(\sin^2 x\right)$ b) $\sin^2(\ln x)$ c) $\sqrt{\cos(3x-1)}$ d) $[1 + \cos(x^2 - 1)]^{3/2}$

Answers

1. a) $36(3x-7)^{11}$ b) $5\cos(5x+2)$ c) $2x\,2^{x^2-1}$ d) $-3e^{2-3x}$

e) $\frac{5}{2\sqrt{5x-3}}$ f) $10(6x+5)^{2/3}$ g) $\frac{4}{(3-x)^5}$ h) 4sin(1 − 4*x*)

2. a) $\frac{\cos x}{\sin x} = \cot x$ b) $\frac{\cos(\ln x)}{x}$ c) $\sin x\, e^{-\cos x}$

d) $e^{-x}\sin(e^{-x})$ e) $3(\cos x - \sin x)(\sin x + \cos x)^2$ f) $\frac{x}{\sqrt{1+x^2}}$

g) $\frac{\sin x}{\cos^2 x} = \tan x \sec x$ h) $\frac{-2(x+1)}{(x^2+2x+1)^4} = \frac{-2}{(x+1)^3}$

3. a) $\frac{2\cos x}{\sin x} = 2\cot x$ b) $\frac{2\sin(\ln x)\cos(\ln x)}{x}$

c) $\frac{-3\sin(3x-1)}{2\sqrt{\cos(3x-1)}}$ d) $-3x\sin(x^2-1)[1+\cos(x^2-1)]^{1/2}$

3.5 Differentiation of trigonometric and inverse trigonometric functions

- Differentiate:
  I. Trigonometric functions(sinx, cosx, tanx, cosecx, secx , cotx) from first principles.
- Inverse trigonometric functions (arcsinx, arccosx, arctanx, arccosecx, arcsecx ,

arccotx) using differentiation formulae.

Differentiation of Inverse Trigonometric Functions

Each of the six basic trigonometric functions have corresponding inverse functions when appropriate restrictions are placed on the domain of the original functions. All the inverse trigonometric functions have derivatives, which are summarized as follows:

(1) If $f(x) = \sin^{-1} x = \arcsin x, -\frac{\pi}{2} \le f(x) \le \frac{\pi}{2}$ then

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

(2) If $f(x) = \cos^{-1} x = \arccos x, 0 \le f(x) \le \pi$ then

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

(3) If $f(x) = \tan^{-1} x = \arctan x, -\frac{\pi}{2} < f(x) < \frac{\pi}{2}$ then

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

(4) If $f(x) = \cot^{-1} x = \text{arccot}\, x, 0 < f(x) < \pi$ then

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

(5) If $f(x) = \sec^{-1} x = \text{arcsec}\, x, 0 \le f(x) \le \pi, f(x) \ne \frac{\pi}{2}$ then

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

(6) If $f(x) = \csc^{-1} x = \text{arccsc}\, x, -\frac{\pi}{2} \le f(x) \le \frac{\pi}{2}, f(x) \ne 0$

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

Example 1: Find $f'(x)$ if $f(x) = \cos^{-1}(5x)$.

$$f'(x) = \frac{-1}{\sqrt{1-(5x)^2}} \cdot 5$$

$$f'(x) = \frac{-5}{\sqrt{1-25x^2}}$$

Example 2: Find y' if $y = \arctan\left(\sqrt{x^3}\right)$.

Because $y = \arctan(x^{3/2})$

$$y' = \frac{1}{1 + (x^{3/2})^2} \cdot \frac{3}{2} x^{1/2}$$

$$= \frac{1}{1 + x^3} \cdot \frac{3}{2} x^{1/2}$$

$$y' = \frac{3\sqrt{x}}{2(1 + x^3)}$$

3.6 Differentiation of Exponential and logarithm functions

- Find the derivative of:
  I. e^x and a^x from first principles.
  II. lnx and $\log_a x$ from first principles.
- Algebraic expression involving product, quotient and power.

3. The Derivative from First Principles

In this section, we will differentiate a function from "first principles". This means we will start from scratch and use algebra to find a general expression for the slope of a curve, at any value *x*.

First principles is also known as "delta method", since many texts use Δ*x* (for "change in *x*) and Δ*y* (for "change in *y*"). This makes the algebra appear more difficult, so here we use *h* for Δ*x* instead. We still call it "delta method".

NOTE

If you want to see how to find slopes (gradients) of tangents directly using derivatives, rather than from first principles, go to Tangents and Normals in the Applications of Differentiation chapter.

Py = f(x)msOpen image in a new page

Slope of the tangent at *P*.

We wish to find an **algebraic method** to find the slope of $y = f(x)$ at *P*, to save doing the numerical substitutions that we saw in the last section (Slope of a Tangent to a Curve - Numerical Approach).

We can approximate this value by taking a point somewhere near to *P*(*x*, *f*(*x*)), say *Q*(*x* + *h*, *f*(*x* + *h*)).

PghQ (x+h, f(x+h))Open image in a new page

Slope of the line *PQ*.

The value \displaystyle\frac{g}{{h}}*hg* is an approximation to the slope of the tangent which we require.

We can also write this slope as \displaystyle\frac{{\text{change in}\ {y}}}{{\text{change in}\ {x}}}change in *x*change in *y* or:

\displaystyle{m}=\frac{{\Delta{y}}}{{\Delta{x}}}*m*=Δ*x*Δ*y*

If we move *Q* closer and closer to *P* (that is, we let *h* get smaller and smaller), the line *PQ* will get closer and closer to the tangent at *P* and so the slope of *PQ* gets closer to the slope that we want.

PQOpen image in a new page

PQOpen image in a new page

Slope of the line *PQ*.

If we let *Q* go all the way to touch *P* (i.e. \displaystyle{h}={0}*h*=0), then we would have the **exact** slope of the tangent.

Differentiation from first principles applet

In the following applet, you can explore how this process works.

We are using the example from the previous page (Slope of a Tangent), $y = x^2$, and finding the slope at the point P(2, 4).

Use the **left-hand slider** to move the point P closer to Q. Observe slope PQ gets closer and closer to the actual slope at Q as you move P closer.

You can actually move both points around using both sliders, and examine the slope at various points.

What is the slope at point (0, 0)?

xy246-21020304050PQ

Derivatives (slope)

The function: \displaystyle{y}={x}^{2}$y=x2$

Gradient at *P*(2.00, 4.00) is 4.00

Gradient of line *PQ* is 6.00

Gradient at *Q*(4.00, 16.00) is 8.00

Expressing the differentiation process using algebra

Now, \displaystyle\frac{g}{{h}}hg can be written:

\displaystyle\frac{g}{{h}}=\frac{{ f{{\left({x}+{h}\right)}}- f{{\left({x}\right)}}}}{{h}}$hg=hf(x+h)-f(x)$

So also, the slope *PQ* will be given by:

\displaystyle{m}=\frac{{{y}_{{2}}-{y}_{{1}}}}{{{x}_{{2}}-{x}_{{1}}}}=\frac{{\Delta{y}}}{{\Delta{x}}}$m=x2-x1y2-y1=\Delta x\Delta y$
\displaystyle=\frac{{ f{{\left({x}+{h}\right)}}- f{{\left({x}\right)}}}}{{h}}$=hf(x+h)-f(x)$

But we require the slope **at** *P*, so we let \displaystyle{h}\rightarrow{0}$h\to 0$ (that is let *h* approach \displaystyle{0}0), then in effect, *Q* will approach *P* and \displaystyle\frac{g}{{h}}hg will approach the required slope.

Continues below ⇩

The Slope of a Curve as a Derivative

Putting this together, we can write the slope of the tangent at *P* as:

\displaystyle\frac{{\left.{d}{y}\right.}}{{\left.{d}{x}\right.}}=\lim_{{{h}\to{0}}}\frac{{ f{{\left({x}+{h}\right)}}- f{{\left({x}\right)}}}}{{h}}$dxdy=h\to 0\lim hf(x+h)-f(x)$

This is called **differentiation from first principles,** (or the **delta method**). It gives the instantaneous rate of change of *y* with respect to *x*.

This is equivalent to the following (where before we were using *h* for Δx):

\displaystyle\frac{{\left.{d}{y}\right.}}{{\left.{d}{x}\right.}}=\lim_{{\Delta{x}\to{0}}}\frac{{\Delta{y}}}{{\Delta{x}}}$dxdy=\Delta x\to 0\lim\Delta x\Delta y$

You will also come across the following way of writing the Delta Method:

\displaystyle\frac{{\left.{d}{y}\right.}}{{\left.{d}{x}\right.}}=\lim_{{\Delta{x}\to{0}}}\frac{{ f{{\left({x}+\Delta{x}\right)}}- f{{\left({x}\right)}}}}{{\Delta{x}}}*dxdy* =Δ*x*→0limΔ*xf*(*x*+Δ*x*)−*f*(*x*)

Notation for the Derivative

IMPORTANT: The **derivative** (also called **differentiation**) can be written in several ways. This can cause some confusion when we first learn about differentiation.

The following are equivalent ways of writing the first derivative of \displaystyle{y}= f{{\left({x}\right)}}*y*=*f*(*x*):

\displaystyle\frac{{\left.{d}{y}\right.}}{{\left.{d}{x}\right.}}*dxdy*
or \displaystyle{f}\text{'}{\left({x}\right)}*f*'(*x*) or \displaystyle{y}\text{'}*y*'.

Example 1

Find \displaystyle\frac{{\left.{d}{y}\right.}}{{\left.{d}{x}\right.}}*dxdy* from first principles if $y = 2x^2 + 3x$.

3.7 Differentiation of hyperbolic and inverse hyperbolic functions Differentiate:

I. Hyperbolic functions (sinhx, coshx, tanhx, cosechx, sechx , cothx).
II. Inverse Hyperbolic functions (arcsinhx, arccoshx, arctanhx, arccosechx, arcsechx , arccothx).

- Apply MAPLE command to differentiate a function.
- **Derivation of the**
  - **Inverse Hyperbolic Trig Functions**
  - $y = \sinh^{-1} x$.
- By definition of an inverse function, we want a function that satisfies the condition
  - $$\begin{aligned} x &= \sinh y \\ &= \frac{e^y - e^{-y}}{2} \end{aligned}$$ by definition of sinh*y*

$$\begin{aligned} &= \frac{e^y - e^{-y}}{2}\left(\frac{e^y}{e^y}\right)\\ &= \frac{e^{2y}-1}{2e^y}.\\ 2e^y x &= e^{2y}-1.\\ e^{2y}-2xe^y-1 &= 0.\\ (e^y)^2 - 2x(e^y)-1 &= 0.\\ e^y &= \frac{2x+\sqrt{4x^2+4}}{2}\\ &= x+\sqrt{x^2+1}.\\ \ln(e^y) &= \ln(x+\sqrt{x^2+1})\\ y &= \ln(x+\sqrt{x^2+1}).\end{aligned}$$

- Thus

$$\sinh^{-1} x = \ln(x+\sqrt{x^2+1})$$

- .

$$f(x) = \sinh^{-1} x.$$

- Next we compute the derivative of

$$\begin{aligned} f'(x) &= \frac{1}{x+\sqrt{x^2+1}}\left(1+\frac{1}{2}(x^2+1)^{-1/2}(2x)\right)\\ &= \frac{1}{\sqrt{x^2+1}}.\end{aligned}$$

- = $\cosh^{-1} x$.

$$\begin{aligned} x &= \cosh y\\ &= \frac{e^y+e^{-y}}{2}\end{aligned}$$

by definition of $\cosh y$

$$\begin{aligned} &= \frac{e^y + e^{-y}}{2}\left(\frac{e^y}{e^y}\right)\\ &= \frac{e^{2y}+1}{2e^y}.\\ 2e^y x &= e^{2y}+1.\\ e^{2y}-2xe^y+1 &= 0.\\ (e^y)^2 - 2x(e^y)+1 &= 0.\\ e^y &= \frac{2x+\sqrt{4x^2-4}}{2}\\ &= x+\sqrt{x^2-1}.\\ \ln(e^y) &= \ln(x+\sqrt{x^2-1}).\\ y &= \ln(x+\sqrt{x^2-1}).\end{aligned}$$

- Thus
- – .

$$\cosh^{-1} x = \ln(x+\sqrt{x^2 \quad 1})$$

- Next we compute the derivative of $f(x) = \cosh^{-1} x$.

$$\begin{aligned} f'(x) &= \frac{1}{x+\sqrt{x^2-1}}\left(1+\frac{1}{2}(x^2-1)^{-1/2}(2x)\right)\\ &= \frac{1}{\sqrt{x^2-1}}.\end{aligned}$$

- = $\tanh^{-1} x$.

$$\begin{aligned} x &= \tanh y\\ &= \frac{e^y-e^{-y}}{e^y+e^{-y}}\end{aligned}$$

by definition of $\tanh y$

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

$$\begin{aligned}
&= \frac{e^y - e^{-y}}{e^y + e^{-y}}\left(\frac{e^y}{e^y}\right)\\
&= \frac{e^{2y}-1}{e^{2y}+1}.\\
x(e^{2y}+1) &= e^{2y}-1.\\
(x-1)e^{2y}+(x+1) &= 0.\\
e^{2y} &= -\frac{x+1}{x-1}.\\
\ln(e^{2y}) &= \ln\left(-\frac{x+1}{x-1}\right).\\
2y &= \ln\left(-\frac{x+1}{x-1}\right).\\
y &= \frac{1}{2}\ln\left(-\frac{x+1}{x-1}\right)\\
&= \frac{1}{2}(\ln(x+1)-\ln(-[x-1]))\\
&= \frac{1}{2}(\ln(x+1)-\ln(1-x)).
\end{aligned}$$

-
- Thus
  - . $\tanh^{-1} x = \frac{1}{2}(\ln(x+1)-\ln(1-x))$

$f(x) = \tanh^{-1} x.$

- Next we compute the derivative of
  - = sech–¹x.

$$\begin{aligned}
x &= \operatorname{sech} y\\
&= \frac{2}{e^y + e^{-y}}
\end{aligned}$$

- by definition of sech*y*

$$\begin{aligned}
&= \frac{2}{e^y+e^{-y}}\left(\frac{e^y}{e^y}\right)\\
&= \frac{2e^y}{e^{2y}+1}.\\
x(e^{2y}+1) &= 2e^y.\\
xe^{2y}-2e^y+x &= 0.\\
e^y &= \frac{-(-2)+\sqrt{(-2)^2-4(x)(x)}}{2x}\\
&= \frac{2+\sqrt{4(1-x^2)}}{2x}\\
&= \frac{2+2\sqrt{1-x^2}}{2x}\\
&= \frac{1+\sqrt{1-x^2}}{x}.\\
y &= \ln\left(\frac{1+\sqrt{1-x^2}}{x}\right)\\
&= \ln(1+\sqrt{1-x^2})-\ln x.
\end{aligned}$$

- Thus
- – –
- Next we compute the derivative of *f*(*x*) = sech–¹x.

$$\begin{aligned}
f'(x) &= \frac{1}{1+\sqrt{1-x^2}}\left(\frac{1}{2}(1-x^2)^{-1/2}(-2x)\right) - \frac{1}{x}\\
&= -\frac{1}{x\sqrt{1-x^2}}.
\end{aligned}$$

- $\operatorname{sech}^{-1}x = \ln(1+\sqrt{1 \quad x^2}) \quad \ln x.$

UNIT 4 HIGHER ORDER DEVIATION AND APPLICATIONS

4.1 Higher order deviation

- Find higher order deviation of algebraic, trigonometric, exponential and logarithmic function.
- Find the 2nd derivative of implicit, inverse trigonometric and parametric function.
- Apply MAPLE command diff repeatedly to find higher order derivative of a function.

ImplicitDifferentiationandtheSecondDerivative

Calculate*y*usingimplicitdifferentiation;simplifyasmuchaspossible.

$$x^2 + 4y^2 = 1$$

Solution

Aswiththedirectmethod,wecalculatethesecondderivativebydifferentiatingtwice.Withimplicitdifferentiationthisleavesuswithaformulafor*y*thatinvolvesyandy,andsimplifyingisaseriousconsideration.

Recallthattotakethederivativeof4y^2withrespectto*x*wefirsttakethederivativewithrespectto*y*andthenmultiplyby*y*;thisisthe"derivativeoftheinsidefunction"mentionedinthechainrule,whilethederivativeoftheoutsidefunctionis8*y*.

So,differentiatingbothsidesof:

$$x^2 + 4y^2 = 1$$

givesus:

$$2x + 8yy = 0.$$

We'renowfacedwithachoice.Wecouldimmediatelyperformimplicitdifferentiationagain,orwecouldsolvefor*y*anddifferentiateagain.

Ifwedifferentiateagainweget:

$$2 + 8yy + 8(y)^2 = 0.$$

Inordertosolvethisfor*y*wewillneedtosolvetheearlierequationfor*y*,soitseemsmostefficienttosolvefor*y*beforetakingasecondderivative.

$$\begin{aligned} 2x + 8yy' &= 0 \\ 8yy' &= -2x \\ y' &= \frac{-2x}{8y} \\ y' &= \frac{-x}{4y} \end{aligned}$$

Differentiatingbothsidesofthisexpression(usingthequotientruleandimpli citdifferentiation),weget:

$$\begin{aligned} y'' &= \frac{(-1)4y - (-x)\cdot 4y'}{(4y)^2} \\ &= \frac{-4y + 4xy'}{16y^2} \\ y'' &= \frac{-y + xy'}{4y^2} \end{aligned}$$

1

Wenowsubstitute $\frac{-x}{4y}$ for$y$:

$$\begin{aligned} y'' &= \frac{-y + xy'}{4y^2} \\ &= \frac{-y + x\frac{-x}{4y}}{4y^2} \\ &= \frac{x\frac{-x}{4y} - y}{4y^2}\cdot\frac{4y}{4y} \\ &= \frac{-x^2 - 4y^2}{16y^3} \\ y'' &= -\frac{1}{16y^3} \end{aligned}$$

(Don'tforgettousetherelation$x^2 + 4y^2 = 1$attheend!)

Howcanwecheckourwork?Ifwerecognize$x^2 + 4y^2 = 1$astheequationofanellipse,wecantestourequation$y = -x/4y$atthepoints(0,1/2)and(1,0).At(0,1/2),$y = -x/4y = 0$whichagreeswiththefactthatthetangentlinetotheellipseishorizontalatthatpoint.At(1,0)$y$isundefined,whichagreeswiththefactthatthetangentlinetotheellipseat(1,0)isvertical.

Oncewehavelearnedhowthevalueofthesecondderivativeisrelatedtotheshapeofthegraph,wecandoasimilartestofourexpressionfor$y$.

4.2 Maclaurin's and Taylor's Expansions

State Maclaurin's and Taylor's theorem (without remainder terms).

- Use the theorems to expand sinx, cos x , tan x, a^x, e^x, $\log_a(1+x)$ and ln(1+x).
- Apply MAPLE command to find Taylor's expansion for a given function.

Taylor and Maclaurin Series

□ If a functionf(x) has continuous derivatives up to (n+1)th order, then this function can be expanded in the following way:

$f(x)=\sum_{n=0}^{\infty} \frac{f^{(n)}(a)(x-a)^n}{n!}=f(a)+f'(a)(x-a)+\frac{f''(a)(x-a)^2}{2!}+\dots+\frac{f^{(n)}(a)(x-a)^n}{n!}+R_n$,

where R_n, called the remainder after n+1 terms, is given by

$R_n=\frac{f^{(n+1)}(\xi)(x-a)^{n+1}}{(n+1)!}, a<\xi<x$.

When this expansion converges over a certain range of x, that is, $\lim_{n\to\infty} R_n=0$, then the expansion is called Taylor Series of f(x) expanded about a.

If a=0, the series is called Maclaurin Series:

$f(x)=\sum_{n=0}^{\infty} \frac{f^{(n)}(0)x^n}{n!}=f(0)+f'(0)x+\frac{f''(0)x^2}{2!}+\dots+\frac{f^{(n)}(0)x^n}{n!}+R_n$.

Some Useful Maclaurin Series

$e^x=\sum_{n=0}^{\infty} \frac{x^n}{n!}=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\dots$

$\cos x=\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\dots$

$\sin x=\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\dots$

$\cosh x=\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}=1+\frac{x^2}{2!}+\frac{x^4}{4!}+\frac{x^6}{6!}+\dots$

$\sinh x=\sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}=x+\frac{x^3}{3!}+\frac{x^5}{5!}+\frac{x^7}{7!}+\dots$

▯ Solved Problems

Click a problem to see the solution.

Example 1

Find the Maclaurin series for cos2x.

Example 2

Obtain the Taylor series for f(x) =3x2−6x+5 about the point x=1.

Example 3

Find the Maclaurin series for ekx, k is a real number.

Example 4

Find the Taylor series of the cubic function x3 about x=3.

Example 5

Find the Maclaurin series for (1+x)μ.

Example 6

Determine the Maclaurin series for f(x)=√1+x.

Example 1.

Find the Maclaurin series for cos2x.

Solution.
We use the trigonometric identity cos2x =1+cos2x2.
As the Maclaurin series for cosx is ∞∑n=0(−1)nx2n(2n)!, we can write:
cos2x=∞∑n=0(−1)n(2x)2n(2n)!=∞∑n=0(−1)n22nx2n(2n)!.
Therefore

1+cos2x=1+∞∑n=0(−1)n22nx2n(2n)!=2+∞∑n=1(−1)n22nx2n(2n)!,
cos2x=1+cos2x2=1+∞∑n=1(−1)n22n−1x2n(2n)!.

Example 2.

Obtain the Taylor series for f(x) =3x2−6x+5 about the point x=1.

Solution.
Compute the derivatives:

f′(x)=6x−6,f′′(x)=6,f′′′(x)=0.
As you can see, f(n)(x)=0 for all n≥3. Then for x=1, we get
f(1)=2,f′(1)=0,f′′(1)=6.
Hence, the Taylor expansion for the given function is

f(x)=∞∑n=0f(n)(1)(x−1)nn!=2+6(x−1)22!=2+3(x−1)2.

Taylor and Maclaurin Series – Page 2

Example 3.

Find the Maclaurin series for e^{kx}, k is a real number.

Solution.
Calculate the derivatives:

$$f'(x)=(e^{kx})'=ke^{kx},\quad f''(x)=(ke^{kx})'=k^2e^{kx},\ \ldots\ f^{(n)}(x)=k^ne^{kx}.$$

Then, at $x=0$ we have

$$f(0)=e^0=1,\quad f'(0)=ke^0=k,\quad f''(0)=k^2e^0=k^2,\ \ldots\ f^{(n)}(0)=k^ne^0=k^n.$$

Hence, the Maclaurin expansion for the given function is

$$e^{kx}=\sum_{n=0}^{\infty}\frac{f^{(n)}(0)x^n}{n!}=1+kx+\frac{k^2x^2}{2!}+\frac{k^3x^3}{3!}+\ldots=\sum_{n=0}^{\infty}\frac{k^nx^n}{n!}.$$

Example 4.

Find the Taylor series of the cubic function x^3 about $x=3$.

Solution.
We denote $f(x)=x^3$. Then

$$f'(x)=(x^3)'=3x^2,\quad f''(x)=(3x^2)'=6x,\quad f'''(x)=(6x)'=6,\quad f^{IV}(x)=0,$$

and further $f^{(n)}(x)=0$ for all $x\geq4$.
Respectively, at the point $x=2$, we have

$$f(2)=8,\quad f'(2)=12,\quad f''(2)=12,\quad f'''(2)=6.$$

Hence, the Taylor series expansion for the cubic function is given by the expression

$$x^3=\sum_{n=0}^{\infty}\frac{f^{(n)}(2)(x-2)^n}{n!}=8+12(x-2)+\frac{12(x-2)^2}{2!}+\frac{6(x-2)^3}{3!}=8+12(x-2)+6(x-2)^2+(x-2)^3.$$

Example 5.

Find the Maclaurin series for $(1+x)^\mu$.

Solution.
Let $f(x)=(1+x)^\mu$, where μ is a real number and $x\neq-1$. Then we can write the derivatives as follows

$$f'(x)=\mu(1+x)^{\mu-1},$$
$$f''(x)=\mu(\mu-1)(1+x)^{\mu-2},$$
$$f'''(x)=\mu(\mu-1)(\mu-2)\cdot(1+x)^{\mu-3},$$
$$f^{(n)}(x)=\mu(\mu-1)(\mu-2)\cdots(\mu-n+1)(1+x)^{\mu-n}.$$

For $x=0$, we obtain

$$f(0)=1,\quad f'(0)=\mu,\quad f''(0)=\mu(\mu-1),\ \ldots\ f^{(n)}(0)=\mu(\mu-1)\cdots(\mu-n+1).$$

Hence, the series expansion can be written in the form

$$(1+x)^\mu=1+\mu x+\frac{\mu(\mu-1)}{2!}x^2+\frac{\mu(\mu-1)(\mu-2)}{3!}x^2+\ldots+\frac{\mu(\mu-1)\cdots(\mu-n+1)}{n!}x^n+\ldots$$

This series is called the binomial series.

Example 6.

Determine the Maclaurin series for $f(x)=\sqrt{1+x}$.

Solution.
Using the binomial series found in the previous example and substituting $\mu=\frac{1}{2}$, we get

√1+x=(1+x)12=1+x2+12(12−1)2!x2+12(12−1)(12−2)3!x3+…=1+x2−1⋅x2222!+1⋅3⋅x3233!−1⋅3⋅5⋅x3244!+…+(−1)n+1⋅1⋅3⋅5⋅⋅⋅(2n−3)xn2nn!.
Keeping only the first three terms, we can write this series as

√1+x≈1+x2−x28.

4.3 Application of Derivatives Discuss geometrical interpretation of derivative.

- Find the equation of tangent and normal to the curve at a given point.
- Find the angle of intersection of the two curves.
- Find the point on a curve where the tangent is Parallel to the given line.

Geometric interpretation of the derivative

Geometrically, the derivative of a function at a given point is the slope of the tangent to at the point .

(See the figure to understand it).

The straight line forms a certain angle that we call .

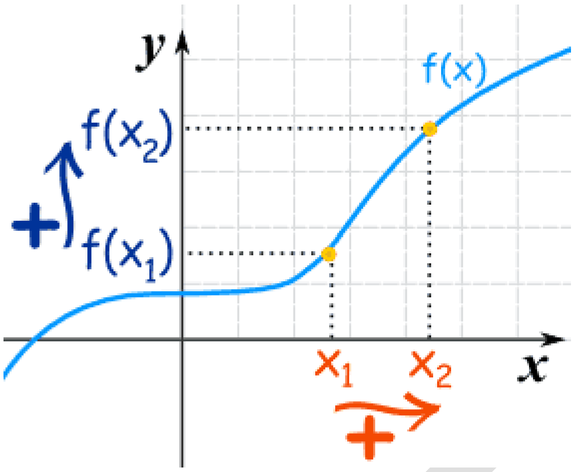
Obviously, this angle will be related to the slope of the straight line, which we have said to be the value of the derivative at the given point.

So we have

\tan\beta = f'(a)

Increasing Functions

A function is "increasing" when the **y-value** increases as the **x-value** increases, like this:



It is easy to see that **y=f(x)** tends to go **up** as it goes **along**.

Flat?

What about that flat bit near the start? Is that OK?

es, it is OK when we say the function is **Increasing**

ut it is **not OK** if we say the function is **Strictly Increasing** (no flatness allowed)

Using Algebra

What if we can't plot the graph to see if it is increasing? In that case we need a definition using algebra.

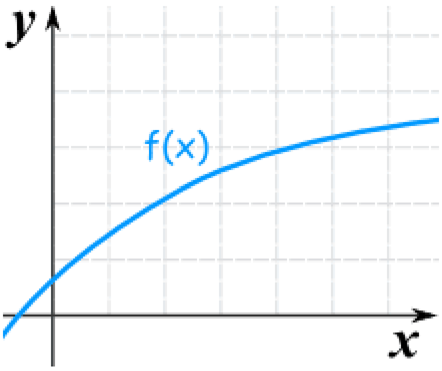
For a function **y=f(x)**:

when $x_1 < x_2$ then $f(x_1) \leq f(x_2)$		**Increasing**
when $x_1 < x_2$ then $f(x_1) < f(x_2)$		**Strictly Increasing**

That has to be true for **any** x_1, x_2, not just some nice ones we might choose.

The important parts are the < and ≤ signs ... remember where they go!

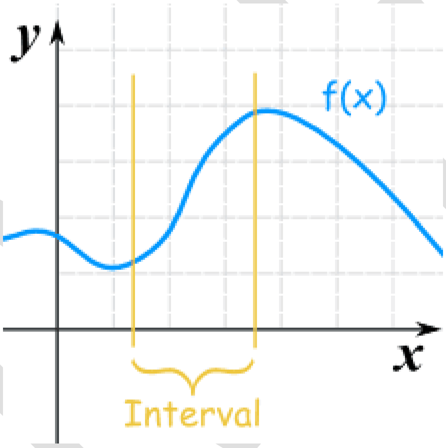
An Example:



This is also an increasing function
even though the rate of increase reduces

For An Interval

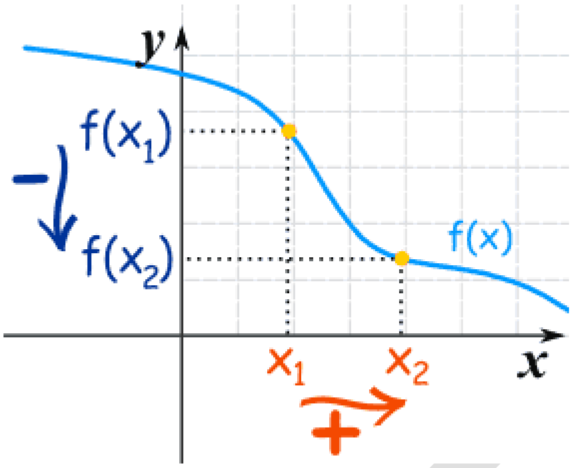
Usually we are only interested in **some interval**, like this one:



This function is **increasing** for the interval shown
(it may be increasing or decreasing elsewhere)

Decreasing Functions

The **y-value** decreases as the **x-value** increases:



For a function **y=f(x)**:

when $x_1 < x_2$ then $f(x_1) \geq f(x_2)$	**Decreasing**
when $x_1 < x_2$ then $f(x_1) > f(x_2)$	**Strictly Decreasing**

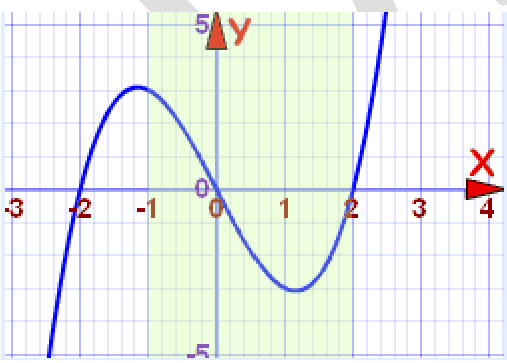
Notice that $f(x_1)$ is now larger than (or equal to) $f(x_2)$.

An Example

Let us try to find where a function is increasing or decreasing.

ample: $f(x) = x^3 - 4x$, for x in the interval $[-1,2]$

Let us plot it, including the interval $[-1,2]$:



Starting from -1 (the beginning of the interval **[−1,2]**):

t $x = \mathbf{-1}$ the function is decreasing,

continues to decrease until **about 1.2** then increases from there, past x = **2**

Without exact analysis we cannot pinpoint where the curve turns from decreasing to increasing, so let us just say:

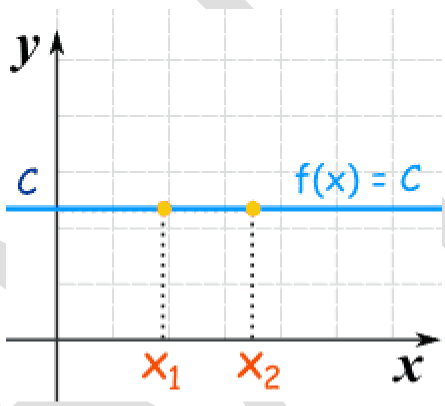
Within the interval **[−1,2]**:

he curve decreases in the interval **[−1, approx 1.2]**

he curve increases in the interval **[approx 1.2, 2]**

Constant Functions

A Constant Function is a horizontal line:



Lines

In fact lines are either increasing, decreasing, or constant.

The equation of a line is:

y = mx + b

The slope m tells us if the function is increasing, decreasing or constant:

m < 0 decreasing

m = 0 constant

m > 0 increasing

One-to-One

Strictly Increasing (and Strictly Decreasing) functions have a special property called "injective" or "one-to-one" which simply means we never get the same "y" value twice.

General Function

"Injective" (one-to-one)

Why is this useful? Because Injective Functions can be **reversed**!

We can go from a "y" value **back to** an "x" value (which we can't do when there is more than one possible "x" value).

4.4 Maximum and Minima Find out increasing and decreasing functions.

- Show that if f (x) is a differentiable function on the open interval (a,b) then:

    I. $f(x)$ is increasing on (a,b) if
        a. $f'(x) > 0, \forall x \in (a,b)$,
    II. $f(x)$ is decreasing on (a,b) if
    III. $f'(x) < 0, \forall x \in (a,b)$
- Evaluate a given function for extreme values.
- State the second derivative rule to find the extreme values of a function at a point.
- Use second derivative rule to examine a given function for extreme values.
- Solve real life problems related to extreme values.
- Apply MAPLE command maximize (minimize) to compute maximum (minimum) value of a function.

1 Maximum-minimum problems

- Tests for absolute extrema
- Problem-solving strategy
- Examples

1.1 Tests for absolute extrema

We stated the extreme value theorem which says that every continuous function on a closed interval has an absoute maximum and absolute minimum. Based on this, we gave a procedure for finding the extreme values of continuous function on a closed interval.

There are two more useful tests.

Theorem 1 *(The first derivative test for absolute maximum.) Suppose that f is defined and continuous on an interval containing c. If* $f^0(x) > 0$ *if* $x > c$ *and* $f^0(x) < 0$ *if* $x < c$*, then c is an absolute minimum.*

Exercise. Restate this test to give a test for an absolute minimum.

A closely related test is the second derivative test. This test does not appear in the text. It states that if a function is concave up, then a critical number is an absolute minimum. More precisely, we have:

Theorem 2 *(Second derivative test for absolute extreme values.) Let f be a twice differentiable function defined on an interval and suppose for some c in the interval,* $f^0(c) = 0$ *and that* $f^{00}(x) > 0$ *for all x in the interval, then f has an absolute minimum at c.*

Proof. To prove this observe that since f^{00} is positive, then f^0 is increasing. Since $f^0(c) = 0$, we can conclude that $f^0(x) > 0$ for $x > c$ and that $f^0(x) < 0$ for $x < c$. Thus, c is a local minimum by the first derivative test.∎

1.2 Strategy

In the examples below, we will follow the following rough guidelines.

1. Read the problem carefully, identify the quantity that we want to make as largeor small as possible. This quantity is called the “objective function”.
2. Draw a diagram and introduce variables for all quantities from the problem.
3. Write an expression for the objective function. This expression may involvemore than one variable.
4. Write relationships among the quantities in the problem and use them to eliminate extra variables from our objective funciton.
5. Write clearly the function (of one variable) to be optimized and state the domain. (The domain for a particular problem may be smaller than the natural domain where the function is defined.)
6. Find the extreme value of the objective funciton using one of the tests above.Explain why you know you have found the maximum.
7. Answer the question. (Are you to give where the maximum occurs, or the extreme value?)

1.3 Examples

Example. Suppose the product of two positive numbers is 5. What is the smallest possible value for the product? The smallest possible value?

Solution. (Minimum value) We let a and b be the two numbers. We are told that a and b satisfy the equation $ab = 5$. Our goal is to maximize the objective function $a + b$.

We solve $ab = 5$ to express b in terms of a (Does it matter if we solve for b or a?) which gives $b = 5/a$. Substituting for b, we obtain the objective function:

$$f(a) = a + \frac{5}{a}$$

and we want to find the smallest value of f for a in $(0,\infty)$. If we compute the derivatives we obtain:

$$f'(a) = 1 - \frac{5}{a^2} \qquad f''(a) = \frac{10}{a^3}.$$

Thus $a = \sqrt{5}$ satisfies $f'(\sqrt{5}) = 0$ and $f''(x) > 0$ for all $x > 0$. We may use the second derivative test for absolute minimum to conclude that $a = \sqrt{5}$ is an absolute minimum on $(0,\infty)$.

The question asks for the minimum value of the sum. This is $f(\sqrt{5}) = 2\sqrt{5}$.

(Maximum value) A sketch the graph of f will indicate that f does not have a maximum value. In fact,

$$\lim_{a\to 0^+} f(a) = \lim_{a\to\infty} f(a) = +\infty.$$

Thus, the sum can be arbitrarily large and does not attain a maximum value.∎

Example. Find the point on the line $x + 3y = 2$ which is closest to $(2,3)$.

Solution. We let (x,y) be an arbitrary point in the plane and then use the distance formula to write down the distance between (x,y) and $(2,3)$.

$$d = \sqrt{(x-2)^2 + (y-3)^2}.$$

If the point (x,y) is to lie on the line, then we have the relation between the coordinates x and y: $x + 3y = 2$. We can solve this equation to give $x = 2 - 3y$ and

substituting for x in d, gives

$$d = \sqrt{((2-3y)-2)^2 + (y-3)^2}.$$

A trick that will simplify the calculations below is to realize that the minimum value of d and the minimum value of d^2 occur at the same point. Thus, our goal is find the minimum value of

$$\begin{aligned} f(y) = d^2 &= ((2-3y)-2)^2 + (y-3)^2 \\ &= 9y^2 + y^2 - 6y + 9 \\ &= 10y^2 - 6y + 9 \end{aligned}$$

for y in $(-\infty,\infty)$.

Computing the derivatives, $f^0(y) = 20y-6$ and $f^{00}(y) = 20$. Thus, we can use the second derivative test for absolute extreme values to conclude that the minimum occurs when $y = 3/10$. Since $x = 2-3y$, the nearest point will be $(x,y) = (11/10,3/10)$.

∎

Remark. It is overkill to use calculus to find the minimum value of a parabola. A method that uses an appropriate level of force is to complete the square:

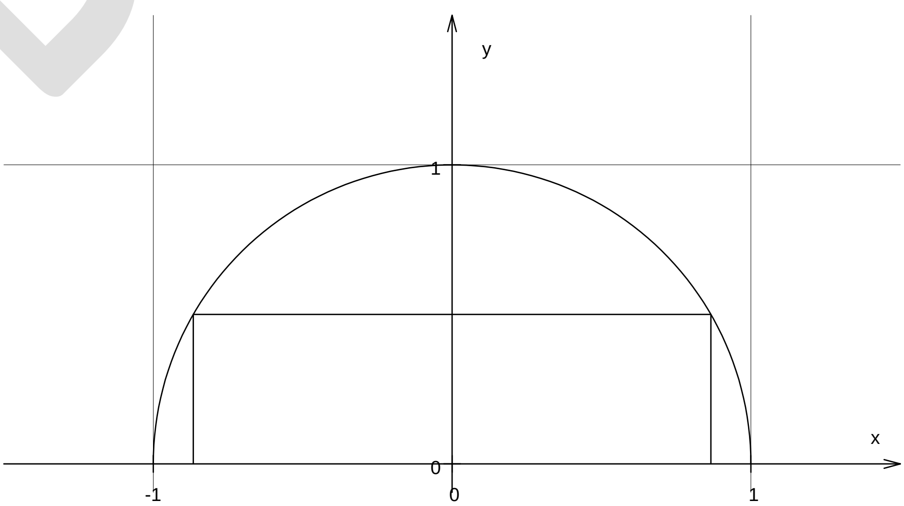
$$\begin{aligned} 10y^2 - 6y + 9 &= 10(y^2 - \frac{6}{10}y + \frac{9}{25}) - \frac{9 \cdot 10}{25} \\ &= 10((y - \frac{3}{5})^2 + \frac{108}{25}. \end{aligned}$$

Since a square of a real number is always positive, we conclude that the minimum value occurs when $y = 3/5$.

Some of you may know that the line segment joining a point in the plane to the nearest point on a given line will be perpendicular to the given line. The proof of this fact uses calculus as in the argument above. You may use this fact to check if our answer is correct.∎

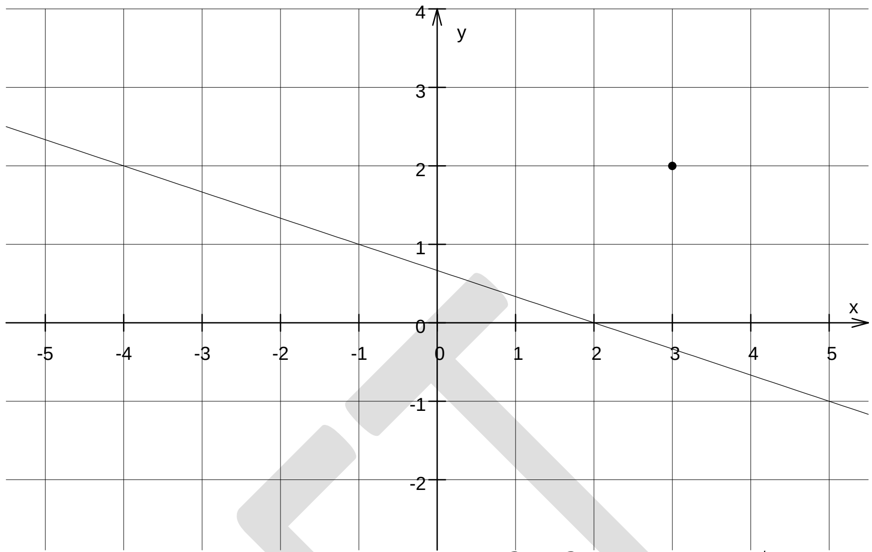
Example. One side of a rectangle rests on the x-axis and two vertices touch the circle $x^2 + y^2 = 1$. Find the largest possible area.

Solution. We suppose that the vertex in the first quadrant is (x,y), then the area of the

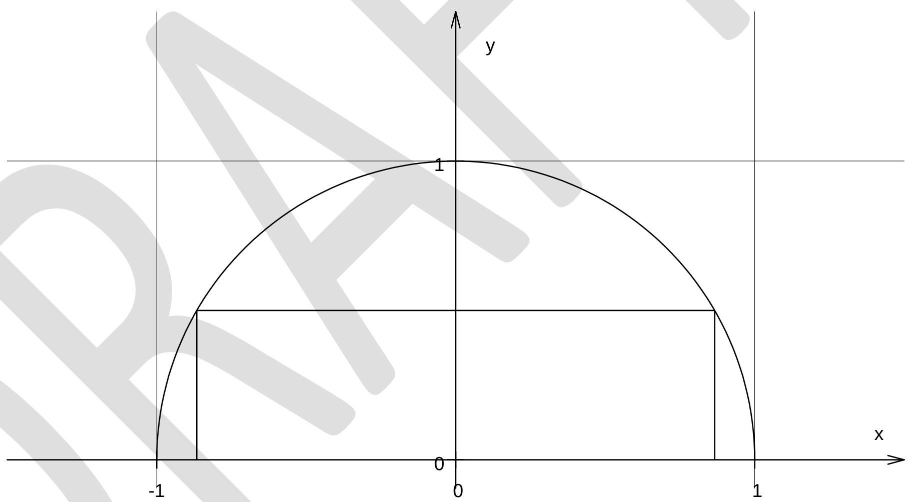


rectangle is

$$A = 2xy.$$



Because the point (x,y) lies on the circle, we have the relation $x^2 + y^2 = 1$. We may



solve this equation to eliminate one of the variables, let us pick y. Then, $y = \sqrt{1 - x^2}$ so that

$$A(x) = 2x\sqrt{1 - x^2}$$

and we need to find the maximum value of the objective funciton $A(x)$ on the interval [0,1]. The domain is restricted to [0,1] since the point (x,y) was chosen in the first quadrant.

Thus, we use the procedure for finding the maximum value of a continuous function√ on a closed interval. We test the endpoints, 0, 1, and the critical number $1/\sqrt{2}$. The maximum occurs at $x = 1/\sqrt{2}$. The largest area is 1.▌

Second solution. When working with circles, one can also use trigonometric functions. If we let the vertex in the first quadrant be $(\cos\theta, \sin\theta)$, then our goal is find the maximum value of

$$A(\theta) = 2\cos\theta\,\sin\theta.$$

This maximum occurs when $\theta = \pi/4$.▪

UNIT 5 DIFFERENTIATION OF VECTOR FUNCTIONS

5.1 Scalar and Vector Functions.

- Define scalar and vector function.
- Explain domain and range of a vector function.

SCALAR AND VECTOR FUNCTIONS, POINT FUNCTIONS, SCALAR POINT FUNCTIONS, VECTOR POINT FUNCTIONS, SCALAR AND VECTOR FIELDS

In vector analysis we deal with scalar and vector functions.

Def. Scalar function. A scalar function is a function that assigns a real number (i.e. a scalar) to a set of real variables. Its general form is

$$u = u(x_1, x_2, \dots, x_n)$$

where $x_1, x_2, \dots, x_n$ are real numbers.

Def. Vector function. A vector function is a function that assigns a vector to a set of real variables. Its general form is

$$\mathbf{F} = f_1(x_1, x_2, \dots, x_n)\,\mathbf{i} + f_2(x_1, x_2, \dots, x_n)\,\mathbf{j} + f_3(x_1, x_2, \dots, x_n)\,\mathbf{k}$$

or equivalently,

$$F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ f_3(x_1, x_2, \dots, x_n) \end{bmatrix}$$

where $x_1, x_2, \dots, x_n$ are real numbers.

Example 1. Function defining a space curve. Let

$\mathbf{R}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}$

or equivalently,

$$R(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}$$

be a radius vector to a point P(x, y, z) in space which moves as t increases in value. It traces out a curve in space. The parametric representation of space curves is

x = x(t)
y = y(t)
z = z(t) .

Example 2. Function defining a surface in space. The function

$\mathbf{R}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k}$

represents a surface in space. Surfaces are represented by parametric equations of the type

x = x(u, v)
y = y(u, v)
z = z(u, v)

If v is regarded as a parameter, u a variable, then this system describes a space curve. For each value of v there is another space curve, thus generating a surface.

Def. Point function. A point function u = f(P) is a function that assigns some number or value u to each point P of some region R of space. Examples of point functions are scalar point functions and vector point functions.

Def. Scalar point function. A scalar point function is a function that assigns a real number (i.e. a scalar) to each point of some region of space. If to each point (x, y, z) of a region R in space there is assigned a real number $u = \Phi(x, y, z)$, then Φ is called a scalar point function.

Examples. 1. The temperature distribution within some body at a particular point in time. 2. The density distribution within some fluid at a particular point in time.

Syn. scalar function of position

Scalar field. A scalar point function defined over some region is called a **scalar field**. A scalar field which is independent of time is called a **stationary** or **steady-state scalar field.**

A scalar field that varies with time would have the representation

$$u = \Phi(x, y, z, t)$$

Def. Vector point function. A vector point function is a function that assigns a vector to each point of some region of space. If to each point (x, y, z) of a region R in space there is assigned a vector $\mathbf{F} = \mathbf{F}(x, y, z)$, then **F** is called a vector point function. Such a function would have a representation

$$\mathbf{F} = f_1(x, y, z)\,\mathbf{i} + f_2(x, y, z)\,\mathbf{j} + f_3(x, y, z)\,\mathbf{k}$$

or equivalently,

$$F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} f_1(x,y,z) \\ f_2(x,y,z) \\ f_3(x,y,z) \end{bmatrix}$$

Syn. vector function of position

Vector field. A vector point function defined over some region is called a **vector field**. A vector field which is independent of time is called a **stationary** or **steady-state vector field.**

A vector field that varies with time would have the representation

$$\mathbf{F} = f_1(x, y, z, t)\,\mathbf{i} + f_2(x, y, z, t)\,\mathbf{j} + f_3(x, y, z, t)\,\mathbf{k}$$

or equivalently,

$$F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} f_1(x,y,z,t) \\ f_2(x,y,z,t) \\ f_3(x,y,z,t) \end{bmatrix}$$

5.2 Limit and Continuity Determine limit of a vector function

- Demonstrate the following properties of limits of a vector function, by using technique for algebra of limits of scalar function to,:
  - I. The limit of the sum (difference) of two vector functions is the sum (difference) of their limits.
  - II. The limit of the dot product of two vector functions is the dot product of their limits.
  - III. The limit of the cross product of two vector functions is the cross product of their limits.
  - IV. The limit of the product of a scalar function and a vector function is the product of their limits.
- Describe continuity of a vector function through examples.

Limits and Continuity

SUMMARY LIMITS AND CONTINUITY

The concept of the limit is one of the most crucial things to understand in order to prepare for calculus. A limit is a number that a function approaches as the independent variable of the function approaches a given value. For example, given the function *f* (*x*) = 3*x*, you could say, "The limit of *f* (*x*) as *x* approaches 2 is 6." Symbolically, this is written $\lim_{x \to 2} f(x) = 6$. In the following sections, we will more carefully define a limit, as well as give examples of limits of functions to help clarify the concept.

Continuity is another far-reaching concept in calculus. A function can either be continuous or discontinuous. One easy way to test for the continuity of a function is to see whether the graph of a function can be traced with a pen without lifting the pen from the paper. For the math that we are doing in precalculus and calculus, a conceptual definition of continuity like this one is probably sufficient, but for higher math, a more technical definition is needed. Using limits, we'll learn a better and far more precise way of defining continuity as well. With an understanding of the concepts of limits and continuity, you are ready for calculus.

Mathematics | Limits, Continuity and Differentiability

1. Limits –

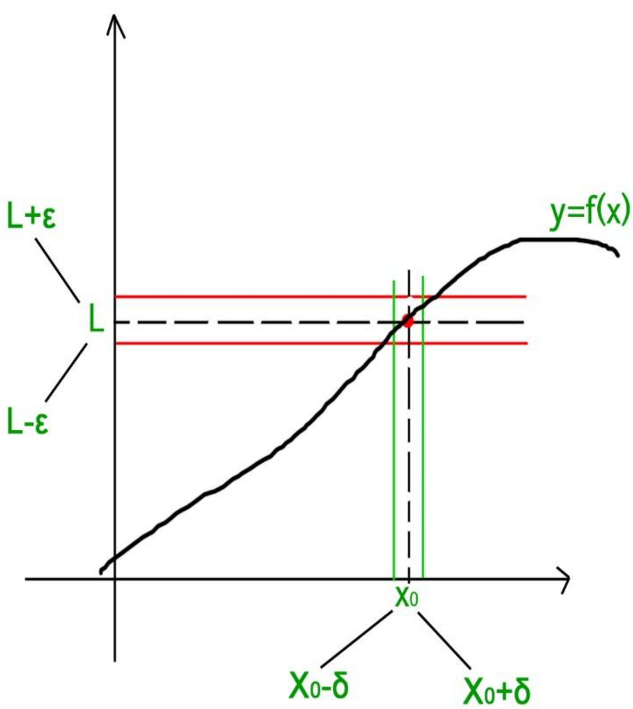
For a function the limit of the function at a point is the value the function achieves at a point which is very close to .

Formally,

Let be a function defined over some interval containing , except that it may not be defined at that point.

We say that, if there is a number for every number such that
whenever

The concept of limit is explained graphically in the following image –



As is clear from the above figure, the limit can be approached from either sides of the number line i.e. the limit can be defined in terms of a number less that or in terms of a number greater than . Using this criteria there are two types of limits –
Left Hand Limit – If the limit is defined in terms of a number which is less than then the limit is said to be the left hand limit. It is denoted as which is equivalent

to where and .
Right Hand Limit – If the limit is defined in terms of a number which is greater than then the limit is said to be the right hand limit. It is denoted as which is equivalent to where and .

Existence of Limit – The limit of a function at exists only when its left hand limit and right hand limit exist and are equal and have a finite value i.e.

Some Common Limits –

L'Hospital Rule –

If the given limit is of the form or i.e. both and are 0 or both and are , then the limit can be solved by **L'Hospital Rule**.
If the limit is of the form described above, then the L'Hospital Rule says that –

where and obtained by differentiating and .
If after differentitating, the form still exists, then the rule can be applied continuously until the form is changed.

- **Example 1 –** Evaluate
- **Solution –** The limit is of the form , Using L'Hospital Rule and differentiating numerator and denominator
- **Example 2 –** Evaluate
- **Solution –** On multiplying and dividing by and re-writing the limit we get –

The Chain Rule

6.5	Integration using partial fractions

- Use partial fractions to find $\int \frac{f(x)}{g(x)}\,dx$, where f(x) and g(x) are algebraic functions such that g(x) $\neq 0$

11. Integration by Partial Fractions

by M. Bourne

Partial fraction decomposition

$$\frac{f(x)}{(x+a)(x+b)} \equiv \frac{A}{x+a} + \frac{B}{x+b}$$

Partial fraction decomposition - linear factors

If the integrand (the expression after the integral sign) is in the form of an algebraic fraction and the integral cannot be evaluated by simple methods, the fraction needs to be expressed in **partial fractions** before integration takes place.

The steps needed to decompose an algebraic fraction into its partial fractions results from a consideration of the reverse process − addition (or subtraction).

Consider the following addition of algebraic fractions:

\displaystyle\frac{1}{{{x}+{2}}}+\frac{5}{{{x}+{3}}}=\frac{{{\left({x}+{3}\right)}+{5}{\left({x}+{2}\right)}}}{{{\left({x}+{2}\right)}{\left({x}+{3}\right)}}}*x*+21+*x*+35=(*x*+2)(*x*+3)(*x*+3)+5(*x*+2)

\displaystyle=\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}=*x*2+5*x*+66*x*+13

In this section, we want to go the other way around. That is, if we were to **start** with the expression

\displaystyle\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}*x*2+5*x*+66*x*+13

and try to find the fractions whose sum gives this result, then the two fractions obtained, i.e.

\displaystyle\frac{1}{{{x}+{2}}}*x*+21 and \displaystyle\frac{5}{{{x}+{3}}}*x*+35,

are called the **partial fractions** of \displaystyle\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}*x*2+5*x*+66*x*+13.

We decompose fractions into partial fractions like this because:

- It makes certain integrals much easier to do, and
- It is used in the [Laplace transform](), which we meet later.

So if we needed to integrate this fraction, we could simplify our integral in the following way:

\displaystyle\int\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}{\left.{d}{x}\right.}∫*x*2+5*x*+66*x*+13 *dx* \displaystyle=\int\frac{1}{{{x}+{2}}}{\left.{d}{x}\right.}+\int\frac{5}{{{x}+{3}}}{\left.{d}{x}\right.}=∫*x*+21*dx*+∫*x*+35*dx*

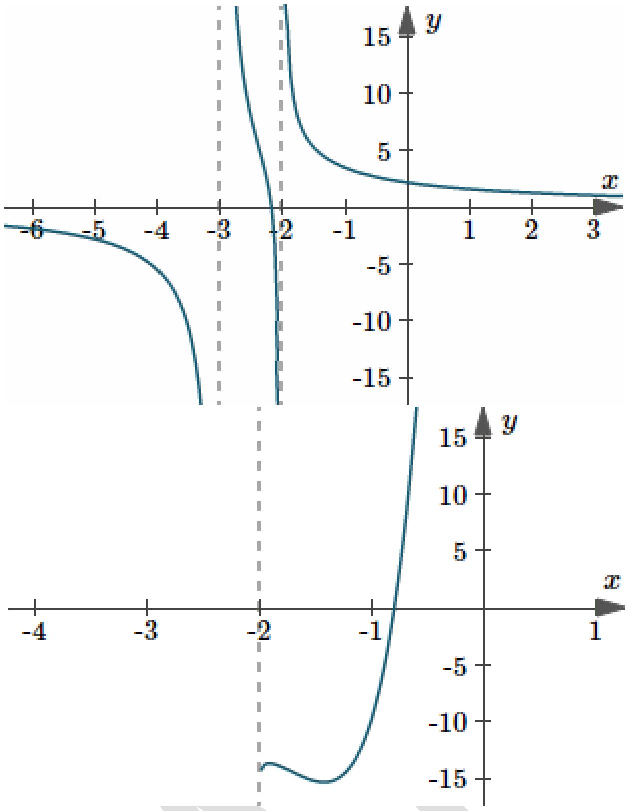
We integrate the two fractions using what we learned in [Basic Logarithmic Form]():

\displaystyle\int\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}{\left.{d}{x}\right.}=\int\frac{1}{{{x}+{2}}}{\left.{d}{x}\right.}+\int\frac{5}{{{x}+{3}}}{\left.{d}{x}\right.}∫*x*2+5*x*+66*x*+13*dx*=∫*x*+21*dx*+∫*x*+35*dx*

\displaystyle= \ln{{\left({x}+{2}\right)}}+{5}\ \ln{{\left({x}+{3}\right)}}+{K}=ln(*x*+2)+5 ln(*x*+3)+*K*

\displaystyle= \ln{{\left[{\left({x}+{2}\right)}{\left({x}+{3}\right)}^{5}\right]}}+{K}=ln[(*x*+2)(*x*+3)5]+*K*

Here are the graphs of the original function, and the integral we just found.



Graph of \displaystyle{y}{\left({x}\right)}=\frac{{{6}{x}+{13}}}{{{x}^{2}+{5}{x}+{6}}}*y*(*x*)=*x*2+5*x*+66*x*+13, and a typical integral solution \displaystyle{y}{\left({x}\right)}=\ln{{\left({x}+{2}\right)}}{\left({x}+{3}\right)}^{5}+{10}*y*(*x*)=ln(*x*+2)(*x*+3)5+10 (where I've used \displaystyle{K}={10}*K*=10).

Next we'll see how to split such a fraction into its partial fractions.

Continues below ⇩

Expressing a Fractional Function In Partial Fractions

RULE 1:

Before a fractional function can be expressed directly in partial fractions, the **numerator must be of at least one degree less than the denominator**.

Example 1

The fraction

\displaystyle\frac{{{2}{x}^{2}+{3}}}{{{x}^{3}-{1}}}*x*3−12*x*2+3

can be expressed in partial fractions whereas the fraction

\displaystyle\frac{{{2}{x}^{3}+{3}}}{{{x}^{3}-{1}}}*x*3−12*x*3+3 cannot be expressed directly in partial fractions.

However, by division

$$\displaystyle\frac{{{2}{x}^{3}+{3}}}{{{x}^{3}-{1}}}$$
$$\displaystyle={2}+\frac{5}{{{x}^{3}-{1}}}$$

and the resulting fraction can be expressed as a sum of partial fractions.

(Note: The denominator of the fraction must be factored before you can proceed.)

RULE 2: Denominator Containing Linear Factors

For each linear factor $\displaystyle{\left({a}{x}+{b}\right)}$ in the denominator of a rational fraction, there is a partial fraction of the form

$$\displaystyle\frac{A}{{{a}{x}+{b}}}$$

where $\displaystyle{A}$ is a constant.

Example 2

Express the following in partial fractions. $\displaystyle\frac{{{3}{x}}}{{{\left({2}{x}+{1}\right)}{\left({x}+{4}\right)}}}$

Answer

RULE 3: Denominator Containing Repeated Linear Factors

If a linear factor is repeated $\displaystyle{n}$ times in the denominator, there will be $\displaystyle{n}$ corresponding partial fractions with degree 1 to $\displaystyle{n}$.

For example, the partial fractions for

$$\displaystyle\frac{{{5}{x}^{3}+{7}{x}-{9}}}{{{\left({x}+{1}\right)}^{4}}}$$

will be of the form:

$$\displaystyle\frac{{{5}{x}^{3}+{7}{x}-{9}}}{{{\left({x}+{1}\right)}^{4}}} \displaystyle\equiv\frac{A}{{{x}+{1}}}+\frac{B}{{{\left({x}+{1}\right)}^{2}}} \displaystyle+\frac{C}{{{\left({x}+{1}\right)}^{3}}} \displaystyle+\frac{D}{{{\left({x}+{1}\right)}^{4}}}$$

(The sign $\equiv$ means "is identically equal to". We normally apply this between 2 expressions when we wish them to be equivalent. It's OK to use the ordinary equals sign, too.)

Example 3

(a) Express the following as a sum of partial fractions.

$$\displaystyle\frac{{x}+{5}}{{\left({x}+{3}\right)}^{2}}$$

Answer

(b) Express the following as a sum of partial fractions.

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

Answer

NOTE: Scientific Notebook can do all this directly for us using **Polynomials/Partial Fractions.**

RULE 4:Denominator Containing a Quadratic Factor

Corresponding to any quadratic factor $\displaystyle{\left({a}{x}^{2}+{b}{x}+{c}\right)}$ in the denominator, there will be a partial fraction of the form

$$\displaystyle\frac{{A}{x}+{B}}{{a}{x}^{2}+{b}{x}+{c}}$$

Example 4

Express the following in partial fractions.

$$\displaystyle\frac{{x}^{3}-{2}}{{x}^{4}-{1}}$$

Answer

Note: Repeated quadratic factors in the denominator are dealt with in a similar way to repeated linear factors.

Example:

$$\displaystyle\frac{{x}^{2}+{1}}{{\left({x}^{2}+{x}+{1}\right)}^{2}}$$

$$\displaystyle\equiv\frac{{{A}{x}+{B}}}{{{x}^{2}+{x}+{1}}}+\frac{{{C}{x}+{D}}}{{{\left({x}^{2}+{x}+{1}\right)}^{2}}}$$

Summary

Denominator containing …	**Expression**	**Form of Partial Fractions**
a. Linear factor	$\displaystyle\frac{{ f{{\left({x}\right)}}}}{{{\left({x}+{a}\right)}{\left({x}+{b}\right)}}}$	$\displaystyle\frac{A}{{{x}+{a}}}+\frac{B}{{{x}+{b}}}$
b. Repeated linear factors	$\displaystyle\frac{{ f{{\left({x}\right)}}}}{{{\left({x}+{a}\right)}^{3}}}$	$\displaystyle\frac{A}{{{x}+{a}}}+\frac{B}{{{\left({x}+{a}\right)}^{2}}}+\frac{C}{{{\left({x}+{a}\right)}^{3}}}$
c. Quadratic term (which cannot be factored)	$\displaystyle\frac{{ f{{\left({x}\right)}}}}{{{\left({a}{x}^{2}+{b}{x}+{c}\right)}{\left({g}{x}+{h}\right)}}}$	$\displaystyle\frac{{{A}{x}+{B}}}{{{a}{x}^{2}+{b}{x}+{c}}}+\frac{C}{{{g}{x}+{h}}}$

Note: In each of the above cases $\displaystyle f{{\left({x}\right)}}$ must be of less degree than the relevant denominator.

6.6 Definite Integrals

- Available Define definite integral as the limit of a sum.
- Describe the fundamental theorem of integral calculus and recognize the following basic properties:

I. $\int_a^a f(x)dx = 0$

II. $\int_a^b f(x)dx = 0,$

III. $\int_a^b f(x)dx = -\int_b^a f(x)dx$

IV. $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)fx, a < c < b,$

V. $\int_{-a}^a f(x)dx = \begin{cases} 2\int_0^a f(x)dx \text{ when } f(-x) = f(x) \\ 0 \quad \text{when } f(-x) = -f(x) \end{cases}$

- Extend techniques of integration using properties to evaluate definite integrals.
- Represent definite integral at the area under the curve.
- Apply definite integral to calculate area under the curve.
- Apply MAPLE command int to evaluate definite and indefinite integrals.

Definite Integral as a Limit of a Sum

Imagine a curve above the x-axis. The function of this graph is a continuous function defined on a closed interval [a, b], where all the values of the function are non-negative. The area bound between the curve, the points ‘x = a’ and ‘x = b’ and the x-axis is the definite integral $\int_a^b f(x)\,dx$ of any such continuous function ‘f’.

Suggested Videos

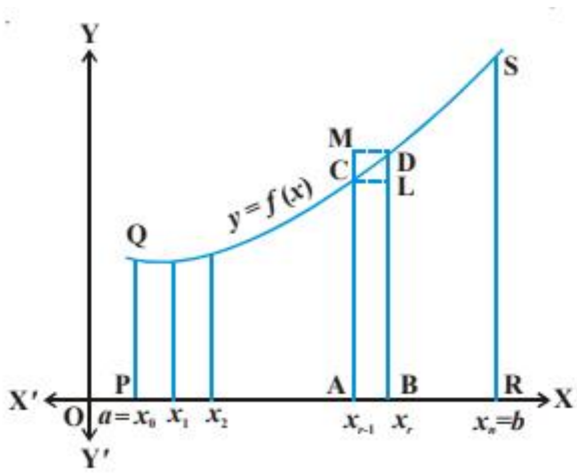
Indefinite Intergration Geometrical Meaning

Indefinite Intergration Introduction

Integration By Changing Integrand Without Substitution

Definite Integral as a Limit of a Sum

Look at the following graph:



To understand this, let's evaluate the area PRSQP between the curve y = f(x), x-axis and the coordinates 'x = a' and 'x = b'. Now, divide the interval [a, b] into 'n' equal sub-intervals denoted as:

$[x_0, x_1], [x_1, x_2], [x_2, x_3]$ …. $[x_{n-1}, x_n]$, where,
$x_0 = a$, $x_1 = a + h$, $x_2 = a + 2h$, $x_3 = a + 3h$ ….. $x_r = a + rh$ and $x_n = b = a + nh$
Or, $n = (b - a)/h$. Note that as $n \to \infty$, $h \to 0$.

Now, the region PRSQP under consideration is the sum of all the 'n' sub-regions, where each sub-region is defined on subintervals $[x_{r-1}, x_r]$, r = 1, 2, 3 … n. Now, look at the region ABDM in the figure above. We can make the following observation:

Area of the rectangle (ABLC) $<$ Area of the region (ABDCA) $<$ Area of the rectangle (ABDM)   (1)

Also, note that as, $h \to 0$ or $x_r - x_{r-1} \to 0$, all these three areas become nearly equal to each other. Hence, we have

$s_n = h [f(x_0) + f(x_1) + f(x_2) + \dots f(x_{n-1})] = h \sum_{r=0}^{n-1} f(x_r) \dots (2)$
and, $S_n = h [f(x_1) + f(x_2) + f(x_3) + \dots f(x_n)] = h \sum_{r=1}^{n} f(x_r) \dots (3)$

Here, s_n and S_n denote the sum of areas of all lower rectangles and upper rectangles raised over subintervals $[x_{r-1}, x_r]$ for $r = 1, 2, 3, \dots, n$, respectively. To bring it into perspective, equation (1) can be re-written as:

$s_n <$ area of the region (PRSQP) $< S_n \dots (4)$

Browse more Topics under Integrals

- Fundamental Theorem of Calculus
- Introduction to Integration
- Properties of Indefinite Integrals
- Properties of Definite Integrals
- Integration by Partial Fractions
- Integration by Parts
- Integration by Substitutions
- Integral of Some Particular Functions
- Integral of the Type e^x[f(x) + f'(x)]dx

As $n \to \infty$, these strips become narrower

Further, it is assumed that the limiting values of (2) and (3) are the same in both cases and the common limiting value is the required area under the curve. Symbolically, we have

$\lim_{n \to \infty} S_n = \lim_{n \to \infty} s_n =$ area of the region (PRSQP) $= \int_a^b f(x)\,dx \dots (5)$

This area is also the limiting value of any area which is between that of the rectangles below the curve and that of the rectangles above the curve. For convenience, we shall take the rectangles having height equal to that of the curve at the left-hand-edge of each sub-interval. Hence, equation (%) is re-written as:

$\int_a^b f(x)\,dx = \lim_{n \to \infty} h [f(a) + f(a + h) + \dots + f(a + \{n - 1\}h)]$
Or, $\int_a^b f(x)\,dx = (b - a) \lim_{n \to \infty} (1/n) [f(a) + f(a + h) + \dots + f(a + \{n - 1\}h)] \dots (6)$

Where, $h = (b - a)/n \to 0$ as $n \to \infty$. This equation is the definition of Definite Integral as the limit of a sum.

Note: The value of the definite integral of a function over any particular interval depends on the function and the interval, but not on the variable of integration that we choose to represent the independent variable. Hence, the variable of integration is called a dummy variable.

Example 1

Find $\int_0^2 (x^2 + 1)\, dx$ as the limit of a sum.

Solution: From equation (6) above, we know that
$\int_a^b f(x)\, dx = (b - a) \lim_{n \to \infty} (1/n) [f(a) + f(a + h) + \ldots. + f(a + \{n - 1\}h)]$
Where, $h = (b - a)/n$

In this example, we have $a = 0$, $b = 2$, $f(x) = (x^2 + 1)$ and $h = (2 - 0)/n = 2/n$. Therefore,

$\int_0^2 (x^2 + 1)\, dx = 2 \lim_{n \to \infty} (1/n) [f(0) + f(2/n) + f(4/n) + \ldots. + f(2\{n - 1\}/n)]$
$= 2 \lim_{n \to \infty} (1/n) [1 + \{(2^2/n^2) + 1\} + \{(4^2/n^2) + 1\} + \ldots. + \{(2n - 2)^2/n^2 + 1\}]$
$= 2 \lim_{n \to \infty} (1/n) [1 + 1 + 1 + \ldots + 1(\text{n-times})] + 1/n^2) [2^2 + 4^2 + \ldots (2n - 2)^2]$
$= 2 \lim_{n \to \infty} (1/n) [n + 2^2/n^2 (1^2 + 2^2 + \ldots (n - 1)^2]$
$= 2 \lim_{n \to \infty} (1/n) [n + 4/n^2 \{(n - 1)\, n\, (2n - 1) / 6\}]$
$= 2 \lim_{n \to \infty} (1/n) [n + 2/3 \{(n - 1) (2n - 1) / n\}]$
$= 2 \lim_{n \to \infty} (1/n) [n + 2/3 (1 - 1/n) (2 - 1/n)]$
As $n \to \infty$, $1/n \to 0$. Therefore, we have
$\int_0^2 (x^2 + 1)\, dx = 2 [1 + 4/3] = 14/3$.

More Solved Examples for You

Question: Find $\int_a^b x\, dx$ as the limit of a sum.

Solution: From equation (6) above, we know that
$\int_a^b f(x)\, dx = (b - a) \lim_{n \to \infty} (1/n) [f(a) + f(a + h) + \ldots. + f(a + \{n - 1\}h)]$
Where, $h = (b - a)/n$

In this example, we have
$a = a$, $b = b$, $f(x) = x$ and $h = (b - a)/n$.

Also, $f(a) = a$
$f(a + h) = a + h$
$f(a + 2h) = a + 2h$

$f(a + 3h) = a + 3h$ ……
$f(a + (n – 1)h) = a + (n – 1)h$

Therefore,
$\int_a^b x\,dx = (b – a) \lim_{n\to\infty} (1/n) [f(a) + f(a + h) + f(a + 2h) + …. + f(a + \{n – 1\}h)]$
$= (b – a) \lim_{n\to\infty} (1/n) [a + (a + h) + (a + 2h) + …. + (a + (n – 1)h)]$
$= (b – a) \lim_{n\to\infty} (1/n) \{[a + a + a + … + a(n\text{-times})] + h + 2h + …. + (n – 1)h\}$
$= (b – a) \lim_{n\to\infty} (1/n) [na + h(1 + 2 + ….. + (n – 1)]$

Now, we know that, $1 + 2 + 3 + … + n = n(n + 1)/2$. Hence,
$1 + 2 + 3 + … + (n – 1) = (n – 1)(n – 1 + 1)/2 = n(n – 1)/2$

Therefore,
$\int_a^b x\,dx = (b – a) \lim_{n\to\infty} (1/n) [na + hn(n – 1)/2]$
$= (b – a) \lim_{n\to\infty} [na/n + hn(n – 1)/2n]$
$= (b – a) \lim_{n\to\infty} [a + (n – 1)h/2]$

Replacing h by $(b – a)/n$, we get
$\int_a^b x\,dx = (b – a) \lim_{n\to\infty} [a + (n – 1) (b – a) / 2n]$
$= (b – a) \lim_{n\to\infty} [a + (n/n – 1/n) \{(b – a)/2\}]$
$= (b – a) \lim_{n\to\infty} [a + (1 – 1/n)\{(b – a)/2\}]$
$= (b – a) [a + (1 – 1/\infty) \{(b – a)/2\}]$
$= (b – a) [a + (1 – 0)\{(b – a)/2\}]$
$= (b – a) [a + (b – a)/2]$
$= (b – a) (2a + b – a)/2$
$= (b – a) (b + a)/2$
$= (b^2 – a^2)/2$

Hence, the definite integral $\int_a^b x\,dx$ as the limit of sum is [(b2 – a2)/2].

Area Under a Curve

Calculating the area under a curve.

Definite Integrals

So far when integrating, there has always been a constant term left. For this reason, such integrals are known as indefinite integrals. With definite integrals, we integrate a function between 2 points, and so we can find the precise value of the integral and there is no need for any unknown constant terms [the constant cancels out].

$$\int_4^8 x + 3x^2\,dx$$

This means the integral of $x + 3x^2$ with respect to x, between x=8 and x=4.

Integrating:

$$\left[\frac{x^2}{2} + \frac{3x^3}{3} + c\right]_4^8$$

$$= \left(\frac{8^2}{2} + \frac{3(8)^3}{3} + c\right) - \left(\frac{4^2}{2} + \frac{3(4)^3}{3} + c\right)$$

$= (32 + 512 + c) - (8 + 64 + c)$
$= 544 + c - 72 - c$
$= \underline{\underline{472}}$

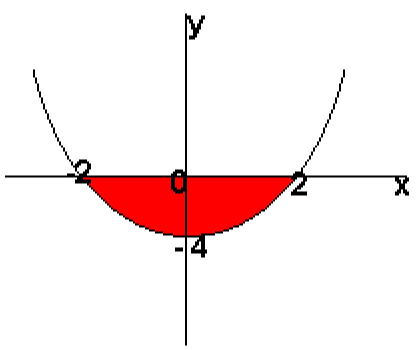
Note that the constant terms disappear, and always will with definite integrals so you don't have to write them in.

The Area Under a Curve

The area under a curve between two points can be found by doing a definite integral between the two points.

To find the area under the curve y = f(x) between x = a and x = b, integrate y = f(x) between the limits of a and b.

Example: What is the area between the curve $y = x^2 - 4$ and the x axis?



The shaded area is the area that we want.

We can easily work out that the curve crosses the x axis when x = - 2 and x = 2. To find the area, therefore, we integrate the function between -2 and 2.

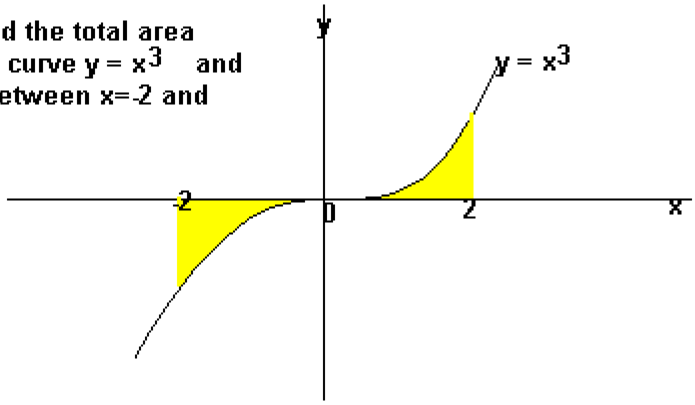
$$\int_{-2}^{2} (x^2 - 4)\,dx = \left[\frac{x^3}{3} - 4x\right]_{-2}^{2} = \left(\frac{8}{3} - 8\right) - \left(\frac{-8}{3} + 8\right)$$

$= 16/3 - 16$
$= \underline{-10.67}$ (2d.p.)

Note: the area is negative because it is below the x-axis. Areas above the x-axis, on the other hand, give positive results.

Areas under the x-axis will come out negative and areas above the x-axis will be positive. This means that you have to be careful when finding an area which is partly above and partly below the x-axis.

Example: find the total area between the curve $y = x^3$ and the x-axis between x=-2 and x=2.



If we simply integrated x^3 between -2 and 2, we would get:

$$\left[\frac{x^4}{4}\right]_{-2}^{2} = 4 - 4 = 0$$

So instead, we have to split the graph up and do two separate integrals.

$$\int_0^2 x^3\,dx = \left[\frac{x^4}{4}\right]_0^2 = 16/4 - 0 = 4$$

$$\int_{-2}^0 x^3\,dx = \left[\frac{x^4}{4}\right]_{-2}^0 = 0 - 16/4 = -4 \quad \text{(so area is 4).}$$

We then add these two up to get: $\underline{\underline{8 \text{ units}^2}}$

You may also be asked to find the area between the curve and the y-axis. To do this, integrate with respect to y.

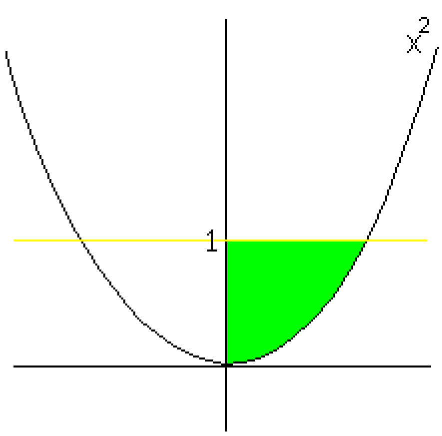
Example

Find the area bounded by the lines y = 0, y = 1 and $y = x^2$.

$$\text{Area} = \int_0^1 y^{1/2}\,dy \qquad \text{since } x = y^{1/2}$$

$$= \left[\frac{2y^{3/2}}{3}\right]_0^1$$

$$= 2/3$$



Suppose an object moves so that its speed, or more properly velocity, is given by v(t)=−t2+5tv(t)=−t2+5t, as shown in figure 7.3.1. Let's examine the motion of this object carefully. We know that the velocity is the derivative of position, so position is given by s(t)=−t3/3+5t2/2+Cs(t)=−t3/3+5t2/2+C. Let's suppose that at time t=0t=0 the object is at position 0, so s(t)=−t3/3+5t2/2s(t)=−t3/3+5t2/2; this function is also pictured in figure 7.3.1.

66	66
66	2222

UNIT 7 PLANE ANALYTIC GEOMETRY-STRAIGHT LINE

7.1 Division of a Line Segment

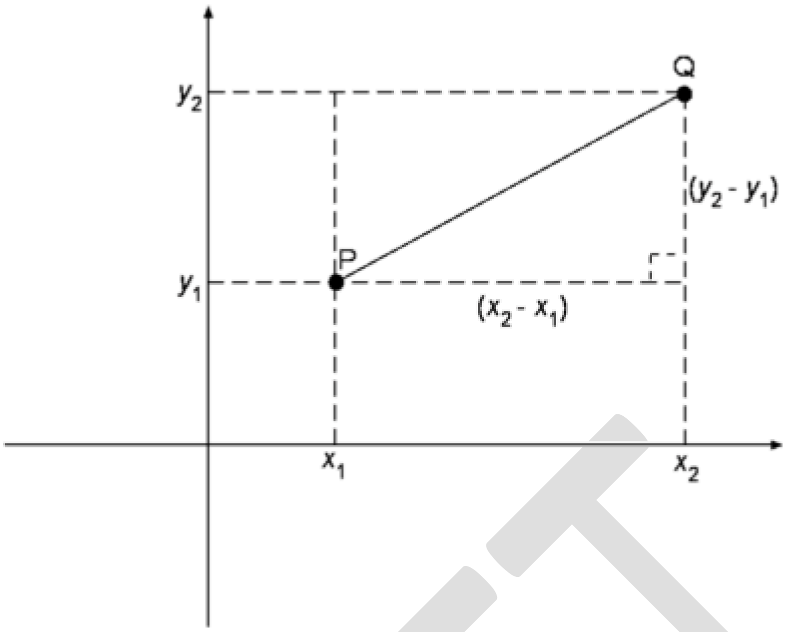
- Explain distance formula to calculate distance between two points given in Cartesian plane.
- Find coordinates of a point that divides the line segment in given ratio (internally and externally).
- Verify that medians and angle bisectors of a triangle are concurrent.

Pythagoras Theorem

Distance Between Two Points

We use the Pythagoras Theorem to derive a formula for finding the distance between two points in 2- and 3- dimensional space.

Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be two points on the Cartesian plane (see picture below).



Then from the Pythagoras Theorem we find that the distance between *P* and *Q* is

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

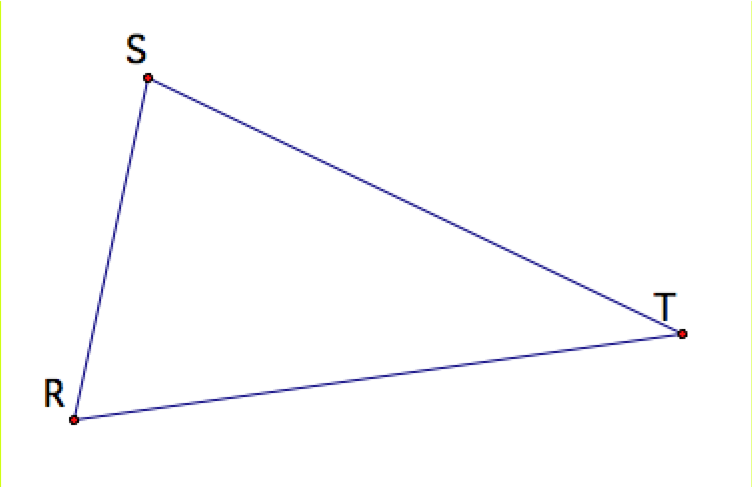
In a similar way, it can be proved that if $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ are two points in the 3-dimensional space, the distance between *P* and *Q* is

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

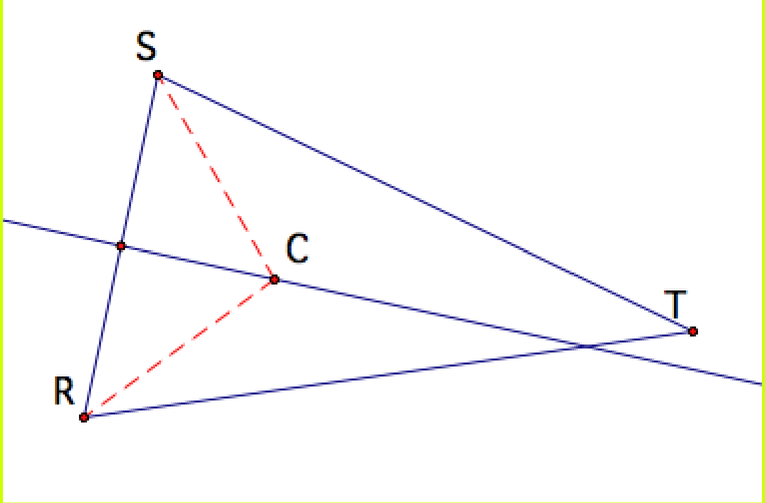
Centers of a Triangle
by
Susan Sexton

Prove that the three perpendicular bisectors of a triangle are concurrent.

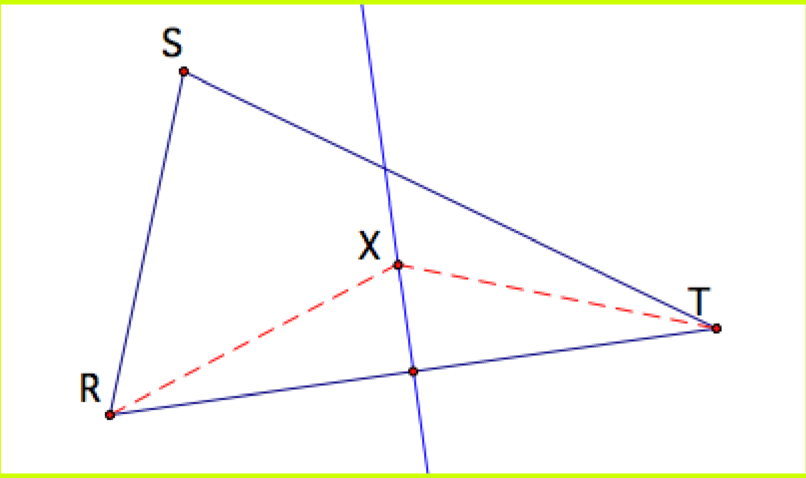
Start with triangle RST.



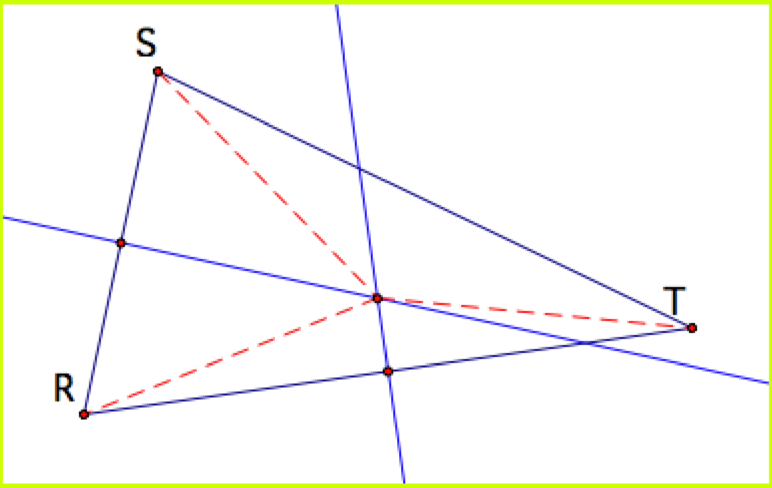
The perpendicular bisector of segment RS is the set of all points that are the same distant (equidistant) from both R and S. Therefore any point that lies on the perpendicular bisector, for instance, point C, will result in SC=RC.



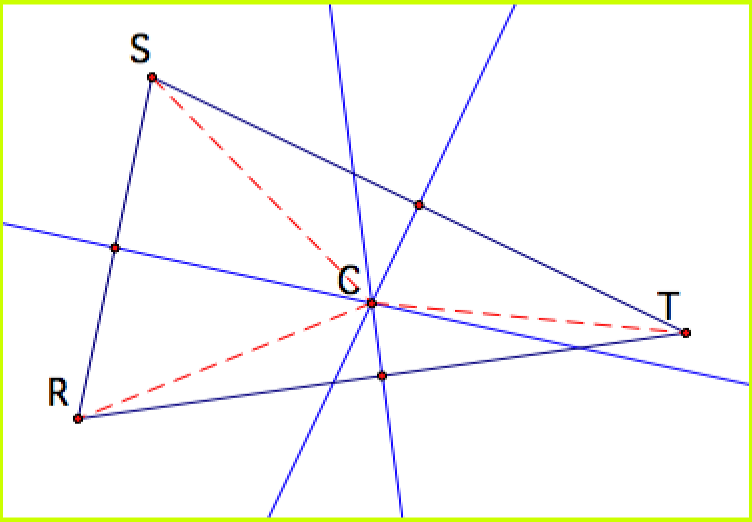
Similarly, the perpendicular bisector of segment RT will be all the points equidistant from R and T. Therefore if X lies on the perpendicular bisector of RT then RX = XT.



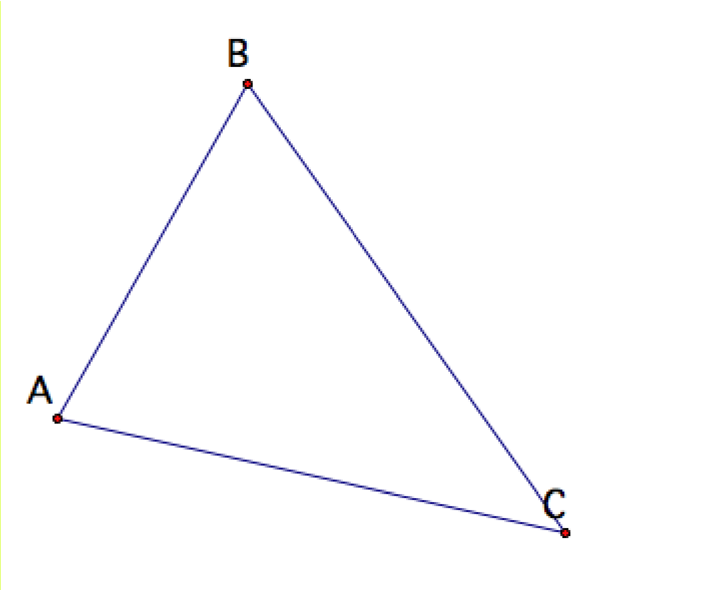
So considering segment RS and segment RT and their respective perpendicular bisectors then the intersection of the two perpendicular bisectors will result in the point that is equidistant from R and S and R and T.



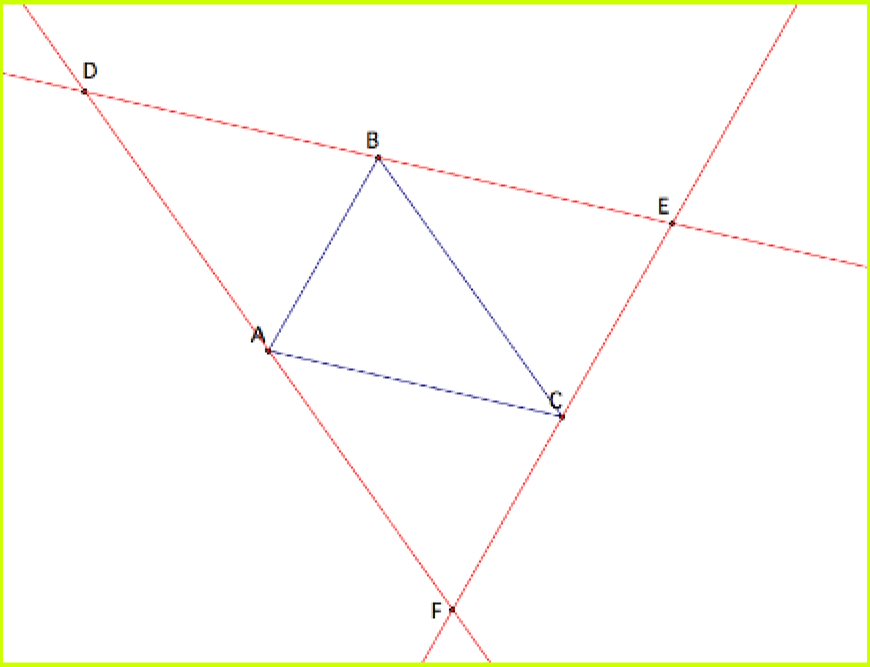
Therefore, using C as the point of intersection, we have RC = SC an RC = TC. So by transitivity we have SC = TC. Since C is equidistant from S and T then C lies on the perpendicular bisector of segment ST. Therefore C is the point of concurrency between the three perpendicular bisectors.



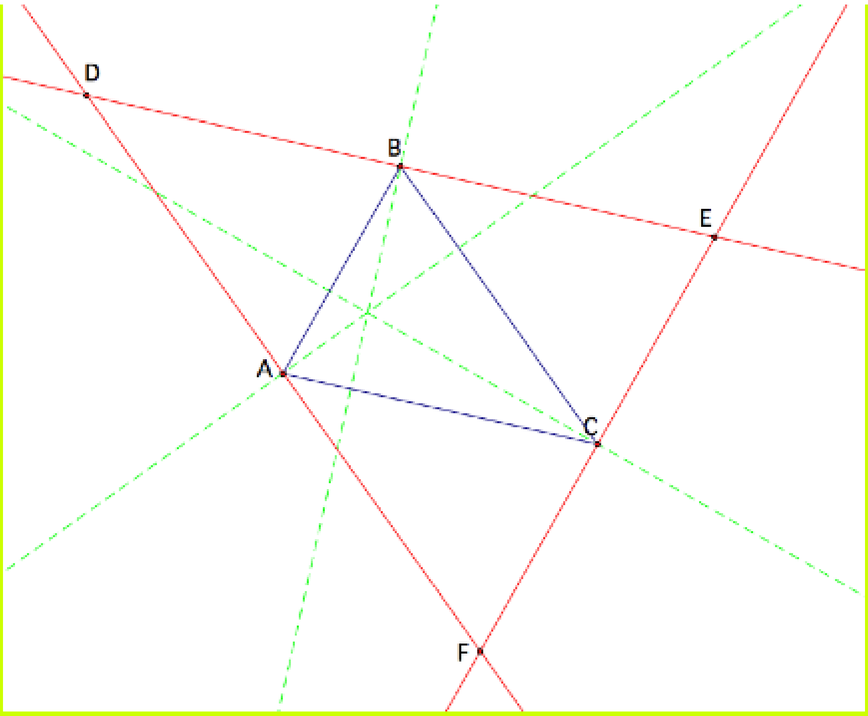
The perpendicular bisectors of a triangle are also the altitudes of another triangle. Consider triangle ABC. I will find another triangle for which triangle ABC is the medial triangle.



Start with finding the line parallel to each segment of the triangle through the vertex opposite to the segment. For example, find the line parallel to segment AB and passing through vertex C. This will result in a new triangle DEF whose medial triangle is ABC.

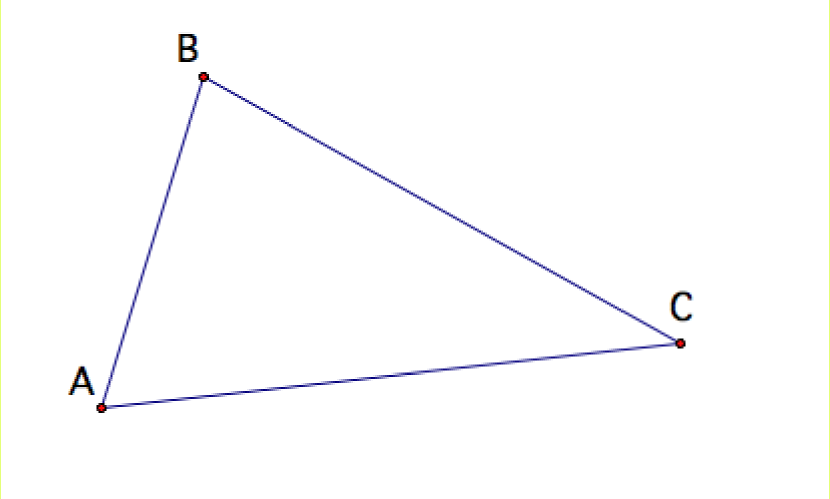


By finding the perpendicular bisectors of triangle DEF, these will also be the altitudes of triangle ABC. Since the perpendicular bisectors are concurrent then the altitudes of triangle ABC will also be concurrent at the same point.

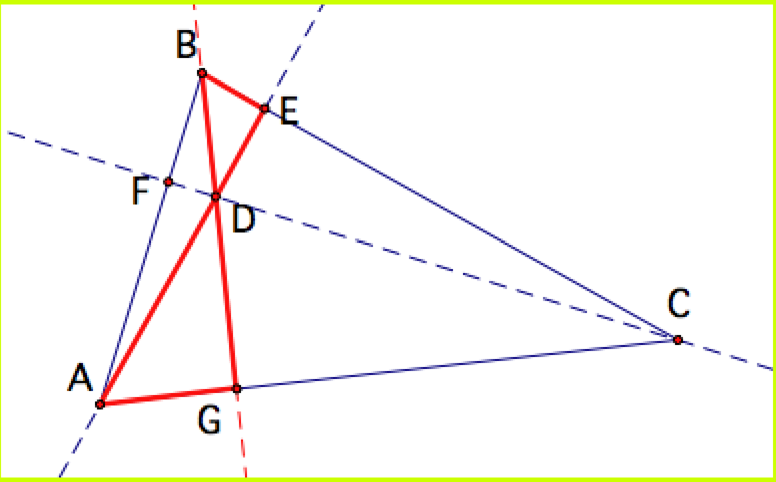


Therefore the altitudes of a triangle are concurrent.

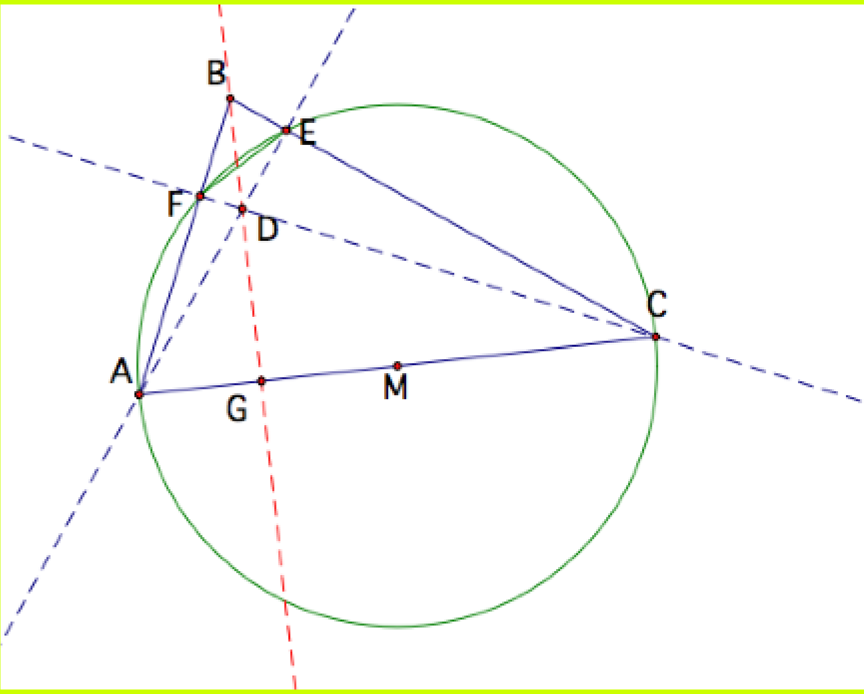
I will prove that the altitudes of a triangle are concurrent using another method . . . one involving circles. Again, consider triangle ABC.



Strategy . . .
Given altitudes CF and AE that intersect at point D, I will show that the line through BD is perpendicular to segment AC implying that angle DGA is a right angle.
I will do this by proving that triangles AGD and BED are similar. Angles BDE and ADG are vertical angles and so are of equal measure. So I am half-way to proving the triangles are similar. I only need to prove angles DAG and DBE are of equal measure. This will prove that the triangles are similar and will result in angle BED = angle AGD. Since angle BED is a right angle because of altitude AE then angle AGD will also be a right angle.

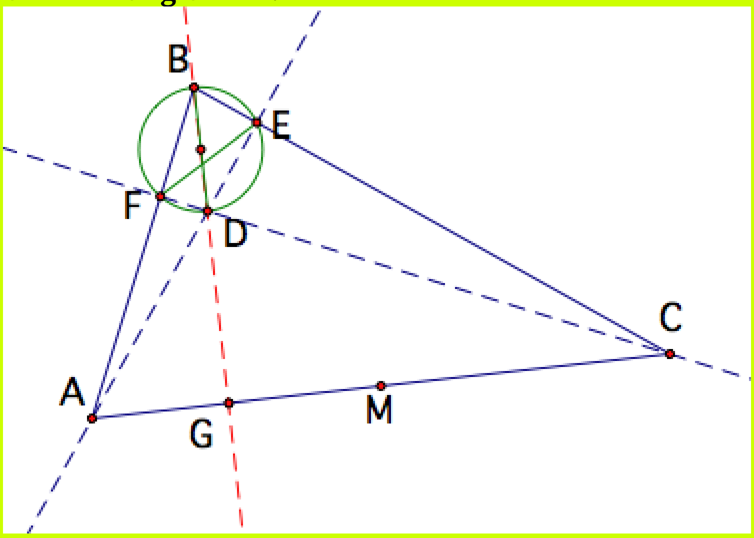


As stated above AE is an altitude so angles AEB and AEC are right angles. This means that triangle AEC is a right triangle with hypotenuse AC. There exists a circle whose diameter is AC and point E on the circle. (This can be proved using the Angle Arc Theorem.)
To find this circle, I will find the midpoint of AC, point M, and construct a circle of radius MC. Also I will construct segment FE. This will create two angles, angle EFC and angle EAC, that both subtend the same arc, arc EC. Therefore, by the Arc Angle Theorem, angle EFC = angle EAC.



Again, using the same idea as above, since angle FEB is a right angle then triangle BED is a right triangle with hypotenuse BD. So there exists a circle whose diameter is BD and point E on the circle. Use the same process as earlier to find this circle. So with segment EF, then angle EFD and angle EBD both subtend arc ED.

Therefore angle EBD = angle EFD.



Since angle EFC is the same as angle EFD then angle EBD = angle EAC. Since angle EAC is the same as angle DAG then angle EBD = angle DAG. Again since angle BDE = angle ADG (since they are vertical angles) then triangles AGD and BED are similar by AA similarity. Therefore angle DGA is a right angle and so BG is perpendicular to AC. Thus BG is an altitude of triangle ABC and the altitudes are concurrent.

There are some ideas inherent in the proofs that were not discussed here. These include the Arc Angle Theorem and why the circumcenter of a triangle is the orthocenter of its medial triangle. These would be neat ideas to explore and prove!

7.2 Slope of a Straight line Determine the slope of a line.

- Derive the formula to find the slope of a line passing through two points.
- Find out the condition that two straight lines with given slopes may be:
  I. Parallel to each other,
- Perpendicular to each other.

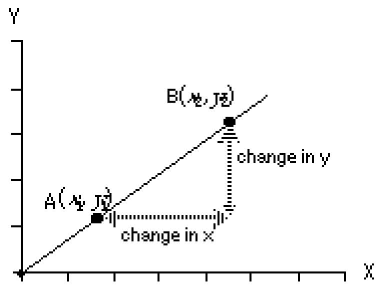
Calculating the Slope

To calculate the slope of a line you need only two points from that line, ($x1$, $y1$) and ($x2$, $y2$).

The equation used to calculate the slope from two points is:

On a graph, this can be represented as:

$$slope = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$



There are three steps in calculating the slope of a straight line when you are not given its equation.

1. **Step One:** Identify two points on the line.
2. **Step Two:** Select one to be ($x1$, $y1$) and the other to be ($x2$, $y2$).
3. **Step Three:** Use the slope equation to calculate slope.

Take a moment to work through an example where we are given two points.

Example

Let's say that points (15, 8) and (10, 7) are on a straight line. What is the slope of this line?

1. **Step One:** Identify two points on the line.

   In this example we are given two points, (15, 8) and (10, 7), on a straight line.

2. **Step Two:** Select one to be ($x1$, $y1$) and the other to be ($x2$, $y2$).

   It doesn't matter which we choose, so let's take (15, 8) to be ($x2$, $y2$). Let's take the point (10, 7) to be the point ($x1$, $y1$).

3. **Step Three:** Use the equation to calculate slope.

   Once we've completed step 2, we are ready to calculate the slope using the equation for a slope:

$$slope = \frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{(8-7)}{(15-10)} = \frac{1}{5}$$

   We said that it really doesn't matter which point we choose as ($x1$, $y1$) and the which to be ($x2$, $y2$). Let's show that this is true. Take the same two points (15, 8) and (10, 7), but this time we will calculate the slope using (15, 8) as ($x1$, $y1$) and (10, 7) as the point ($x2$, $y2$). Then substitute these into the equation for slope:

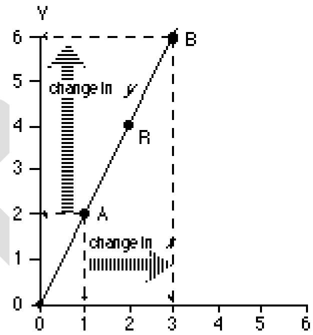
$$slope = \frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{(7-8)}{(10-15)} = \frac{-1}{-5} = \frac{1}{5}$$

We get the same answer as before!

Often you will not be given the two points, but will need to identify two points from a graph. In this case the process is the same, the first step being to identify the points from the graph. Below is an example that begins with a graph.

Example

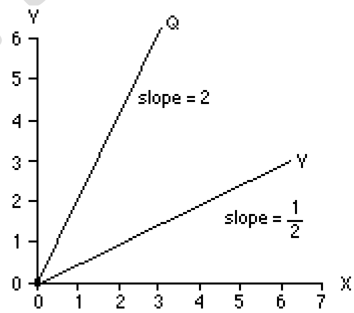
What is the slope of the line given in the graph?
The slope of this line is 2.



[detailed solution to example]

Now, take a moment to compare the two lines which are on the same graph.

Notice that the line with the greater slope is the steeper of the two. The greater the slope, the steeper the line. Keep in mind, you can only make this comparison between lines on a graph if: (1) both lines are drawn on the same set of axes, or (2) lines are drawn on different graphs (i.e., using different sets of axes) where both graphs have the same scale.



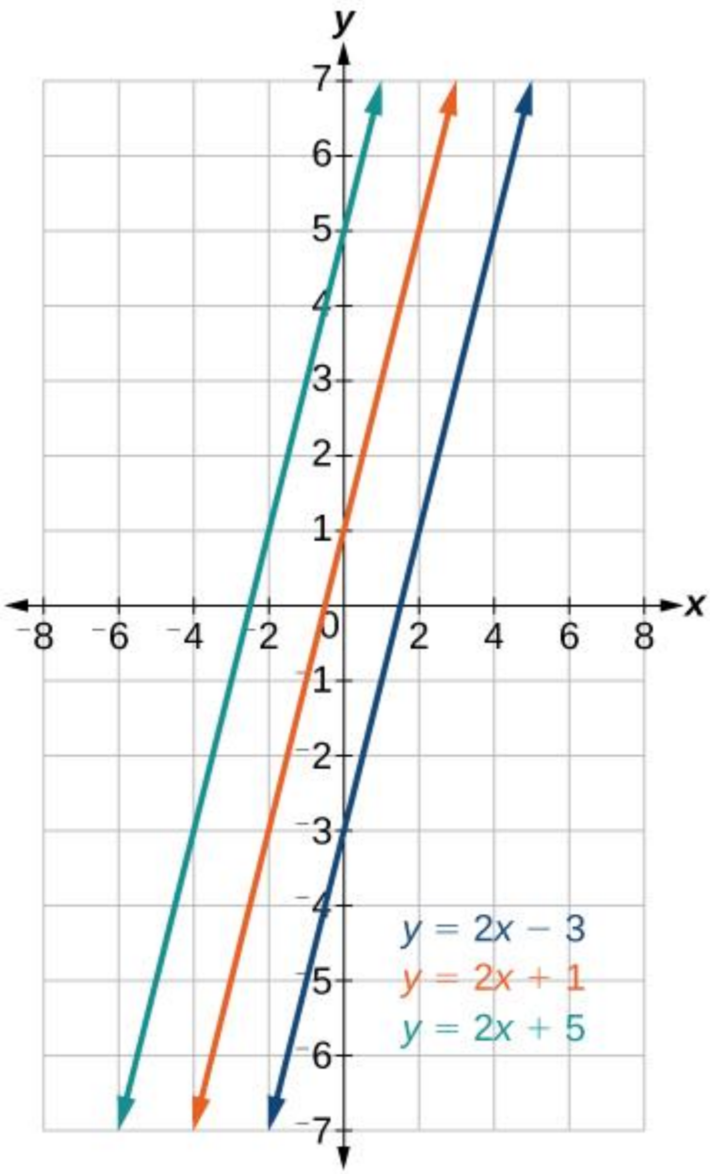
You are now ready to try a practice problem. If you have already completed the first practice problem for this unit you may wish to try the additional practice.

Parallel and Perpendicular Lines

LEARNING OUTCOMES

- Determine whether two lines are parallel or perpendicular.
- Find the equations of parallel and perpendicular lines.
- Write the equations of lines that are parallel or perpendicular to a given line.

Parallel lines have the same slope and different *y*-intercepts. Lines that are **parallel** to each other will never intersect. For example, the figure below shows the graphs of various lines with the same slope, $m=2m=2$.

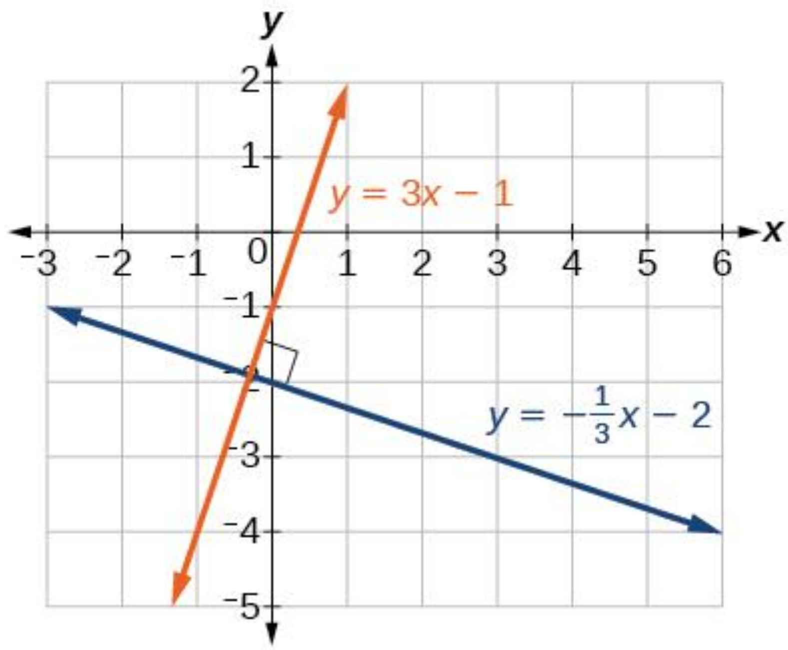


Parallel lines have slopes that are the same.

All of the lines shown in the graph are parallel because they have the same slope and different *y*-intercepts.

Lines that are **perpendicular** intersect to form a $90^\circ 90^\circ$ angle. The slope of one line is the negative **reciprocal** of the other. We can show that two lines are perpendicular if the product of the two slopes is $-1:m1\cdot m2=-1-1:m1\cdot m2=-1$. For example, the figure below shows the graph of two perpendicular lines. One line has a slope of 3; the other line has a slope of $-13-13$.

$$m1\cdot m2=-1\ 3\cdot(-13)=-1\ m1\cdot m2=-1\ 3\cdot(-13)=-1$$



Perpendicular lines have slopes that are negative reciprocals of each other.

EXAMPLE: GRAPHING TWO EQUATIONS, AND DETERMINING WHETHER THE LINES ARE PARALLEL, PERPENDICULAR, OR NEITHER

Graph the equations of the given lines and state whether they are parallel, perpendicular, or neither: 3y=−4x+33y=−4x+3 and 3x−4y=83x−4y=8.

Show Solution

TRY IT

Graph the two lines and determine whether they are parallel, perpendicular, or neither: 2y−x=102y−x=10 and 2y=x+42y=x+4.

Show Solution

If we know the equation of a line, we can use what we know about slope to write the equation of a line that is either parallel or perpendicular to the given line.

Writing Equations of Parallel Lines

Suppose we are given the following equation:

y=3x+1y=3x+1

We know that the slope of the line formed by the function is 3. We also know that the *y*-intercept is (0, 1). Any other line with a slope of 3 will be parallel to y=3x+1y=3x+1. So all of the following lines will be parallel to the given line.

y=3x+6y=3x+1y=3x+23y=3x+6y=3x+1y=3x+23

Suppose then we want to write the equation of a line that is parallel to $y=3x+6y=3x+6$ and passes through the point (1, 7). We already know that the slope is 3. We just need to determine which value for *b* will give the correct line. We can begin with point-slope form of a line and then rewrite it in slope-intercept form.

$$y-y1=m(x-x1)y-7=3(x-1)y-7=3x-3y=3x+4y-y1=m(x-x1)y-7=3(x-1)y-7=3x-3y=3x+4$$

So $y=3x+4y=3x+4$ is parallel to $y=3x+1y=3x+1$ and passes through the point (1, 7).

HOW TO: GIVEN THE EQUATION OF A LINE, WRITE THE EQUATION OF A LINE PARALLEL TO THE GIVEN LINE THAT PASSES THROUGH A GIVEN POINT

1. Find the slope of the line.
2. Substitute the given values into either point-slope form or slope-intercept form.
3. Simplify.

EXAMPLE: FINDING A LINE PARALLEL TO A GIVEN LINE

Find a line parallel to the graph of $y=3x+6y=3x+6$ that passes through the point (3, 0).
Show Solution

TRY IT

Check your work with an online graphing tool.

Launch Desmos Calculator

Writing Equations of Perpendicular Lines

We can use a very similar process to write the equation for a line perpendicular to a given line. Instead of using the same slope, however, we use the negative reciprocal of the given slope. Suppose we are given the following line:

$$y=2x+4y=2x+4$$

The slope of the line is 2 and its negative reciprocal is $-12-12$. Any function with a slope of $-12-12$ will be perpendicular to $y=2x+4y=2x+4$. So all of the following lines will be perpendicular to $y=2x+4$.

$$y=-12x+4y=-12x+2y=-12x-12y=-12x+4y=-12x+2y=-12x-12$$

As before, we can narrow down our choices for a particular perpendicular line if we know that it passes through a given point. Suppose then we want to write the equation of a line that is perpendicular to $y=2x+4y=2x+4$ and passes through the point (4, 0). We already know that the slope is $-12-12$. Now we can use the point to find the *y*-intercept by substituting the given values into slope-intercept form and solving for *b*.

$$y=mx+b0=-12(4)+b0=-2+b2=bb=2y=mx+b0=-12(4)+b0=-2+b2=bb=2$$

The equation for the function with a slope of $-12-12$ and a *y*-intercept of 2 is $y=-12x+2y=-12x+2$.

So $y=-12x+2y=-12x+2$ is perpendicular to $y=2x+4y=2x+4$ and passes through the point (4, 0). Be aware that perpendicular lines may not look obviously perpendicular on a graphing calculator unless we use the square zoom feature.

Q & A

A horizontal line has a slope of zero and a vertical line has an undefined slope. These two lines are perpendicular, but the product of their slopes is not –1. Doesn't this fact contradict the definition of perpendicular lines?
No. For two perpendicular linear functions, the product of their slopes is –1. As you will learn later, a vertical line is not a function so the definition is not contradicted.

HOW TO: GIVEN THE EQUATION OF A LINE, WRITE THE EQUATION OF A LINE PERPENDICULAR TO THE GIVEN LINE THAT PASSES THROUGH A GIVEN POINT

1. Find the slope of the given line.
2. Determine the negative reciprocal of the slope.
3. Substitute the slope and point into either point-slope form or slope-intercept form.
4. Simplify.

EXAMPLE: FINDING THE EQUATION OF A PERPENDICULAR LINE

Find the equation of a line perpendicular to $y=3x+3y=3x+3$ that passes through the point (3, 0).
Show Solution

TRY IT

Given the line $y=2x-4y=2x-4$, write an equation for the line passing through (0, 0) that is

1. parallel to *y*
2. perpendicular to *y*

Show Solution

Check your work with an online graphing tool.

Launch Desmos Calculator

HOW TO: GIVEN TWO POINTS ON A LINE, WRITE THE EQUATION OF A PERPENDICULAR LINE THAT PASSES THROUGH A THIRD POINT

1. Determine the slope of the line passing through the points.
2. Find the negative reciprocal of the slope.
3. Substitute the slope and point into either point-slope form or slope-intercept form.
4. Simplify.

EXAMPLE: FINDING THE EQUATION OF A PERPENDICULAR LINE

A line passes through the points (–2, 6) and (4, 5). Find the equation of a perpendicular line that passes through the point (4, 5).

Show Solution

TRY IT

A line passes through the points (–2, –15) and (2, –3). Find the equation of a perpendicular line that passes through the point (6, 4).

Show Solution

Writing the Equations of Lines Parallel or Perpendicular to a Given Line

As we have learned, determining whether two lines are parallel or perpendicular is a matter of finding the slopes. To write the equation of a line parallel or perpendicular to another line, we follow the same principles as we do for finding the equation of any line. After finding the slope, use **point-slope form** to write the equation of the new line.

EXAMPLE: WRITING THE EQUATION OF A LINE PARALLEL TO A GIVEN LINE

Write the equation of line parallel to a $5x+3y=15x+3y=1$ which passes through the point $(3,5)(3,5)$.
Show Solution

TRY IT

Find the equation of the line parallel to $5x=7+y5x=7+y$ which passes through the point $(-1,-2)(-1,-2)$.
Show Solution

EXAMPLE: FINDING THE EQUATION OF A PERPENDICULAR LINE

Find the equation of the line perpendicular to $5x-3y+4=05x-3y+4=0$ which goes through the point $(-4,1)(-4,1)$.
Show Solution

7.3 Equation of a straight line parallel to Co-ordinate Axes Find the equation of a straight line parallel to

I. Y-axis at a distance α from it.

- X-axis at a distance b from it

Equation of a Line Parallel to x-axis

Search

To find the equation of x-axis and of a line parallel to x-axis:

Let AB be a straight line parallel to x-axis at a distance b units from it. Then, clearly, all points on the line AB have the same ordinate b. Thus, AB can be considered as the locus of a point at a distance b from x-axis and all points on the line AB satisfy the condition y = b.

Equation of a Line Parallel to x-axis

Thus, if P(x, y) is any point on AB, then y = b.

Hence, the equation of a straight line parallel to x-axis at a distance b from it is y = b.

The equation of x-axis is y = 0, since, x-axis is a parallel to itself at a distance 0 from it.

Or

Let P (x,y) be any point on the x-axis. Then clearly, for all position of P we shall the same ordinate 0 or, y = 0.

Therefore, the equation of x-axis is y = 0.

If a straight line is parallel and below to x-axis at a distance b, then its equation is y = -b.

Solved examples to find the equation of x-axis and equation of a line parallel to x-axis:

1. Find the equation of a straight line parallel to x-axis at a distance of 10 units above the x-axis.

Solution:

We know that the equation of a straight line parallel to x-axis at a distance b from it is y = b.

Therefore, the equation of a straight line parallel to x-axis at a distance 10 units above the x-axis is y = 10.

2. Find the equation of a straight line parallel to x-axis at a distance of 7 units below the x-axis.

Solution:

We know that If a straight line is parallel and below to x-axis at a distance b, then its equation is y = -b.

Therefore, the equation of a straight line parallel to x-axis at a distance 7 units below the x-axis is y = -7.

● **The Straight Line**

- Straight Line
- Slope of a Straight Line
- Slope of a Line through Two Given Points
- Collinearity of Three Points
- Equation of a Line Parallel to x-axis
- Equation of a Line Parallel to y-axis
- Slope-intercept Form
- Point-slope Form
- Straight line in Two-point Form
- Straight Line in Intercept Form
- Straight Line in Normal Form
- General Form into Slope-intercept Form
- General Form into Intercept Form
- General Form into Normal Form
- Point of Intersection of Two Lines
- Concurrency of Three Lines
- Angle between Two Straight Lines
- Condition of Parallelism of Lines
- Equation of a Line Parallel to a Line
- Condition of Perpendicularity of Two Lines
- Equation of a Line Perpendicular to a Line
- Identical Straight Lines
- Position of a Point Relative to a Line
- Distance of a Point from a Straight Line

- [Equations of the Bisectors of the Angles between Two Straight Lines]
- [Bisector of the Angle which Contains the Origin]
- [Straight Line Formulae]
- [Problems on Straight Lines]
- [Word Problems on Straight Lines]
- [Problems on Slope and Intercept]

7.4 Standard Form of Equation of a Straight Line Determine intercepts of a straight line.

- Derive equation of a straight line in
  I. Slope-intercepts form,
  II. Point-slope form,
  III. Two-point form,
  IV. Intercepts form ,
  V. Symmetric form,
  VI. Normal form.
- Verify that a linear equation in two variables represents a straight line.
- Reduce the general form of the equation of a straight line to the other standard forms.

Intercepts of Linear Equations

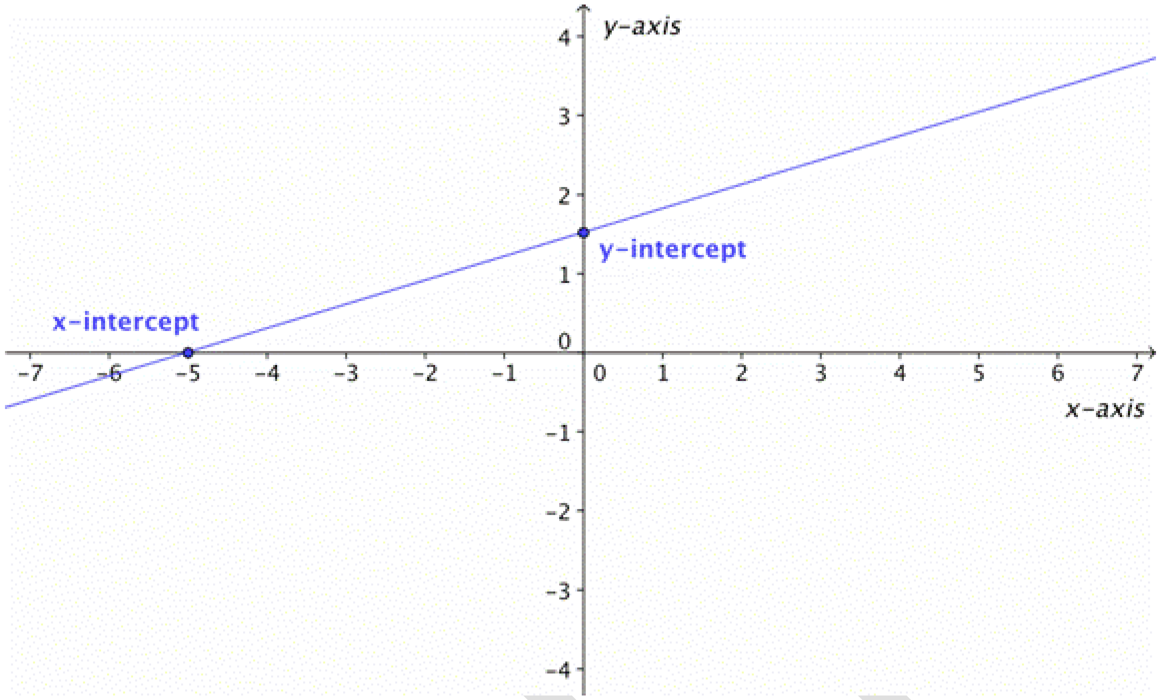
Learning Objective(s)

- Calculate the intercepts of a line.
- Use the intercepts to plot a line.

Introduction

The **intercepts** of a line are the points where the line intercepts, or crosses, the horizontal and vertical axes.

The straight line on the graph below intercepts the two coordinate axes. The point where the line crosses the *x*-axis is called the **[*x*-intercept]**. The **[*y*-intercept]** is the point where the line crosses the *y*-axis.



Notice that the *y*-intercept occurs where x = 0, and the *x*-intercept occurs where y = 0.

Calculating Intercepts

We can use the characteristics of intercepts to quickly calculate them from the equation of a line. Just see how easy it is, as we find the *x*- and *y*-intercepts for the line $3y + 2x = 6$.

To find the *y*-intercept, we substitute 0 for *x* in the equation, because we know that every point on the *y*-axis has an *x*-coordinate of 0. Once we do that, we can solve to find the value of *y*. When we make *x* = 0, the equation becomes $3y + 2(0) = 6$, which works out to *y* = 2. So when *x* = 0, *y* = 2. The coordinates of the *y*-intercept are (0, 2).

Example			
Problem	**3*y* + 2*x***	**=**	**6**
	3*y* + 2(0)	=	6

	$3y$	=	6
	$\frac{3y}{3}$	=	$\frac{6}{3}$
Answer	y	=	2

Now we'll follow the same steps to find the *x*-intercept. We'll let $y = 0$ in the equation, and solve for *x*. When $y = 0$, the equation for the line becomes $3(0) + 2x = 6$, and that works out to $x = 3$. When $y = 0$, $x = 3$. The coordinates of the *x*-intercept are (3, 0).

Example			
Problem	$\mathbf{3y + 2x}$	**=**	**6**
	$3(0) + 2x$	=	6
	$2x$	=	6
	$\frac{2x}{2}$	=	$\frac{6}{2}$
Answer	x	=	3

See, I told you that it would be easy.

What is the *y*-intercept of a line with the equation $y = 5x - 4$?

A) $\left(\frac{4}{5}, 0\right)$

B) (-4, 0)

C) (0, -4)

D) (5, -4)

Show/Hide Answer

Using Intercepts to Graph Lines

Knowing the intercepts of a line is a useful thing. For one thing, it makes it easy to draw the graph of a line—we just have to plot the intercepts and then draw a line through them. Let's do it with the equation $3y + 2x = 6$. We figured out that the intercepts of the line this equation represents are (0, 2) and (3, 0). That's all we need to know:

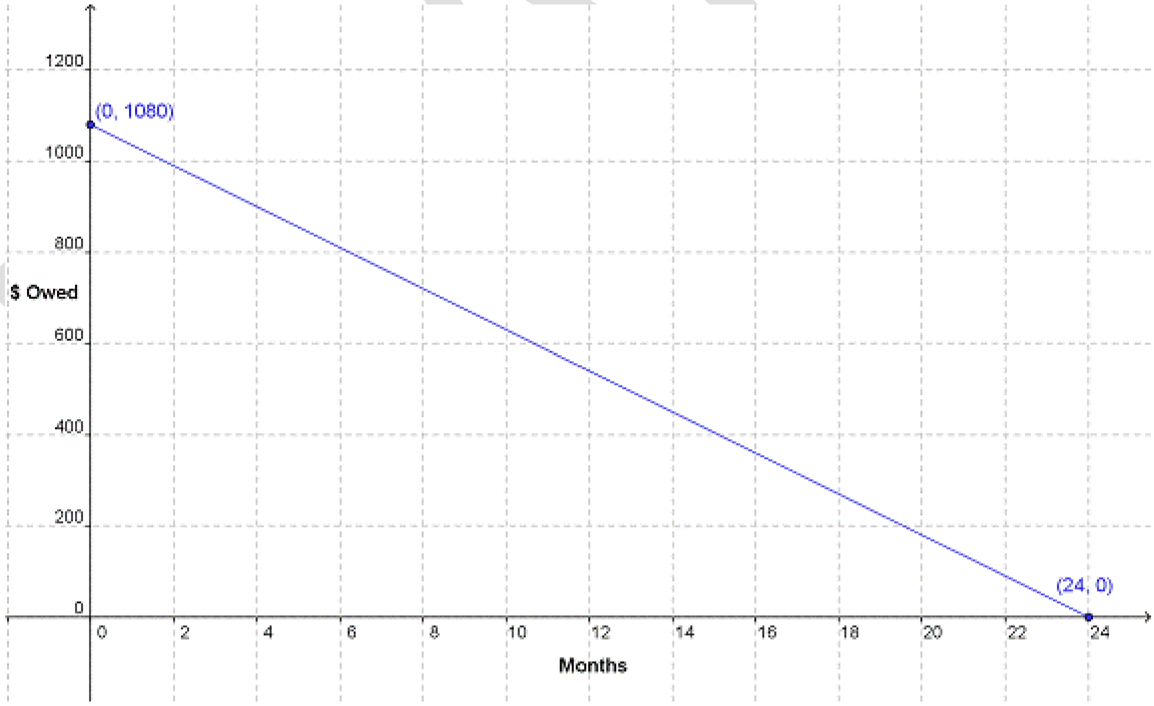
And there we have the line.

Intercepts and Problem-solving

Intercepts are also valuable tools for predicting or tracking a process. At each intercept, one of the two quantities being plotted reaches zero. That means that the intercepts of a line can be used to mark the beginning and the end of a process.

Imagine a student named Morgan who is buying a laptop for $1,080 to use for school. Morgan is going to use the computer store's finance plan to make this purchase—she'll pay $45 per month for 24 months.

She wants to know how much she will still owe after each month of the plan. She can keep track of her debt by making a graph. The *x*-axis will be the number of months and the *y*-axis will represent the amount of money she still owes on the finance plan. Morgan knows two points in her pay-off schedule. The day she buys the computer, she'll be at 0 months passed and $1,080 owed. The day she pays it off completely, she'll be at 24 months passed and $0 owed. With these two points, she can draw a line, running from the *y*-intercept at (0, 1080) to the *x*-intercept at (24, 0).



Morgan can now use this graph to figure out how much money she still owes after any number of months.

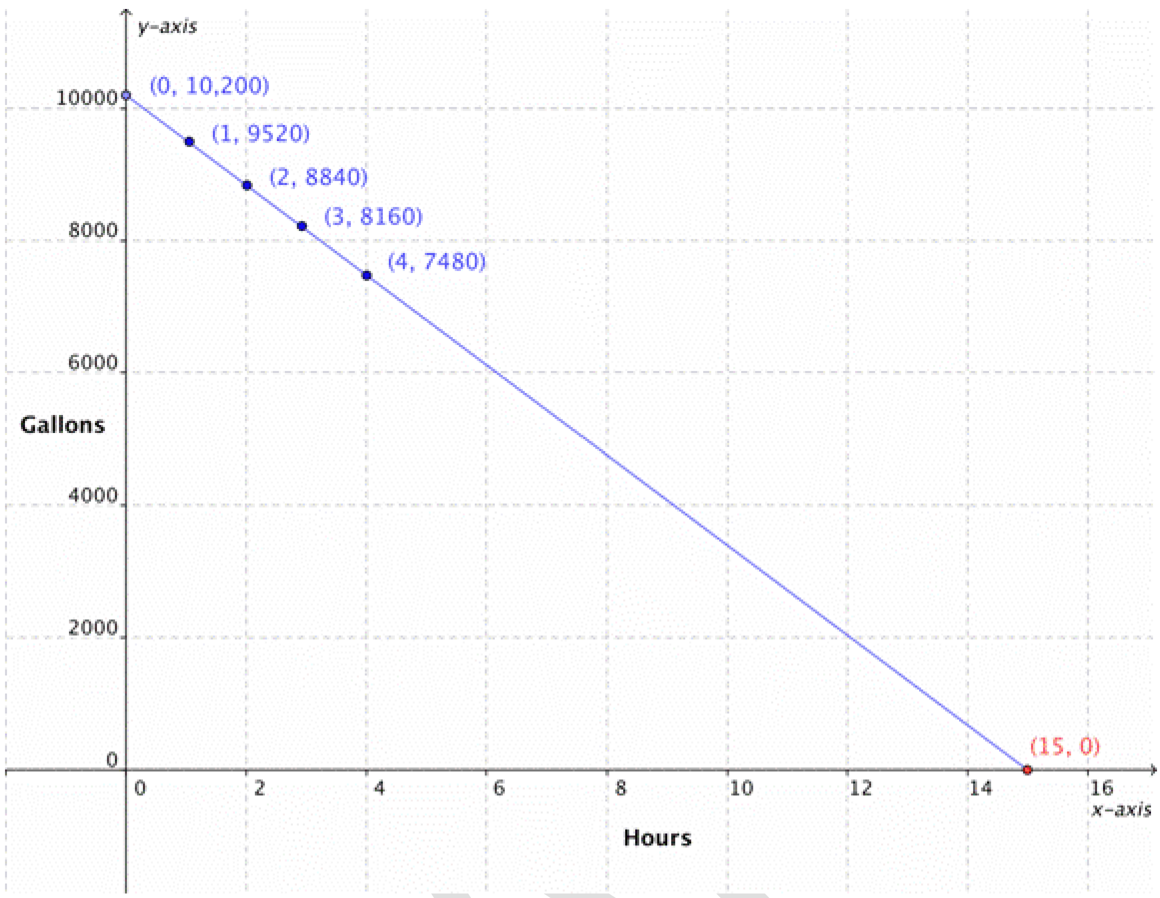
Let’s look at another situation involving intercepts, this time when we know only one intercept and want to find the other. Joe is a lifeguard at the local swimming pool. It’s the end of the summer, and the pool is being drained. Joe has to wait by the pool until it’s completely empty, so no one falls in and drowns. How can poor Joe figure out how long that’s going to take?

If Joe has taken an algebra course, he’s got it made. The pool contains 10,200 gallons of water. It drains at a rate of 680 gallons per hour. Joe can use that information to make a table of how much water will be left in the pool hour by hour.

x, Time (hours)	**y, Volume of Water (gallons)**
0	10,200
1	9,520
2	8,840
3	8,160
4	7,480

Once he’s calculated a few data points, Joe can use a graph and intercepts as a short-cut to find out how long it will be until the pool is dry. Joe’s starting point is the *y*-intercept, where the pool is full at 10,200 gallons and the elapsed time is 0. Next, he plots the volume of the pool at 1, 2, 3, and finally 4 hours.

Now all Joe needs to do is connect the points with a line, and then extend the line until it meets the *x*-axis.



The line intercepts the *x*-axis when x = 15. So now Joe knows—the pool will take 15 hours to drain completely. It's going to be a long day.

Summary

We've now seen the usefulness of the intercepts of a line. When we know where a line crosses the *x*- and *y*-axes, we can easily produce the graph or the equation for that line. When we know one of the intercepts and the slope of a line, we can find the beginning or predict the end of a process.

Different Forms of Line

Table of Content

- [Related Resources]

Straight line is an extremely important and a vast topic of the mathematics syllabus of IIT JEE. Students are advised to go into the intricacies of topics in order to master them. In this section, we shall discuss the equation of straight line in various forms and illustrate the concepts along with certain examples as well. Aspirants must try to grasp these concepts in order to excel in competitions like the IIT JEE.

What do you mean by the equation of a Straight Line?

A certain kind of relation between the variables x and y which is met by the coordinates of every point lying on a line is termed as the equation of a straight line.

Remark:

- Any linear equation in two variables x and y always represents a straight line.
- Equation of a line which is parallel to x-axis or perpendicular to y-axis at a distance 'c' from it is given as y = c.
- Equation of a line which is parallel to y-axis or perpendicular to x-axis at a distance 'c' from it is given as x = c.

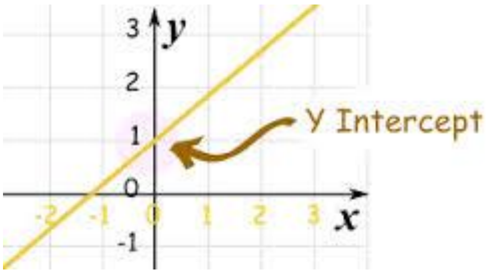
?We now discuss the various forms of lines one by one:

General Form

The most general equation of a straight line is ax + by + c = 0, where a, b and c are any real numbers such that both a and b can't be zero simultaneously.

Slope Intercept Form

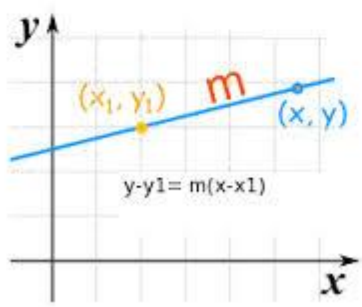
If we have a straight line whose slope is 'm' and which makes an intercept 'c' on the y-axis then its equation is given by y = mx + c.



As shown in the figure above, the y-intercept here is c.

Slope One Point Form

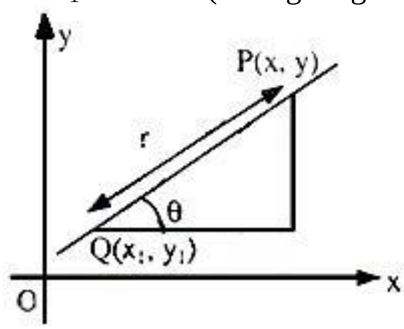
The equation of a straight line having slope as 'm' and which passes through the point (x_1, y_1) is given by $(y-y_1) = m(x-x_1)$.



Parametric Form

Consider line PQ with coordinates P(x, y) and $Q(x_1, y_1)$. Then Co-ordinates of any points P(x, y) are

$x = x_1 + r \cos \theta$ (see figure given below)



$y = y_1 + r \sin \theta$

Equation of the line is obtained as follows:-

$\Rightarrow x-x_1/\cos \theta = y-y_1/\sin \theta = r$

This is parametric form of the equation of a straight line.

Note:

1. Note that 'r' is positive if the point (x, y) lies on the right of (x_1, y_1) and negative if the point (x, y) lies on the left of (x_1, y_1).

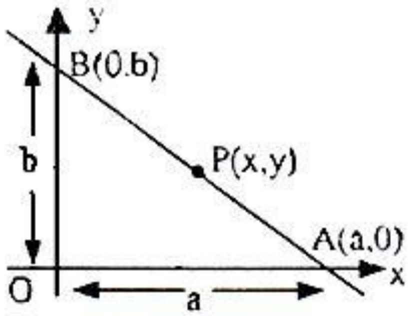
Two Point Form

If we have two given points say (x_1, y_1) and (x_2, y_2), then the line passing through them is given by the formula

$(y-y_1) = m(x-x_1)$ or $(y-y_1) = (y_2-y_1)/(x_2-x_1).\ (x-x_1)$.

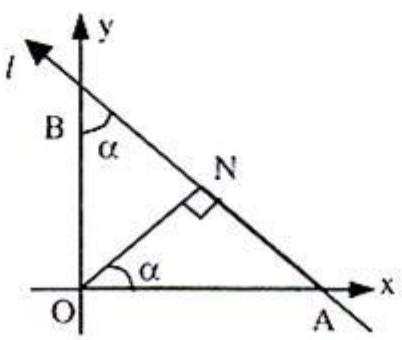
Intercept Form

If intercepts of a line on x and y-axis are known then equation of the line can also be found in two-intercept form. Intercepts are OA and OB on x and y-axis respectively, where A(a, 0) and B(0, b) are two points through which line is passing. Treating it as a special case of two-point form, one can write a unique equation of the line as
$y–0/x–a = 0–b/a–0$, where P(x, y) is any point on the line (figure given below)
If we are given the intercepts of a line on the x and y axis respectively as ‘a’ and ‘b’ then the equation of the straight line is given by $x/a + y/b = 1$.
$\Rightarrow y/b = –x/a + 1$.
$\Rightarrow x/a + y/b = 1$.
This is intercept from of the equation of a straight line.



Normal Form

$x \cos \alpha + y \sin \alpha = a$ is the equation of the straight line in perpendicular form, where ‘p’ is the length of the perpendicular from the origin O on the line and this perpendicular makes an angle α with the positive direction of x-axis.



Consider line l as shown in figure given above
ON ⊥ l and |ON| = p
We have in triangle ONA
$OA = p/\cos \alpha$
A and B are intercept points of line l. So intercepts on x and y-axes are $p/\cos \alpha$ and $p/\sin \alpha$ respectively. So equation of line l will be (both intercept form)
$x \cos \alpha/p + y \sin \alpha/p = 1$
$\Rightarrow x \cos \alpha + y \sin \alpha = p$

This is the equation of a straight line in normal form, where p is perpendicular distance of the line from origin.

Note:

P is always measured away from the origin and is always positive in value, αa is a positive angle less than 360^o measured from the positive direction OX of the x-axis to the normal from the origin to the line.

Reduction of General form of Equations into Standard Forms

The general form of equation is ax + by + c = 0, then we can derive the other forms as:

Slope-intercept form: This form is given by y = -a/b x -c/b, where slope = – a/b and intercept is -c/b

Intercept form: This form is given as x/(-c/a) + y/(-c/b) = 1, here x-intercept is (-c/a) and y-intercept is (-c/b).

Normal Form: In order to change the general form of equation to the normal form, first take 'c' to the right side and try to make it positive. Then the whole equation should be divided by $\sqrt{a^2 + b^2}$ like

$-ax/\sqrt{(a^2 + b^2)} – by/\sqrt{(a^2 + b^2)} = c/\sqrt{(a^2 + b^2)}$, where

$\cos \alpha = -a/\sqrt{(a^2 + b^2)}$, $\sin \alpha = -b/\sqrt{(a^2 + b^2)}$ and $p = c/\sqrt{(a^2 + b^2)}$.

Now, we discuss some of illustrations based on this concept:

Illustration:

The ends of a rod of length l move on two mutually perpendicular lines. Find the locus of the point on the rod, which divides it in the ratio 2 : 1.

Solution:

Suppose the two perpendicular lines are x = 0 and y = 0 and let the end of the rod lie at the point (0, a) and (b, 0).

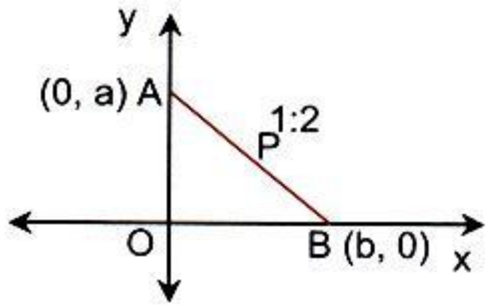
The point P has coordinates given by

h = b+2.0/2+1, k = 2.a+1.0/2+1

⇒ a = 3k/2, b = 3h.

Also $l^2 = a^2 + b^2$.

⇒ $l^2 = (3k/2)^2 + (3h)^2$.



Thus the required locus is

$x^2 + y^2/4 = l^2/9$, which represents an ellipse.

Similarly, take the ratio AP : BP as 2 : 1 and proceed. We get the result as $y^2 + x^2/4 = l^2/9$.

Illustration:

Find the locus of the point of intersection of the lines x cos α + y sin α = a and x sin α – y cos α = b. where α is a variable.

Solution:

Let P(h, k) be the point of intersection of the given lines.

Then h cos α + k sin α = a. … (1)

h sin α – k cos α = b. … (2)

Here a is a variable. So we have to eliminate a.

Squaring and adding (1) and (2),

We get, $(h\cos\alpha + k\sin\alpha)^2 + (h\sin\alpha - k\cos\alpha)^2 = a^2 + b^2$

$\Rightarrow h^2 + k^2 = a^2 + b^2$.

Hence locus of (h, k) is $x^2 + y^2 = a^2 + b^2$.

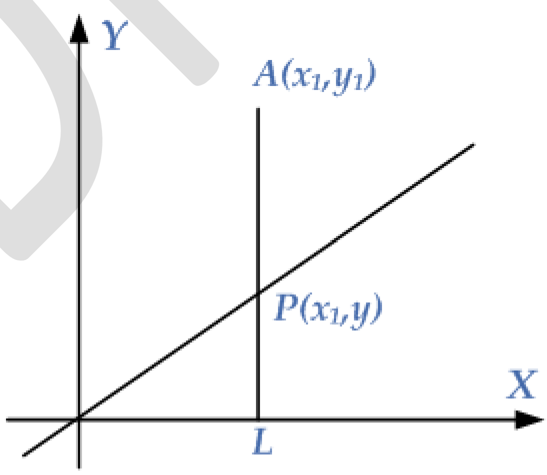
7.5 Distance of a point from a line Recognize a point with respect to position of a line.

- Find the perpendicular distance from a point to the given straight line.

Position of Point with Respect to a Line

The general equation or standard equation of a straight line is given by

ax+by+c=0 – – – (i)ax+by+c=0 – – – (i)



Let A(x1,y1)A(x1,y1) be a point which does not lie on the line (i). Now draw a perpendicular line from AA to the X-axis through the point PP as shown in the given diagram. It is clear from the diagram that x1x1 is the abscissa of the point PP. If yy is the ordinate of PP, then (x1,y)(x1,y) are the coordinates of PP. Since PP lies on the line (i), it must satisfy the equation of the line, i.e.

ax1+by+c=0⇒by=–ax1–c⇒y=–ax1+cb – – – (ii)ax1+by+c=0⇒by=–ax1–c⇒y=–ax1+cb – – – (ii)

Next we consider the difference y1–yy1–y, i.e.

y1–y=y1–(–ax1+cb)=y1+ax1+cb=by1+ax1+cb⇒y1–y=ax1+by1+cb – – – (iii)y1–y=y1–(–ax1+cb)=y1+ax1+cb=by1+ax1+cb⇒y1–y=ax1+by1+cb – – – (iii)

(a) If the point AA is above the line, then y1–y>0y1–y>0. From equation (iii), we note that y1–y>0y1–y>0 only if ax1+by1+cb>0ax1+by1+cb>0 .
But ax1+by1+cb>0ax1+by1+cb>0 if ax1+by1+c>0ax1+by1+c>0 and b>0b>0 or ax1+by1+cb>0ax1+by1+cb>0 if ax1+by1+c<0ax1+by1+c<0 and b<0b<0.
We conclude that the point A(x1,y1)A(x1,y1) is above the line ax+by+c=0ax+by+c=0 if
(i) ax1+by1+c>0ax1+by1+c>0 and b>0b>0
(ii) ax1+by1+c<0ax1+by1+c<0 and b<0b<0
(b) If the point AA is below the line, then y1–y<0y1–y<0. From equation (iii), we note that y1–y<0y1–y<0 only if ax1+by1+cb<0ax1+by1+cb<0 .
But ax1+by1+cb<0ax1+by1+cb<0 if ax1+by1+c<0ax1+by1+c<0 and b>0b>0 or ax1+by1+cb<0ax1+by1+cb<0 if ax1+by1+c>0ax1+by1+c>0 and b<0b<0.
We conclude that the point A(x1,y1)A(x1,y1) is below the line ax+by+c=0ax+by+c=0 if
(i) ax1+by1+c<0ax1+by1+c<0 and b>0b>0
(ii) ax1+by1+c>0ax1+by1+c>0 and b<0b<0
NOTE: The point A(x1,y1)A(x1,y1) will be on the line ax+by+c=0ax+by+c=0 if ax1+by1+c=0ax1+by1+c=0.
Example: Determine whether the point (1,3)(1,3) lies below or above the line 3x–2y+1=03x–2y+1=0.
Comparing the given line 3x–2y+1=03x–2y+1=0 with the general equation of line ax+by+c=0ax+by+c=0, we have a=3a=3, b=–2<0b=–2<0 and c=1c=1.
Since (1,3)(1,3) is the given point, then x1=1,y1=3x1=1,y1=3.
Now:

ax1+by1+c=3(1)+(–2)(3)+1=–2<0ax1+by1+c=3(1)+(–2)(3)+1=–2<0

Since ax1+by1+c<0ax1+by1+c<0 and b<0b<0, the given point lies above the line.

Read more: https://www.emathzone.com/tutorials/geometry/position-of-point-with-respect-to-line.html#ixzz6CON7HJHc

Perpendicular Distance from a Point to a Line

Later, on this page...

Example using perpendicular distance formula

(BTW - we don't really need to say 'perpendicular' because the distance from a point to a line always means the shortest distance.)

This is a great problem because it uses all these things that we have learned so far:

- distance formula
- slope of parallel and perpendicular lines
- rectangular coordinates
- different forms of the straight line
- solving simultaneous equations

Need Graph Paper?

Graph paper icon
Download graph paper

The distance from a point (m, n) to the line $Ax + By + C = 0$ is given by:

$$d=\frac{\left|Am+Bn+C\right|}{\sqrt{A^{2}+B^{2}}}$$

There are some examples using this formula following the proof.

Continues below ⇩

Proof of the Perpendicular Distance Formula

Let's start with the line $Ax + By + C = 0$ and label it DE. It has slope $-\frac{A}{B}$.

xyAx + By + C = 0DEOpen image in a new page

Line DE with slope $-A/B$.

We have a point P with coordinates (m, n). We wish to find the perpendicular distance from the point P to the line DE (that is, distance PQ).

xyP (m, n)QDEOpen image in a new page

Perpendicular to straight line.

We now do a trick to make things easier for ourselves (the algebra is really horrible otherwise). We construct a line parallel to DE through (m, n). This line will also have slope $-\frac{A}{B}$, since it is parallel to DE. We will call this line FG.

xyP (m, n)QDEFGOpen image in a new page

Perpendicular and parallel constructions.

Now we construct another line parallel to PQ passing through the origin.

This line will have slope \displaystyle\frac{B}{{A}}*AB*, because it is perpendicular to DE.

Let's call it line RS. We extend it to the origin \displaystyle{\left({0},{0}\right)}(0,0).

We will find the distance RS, which I hope you agree is equal to the distance PQ that we wanted at the start.

xyP (m, n)QDEFGRS(0, 0)Open image in a new page

Perpendicular through origin.

Since FG passes through (*m*, *n*) and has slope \displaystyle-\frac{A}{{B}}−*BA*, its equation is \displaystyle{y}-{n}=-\frac{A}{{B}}{\left({x}-{m}\right)}*y*−*n*=−*BA* (*x*−*m*) or

\displaystyle{y}=\frac{{-{A}{x}+{A}{m}+{B}{n}}}{{B}}*y*=*B*−*Ax*+*Am*+*Bn*.

Line RS has equation \displaystyle{y}=\frac{B}{{A}}{x}.*y*=*ABx*.

Line FG intersects with line RS when

\displaystyle\frac{B}{{A}}{x}=\frac{{-{A}{x}+{A}{m}+{B}{n}}}{{B}}*AB* *x*=*B*−*Ax*+*Am*+*Bn*

Solving this gives us

\displaystyle{x}=\frac{{{A}{\left({A}{m}+{B}{n}\right)}}}{{{A}^{2}+{B}^{2}}}*x*= *A*2+*B*2*A*(*Am*+*Bn*)

So after substituting this back into \displaystyle{y}=\frac{B}{{A}}{x},*y*=*ABx*, we find that point R is

\displaystyle{\left(\frac{{{A}{\left({A}{m}+{B}{n}\right)}}}{{{A}^{2}+{B}^{2}}},\frac{{{B}{\left({A}{m}+{B}{n}\right)}}}{{{A}^{2}+{B}^{2}}}\right)}(*A*2+*B*2*A*(*Am*+*Bn*),*A*2+*B*2*B*(*Am*+*Bn*))

Point S is the intersection of the lines \displaystyle{y}=\frac{B}{{A}}{x}*y*=*AB* *x* and *Ax* + *By* + *C* = 0, which can be written \displaystyle{y}=-\frac{{{A}{x}+{C}}}{{B}}*y*=−*BAx*+*C*.

This occurs when (that is, we are solving them simultaneously)

\displaystyle-\frac{{{A}{x}+{C}}}{{B}}=\frac{B}{{A}}{x}−*BAx*+*C*=*ABx*

Solving for *x* gives

\displaystyle{x}=\frac{{-{A}{C}}}{{{A}^{2}+{B}^{2}}}x=A2+B2−AC

Finding *y* by substituting back into

\displaystyle{y}=\frac{B}{{A}}{x}y=ABx

gives

\displaystyle{y}=\frac{B}{{A}}{\left(\frac{{-{A}{C}}}{{{A}^{2}+{B}^{2}}}\right)}=\frac{{-{B}{C}}}{{{A}^{2}+{B}^{2}}}y=AB(A2+B2−AC)=A2+B2−BC

So S is the point

\displaystyle{\left(\frac{{-{A}{C}}}{{{A}^{2}+{B}^{2}}},\frac{{-{B}{C}}}{{{A}^{2}+{B}^{2}}}\right)}(A2+B2−AC,A2+B2−BC)

The distance RS, using the distance formula,

\displaystyle{d}=\sqrt{{{\left({x}_{{2}}-{x}_{{1}}\right)}^{2}+{\left({y}_{{2}}-{y}_{{1}}\right)}^{2}}}d=(x2−x1)2+(y2−y1)2

is

\displaystyle{d}=\sqrt{{{\left(\frac{{-{A}{C}}}{{{A}^{2}+{B}^{2}}}-\frac{{{A}{\left({A}{m}+{B}{n}\right)}}}{{{A}^{2}+{B}^{2}}}\right)}^{2}+{\left(\frac{{-{B}{C}}}{{{A}^{2}+{B}^{2}}}-\frac{{{B}{\left({A}{m}+{B}{n}\right)}}}{{{A}^{2}+{B}^{2}}}\right)}^{2}}}d=(A2+B2−AC−A2+B2A(Am+Bn))2+(A2+B2−BC−A2+B2B(Am+Bn))2

\displaystyle=\sqrt{{\frac{{{\left\lbrace-{A}{\left({A}{m}+{B}{n}+{C}\right)}\right\rbrace}^{2}+{\left\lbrace-{B}{\left({A}{m}+{B}{n}+{C}\right)}\right\rbrace}^{2}}}{{\left({A}^{2}+{B}^{2}\right)}^{2}}}}=(A2+B2)2{−A(Am+Bn+C)}2+{−B(Am+Bn+C)}2

\displaystyle=\sqrt{{\frac{{{\left({A}^{2}+{B}^{2}\right)}{\left({A}{m}+{B}{n}+{C}\right)}^{2}}}{{\left({A}^{2}+{B}^{2}\right)}^{2}}}}=(A2+B2)2(A2+B2)(Am+Bn+C)2

\displaystyle=\sqrt{{\frac{{{\left({A}{m}+{B}{n}+{C}\right)}^{2}}}{{{A}^{2}+{B}^{2}}}}}=A2+B2(Am+Bn+C)2

\displaystyle=\frac{{{\left|{A}{m}+{B}{n}+{C}\right|}}}{{\sqrt{{{A}^{2}+{B}^{2}}}}}=A2+B2|Am+Bn+C|

The absolute value sign is necessary since distance must be a positive value, and certain combinations of *A*, *m* , *B*, *n* and *C* can produce a negative number in the numerator.

So the distance from the point (*m*, *n*) to the line *Ax* + *By* + *C* = 0 is given by:

\displaystyle{d}=\frac{{{\left|{A}{m}+{B}{n}+{C}\right|}}}{{\sqrt{{{A}^{2}+{B}^{2}}}}}*d*=*A*2+*B*2|*Am*+*Bn*+*C*|

Example 1

Find the perpendicular distance from the point (5, 6) to the line −2*x* + 3*y* + 4 = 0, using the formula we just found.

Answer

Example 2

Find the distance from the point \displaystyle{\left(-{3},{7}\right)}(−3,7) to the line

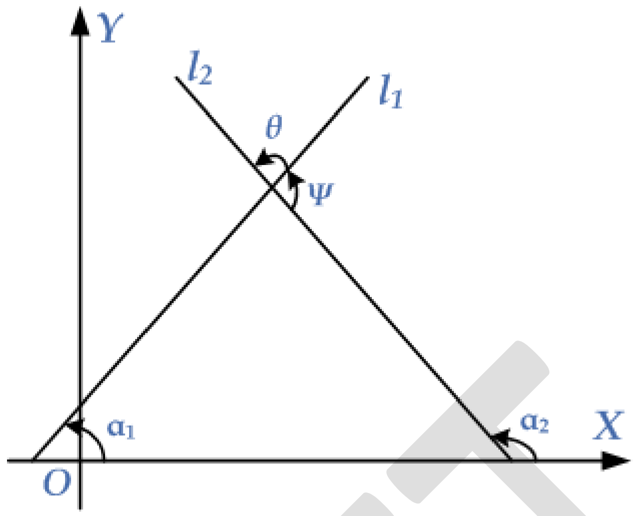
\displaystyle{y}=\frac{6}{{5}}{x}+{2}*y*=56*x*+2

7.6 Angle Between Lines Show that the angle between two coplanar intersecting straight lines.

- Determine the equation of family of lines passing through the point of intersection of two given lines
- Calculate angles of the triangle when the slopes of the sides are given.

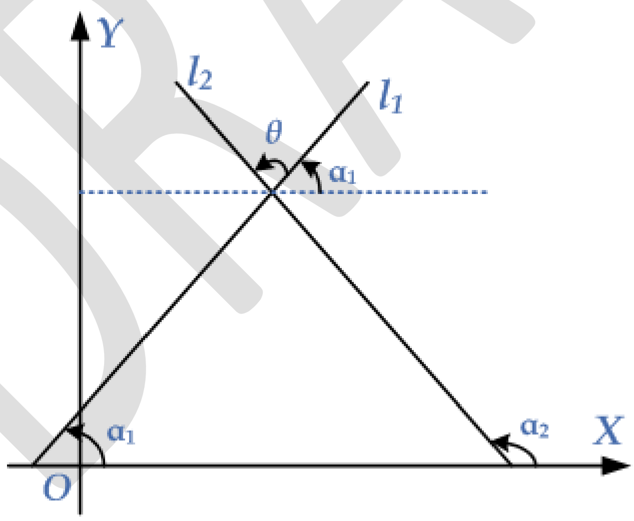
Angle of Intersection of Two Lines

Let l1l1 and l2l2 be two coplanar and non-parallel lines with inclination α1α1 and α2α2 respectively, as shown in the given diagram. The angle of intersection of lines l1l1 and l2l2 is the angle θθ through which line l1l1 is rotated counter-clockwise about the point of intersection so that it coincides with l2l2.



The angle θθ is the angle of the intersection of lines l1l1 and l2l2 measured from l2l2 to l1l1. The angle ψψ is also the angle of intersection of lines l1l1 and l2l2 measured from l2l2 to l1l1. If the lines are not perpendicular, then one angle between them is an acute angle.

<u>Theorem</u>: The angle θθ of the intersection of two non-vertical lines l1l1 and l2l2 from l2l2 to l1l1 is given by tanθ=m2–m11+m1m2tanθ=m2–m11+m1m2, where m1m1 and m2m2 are the slopes of lines l1l1 and l2l2 respectively.



Proof: Let l1l1 and l2l2 be two coplanar and non-parallel lines with inclination α1α1 and α2α2 respectively, as shown in the given diagram. It is clear from the diagram that

α1+θ=α2⇒θ=α2–α1⇒tanθ=tan(α2–α1)⇒tanθ=tanα2–tanα11+tanα1tanα2 – – – (i)α1+θ=α2⇒θ=α2–α1⇒tanθ=tan(α2–α1)⇒tanθ=tanα2–tanα11+tanα1tanα2 – – – (i)

Since α1α1 and α2α2 are the inclination of lines l1l1 and l2l2 respectively, their slopes are m1=tanα1m1=tanα1, m2=tanα2m2=tanα2. Putting these values of tanα1tanα1, tanα2tanα2 in equation (i), we have

⇒tanθ=m2–m11+m1m2⇒tanθ=m2–m11+m1m2

If the lines l1l1 and l2l2 are perpendicular, then θ=90∘θ=90∘, and using above formula, we have

1+m1m2=m2–m1tanθ⇒1+m1m2=m2–m1tan90∘⇒1+m1m2=m2–m1∞⇒1+m1m2=01+m1m2=m2–m1tanθ⇒1+m1m2=m2–m1tan90∘⇒1+m1m2=m2–m1∞⇒1+m1m2=0

m1m2=–1m1m2=–1

This is the condition for two lines to be perpendicular

Read more: https://www.emathzone.com/tutorials/geometry/angle-of-intersection-of-two-lines.html#ixzz6CONWDnOw

In a right angled triangle, the cosine of an angle is:

The length of the adjacent side divided by the length of the hypotenuse.

The abbreviation is cos

cos(θ) = adjacent / hypotenuse

Well Done!Another

cos θ =

Sine, Cosine, Tangent

7.7 Concurrency of Straight Lines Show the condition of concurrency of three straight lines.

- Find the equation of median, altitude and right bisector of a triangle:
- Show that: Three right bisectors; Three medians; Three altitudes, of a triangle are concurrent.

Altitudes, Medians and Angle Bisectors of a Triangle

The altitudes, medians and angle bisectors of a Triangle are defined and problems along with their solutions are presented.

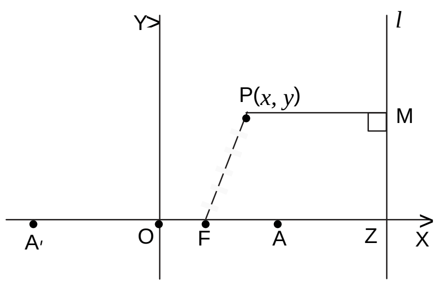
4.4 ELLIPSE

4.4.1 To find the equation of an ellipse in the standard form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \text{ where } b^2 = a^2(1-e^2).$$

Let F be a focus, *l* be a directrix and Z be the foot of perpendicular from F to the line *l*. Take FZ as *x*-axis with positive direction from F to Z. Divide the segment FZ internally and externally in the ratio *e* : 1 at the points A, A′ so that

F A = *e* .AZ …(*i*) and A′F = *e* .A′Z …(*ii*)



As *e*< 1 and A′F = *e* .A′Z

⇒ A′ is closer to F than to Z ⇒ A′ lies towards the left of F.

Let O be the mid-point of A′A, take O as origin,

then the line through O and perpendicular to FZ *Fig. 4.11.* becomes *y*-axis (shown in figure 4.11).

Let A′A = 2 *a* (*a*> 0), then A′O = OA = *a* so the points A, A′ are (*a*, 0), (– *a*, 0) respectively.

The relations (*i*) and (*ii*) can be written as

OA – OF = *e* (OZ – OA) …(*iii*)

and AO′ + OF = *e* (A′O + OZ)

i.e. OA + OF = *e* (OA + OZ) …(*iv*)(□ □ A′O = OA)

Adding (*iii*) and (*iv*), we get

$$2 \,.\, OA = 2\,e\,.\,OZ \Rightarrow a = e\,.\,OZ \Rightarrow OZ = \frac{a}{e}.$$

Subtracting (*iii*) from (*iv*), we get

$$2\,.\,OF = 2\,e.\,OA \Rightarrow OF = a\,e.$$

∴ The co-ordinate of focus F are $(ae, 0)$ and the equation of the directrix is

$$x = \frac{a}{e} \text{ i.e. } x - \frac{a}{e} = 0.$$

Let P (x, y) be any point in the plane of the line l and the point F, and | MP | be the perpendicular distance from P to the line l, then P lies on ***ellipse*** iff

$$|FP| = e\,|MP| \qquad (0 < e < 1)$$

$$\Leftrightarrow \sqrt{(x - ae)^2 + y^2} = e.\frac{\left|x - \frac{a}{e}\right|}{1}$$

$$\Leftrightarrow \sqrt{(x - ae)^2 + y^2} = |\,e\,x - a\,|$$

$$\Leftrightarrow x^2 + a^2 e^2 - 2a\,e\,x + y^2 = e^2 x^2 + a^2 - 2\,a\,e\,x$$

$$\Leftrightarrow (1 - e^2)\,x^2 + y^2 = a^2\,(1 - e^2)$$

$$\Leftrightarrow \frac{x^2}{a^2} + \frac{y^2}{a^2(1 - e^2)} = 1 \qquad (\because 0 < e < 1 \Rightarrow a^2(1 - e^2) \neq 0)$$

$$\Leftrightarrow \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \text{ where } b^2 = a^2\,(1 - e^2).$$

Find the equation of the hyperbola satisfying the following conditions :

(i) Vertices (± 7, 0), $e = \frac{4}{3}$ (*ii*) Vertices (0, ± 5), foci (0, ± 8)

(*iii*) Foci (0, ± 4), length of transverse axis 6.

5. The focus of a parabola is (1, 5) and its directrix is $x + y + 2 = 0$. Find the equation of the parabola, its vertex and length of latus-rectum.

6. Find the equation of the ellipse with axes along x-axis and y-axis, which passes through the points (4, 3) and (6, 2).

7. Show that the equation $5x^2 + 30x + 2y + 59 = 0$ represents a parabola. Find its vertex, focus, length of latus-rectum and equations of directrix and axis of the parabola.

8. Find the equation of the parabola with its axis parallel to y-axis and passing through the points (4, 5), (– 2, 11) and (–4, 21).

9. Identify the curves represented by the following equations :

(i) $2x^2 - 10xy + 12y^2 + 10x - 16y - 3 = 0$

ANSWERS

(ii) $9x^2 - 24xy + 16y^2 - 6x + 8y - 5 = 0$ (*iii*) $6x^2 - 5xy - 6y^2 + 14x + 6y - 4 = 0$.

10. Show that the line $y = x + \sqrt{\frac{5}{6}}$ touches the ellipse $2x^2 + 3y^2 = 1$. Also find the co-ordinates of the point of contact.

11. Determine k so that the line $2x + y + k = 0$ may touch the hyperbola $3x^2 - y^2 = 3$.

12. Find the equation of tangents to the ellipse $4x^2 + 5y^2 = 20$ which are perpendicular to the line $3x + 2y - 5 = 0$.

EXERCISE 4.1

1. (*i*) Symmetric about y-axis only (*ii*) Symmetric about x-axis only
(*iii*) Symmetric about x-axis, y-axis and about origin (*iv*)
Symmetric about x-axis, y-axis and about origin
(*v*) Symmetric about origin only (*vi*) Symmetric about origin only.

2. May not be true; for example, $xy - 1 = 0$.

EXERCISE 4.2

1. $x^2 - 2xy + y^2 - 4x + 2 = 0$. **2.** $x^2 - 2xy + y^2 - 8x + 20y + 46 = 0$.

3. $9x^2 + 36xy + 36y^2 + 276x - 138y - 161 = 0$.

4. $4x^2 - 12xy + 9y^2 - 58x - 30y + 64 = 0$.

$a^2x^2 - 2abxy + b^2y^2 - 2a^3x - 2b^3y + (a^4 - a^2b^2 + b^4) = 0$.

5. $9x^2 + 24xy + 16y^2 + 34x + 112x + 121 = 0$; $3x + 4y + 11 = 0$.

6. $16x^2 - 24xy + 9y^2 - 38x - 34y + 46 = 0$; $4x - 3y - 1 = 0$; $\left(\frac{7}{10}, \frac{3}{5}\right)$.

7. $(2\sqrt{3}, 0)$; $x + 2\sqrt{3} = 0$; 1.

8. (*i*) $\frac{2}{\sqrt{3}}$; $\left(\frac{\sqrt{3}}{2}, 0\right)$; $2x + \sqrt{3} = 0$ (*ii*) 4; $(-1, 0)$; $x - 1 = 0$ (*iii*) $\frac{4}{3}$; $\left(0, \frac{1}{3}\right)$; $3y + 1 = 0$ (*iv*) 12; $(0, -3)$

10. (2, 6). **11.** $\frac{4}{3}$; $\left(\frac{1}{3}, 0\right)$

12. (*i*) $y^2 = -16x$ (*ii*) $x^2 = -8y$ (*iii*) $2y^2 = 9x$.

13. $16x^2 - 24xy + 9y^2 - 444x - 292y + 424 = 0$.

14. $x^2 + 2xy + y^2 + 32x - 32y + 128 = 0$.

15. $(9, -8)$; $25x^2 + 30xy + 9y^2 - 618x + 554y + 4929 = 0$.

9.2 Ellipse

9.2.1 Standard for form of Equation of an Ellipse

9.2.2 Equations of Tangent and Normal

- Define ellipse and its element (i.e. center, foci, vertices covertices, directories, major and minor axes, eccentricity, focal chord and latera recta).
- Explain that circle is a special case of an ellipse.
- Derive the standard form of equation of an ellipse and identify its elements.
- Find the equation to the ellipse with the following given element:
  I. Major and minor axes,
  II. Two points,

III. Foci, vertices or length of a latera reta,
IV. Foci ,minor axes or length of latus rectum.

- Transform a given equation to the standard form of equation of an ellipse, find its elements and draw the graph.
- Identify tangent and normal to an ellipse
- Find points of intersection of an ellipse with a line including the condition of tangency.
- Find the Equation of a tangent in slope form.
- Obtain the equation of an tangent and a normal to an ellipse at a point

Parabola focus & directrix review

Review your knowledge of the focus and directrix of parabolas.

Google ClassroomFacebookTwitter

Email

What are the focus and directrix of a parabola?

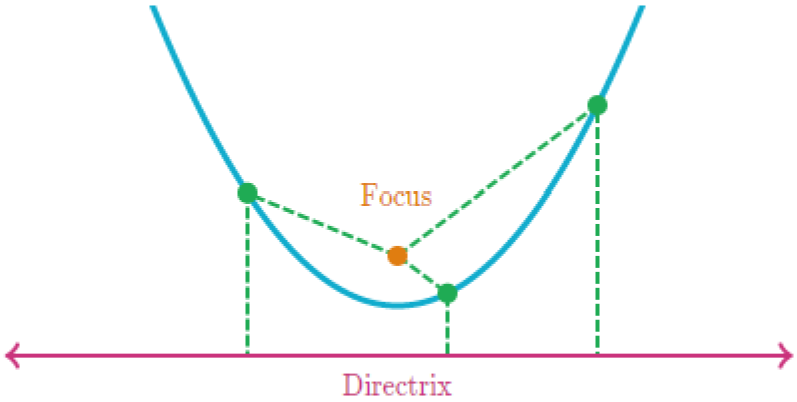
Parabolas are commonly known as the graphs of quadratic functions. They can also be viewed as the set of all points whose distance from a certain point (the **focus**) is equal to their distance from a certain line (the **directrix**).

\goldD{\text{Focus}}Focus\maroonD{\text{Directrix}}Directrix

Want to learn more about focus and directrix of a parabola? Check out this video.

Parabola equation from focus and directrix

Given the focus and the directrix of a parabola, we can find the parabola's equation. Consider, for example, the parabola whose focus is at (-2,5)(−2,5)left parenthesis, minus, 2, comma, 5, right parenthesis and directrix is y=3*y*=3y, equals, 3. We start by assuming a general point on the parabola (x,y)(x,y)left parenthesis, x, comma, y, right parenthesis.



Using the distance formula, we find that the distance between (x,y)(x,y)left parenthesis, x, comma, y, right parenthesis and the focus (-2,5)(−2,5)left parenthesis, minus, 2, comma, 5, right parenthesis is \sqrt{(x+2)^2+(y-5)^2}(x+2)2+(y−5)2square root of, left parenthesis, x, plus, 2, right parenthesis, squared, plus, left parenthesis, y, minus, 5, right parenthesis, squared, end square root, and the distance between (x,y)(x,y)left parenthesis, x, comma, y, right parenthesis and the directrix y=3y=3y, equals, 3 is \sqrt{(y-3)^2}(y−3)2square root of, left parenthesis, y, minus, 3, right parenthesis, squared, end square root. On the parabola, these distances are equal:

\begin{aligned} \sqrt{(y-3)^2} &= \sqrt{(x+2)^2+(y-5)^2} \\\\ (y-3)^2 &= (x+2)^2+(y-5)^2 \\\\ \blueD{y^2}-6y\goldD{+9} &= (x+2)^2\blueD{+y^2}\maroonD{-10y}+25 \\\\ -6y\maroonC{+10y}&=(x+2)^2+25\goldD{-9} \\\\ 4y&=(x+2)^2+16 \\\\ y&=\dfrac{(x+2)^2}{4}+4\end{aligned}(y−3)2(y−3)2y2−6y+9−6y+10y4yy=(x+2)2+(y−5)2 =(x+2)2+(y−5)2=(x+2)2+y2−10y+25=(x+2)2+25−9=(x+2)2+16=4(x+2)2+4

9.3 Hyperbola

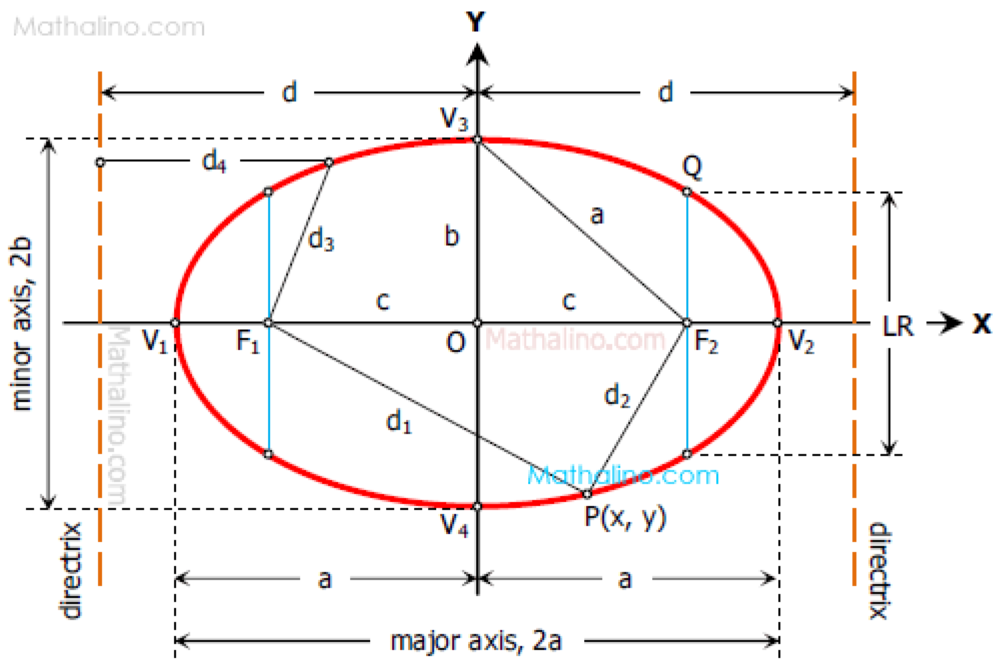
9.3.1

9.3.2 Standard Form of Equation of Hyperbola

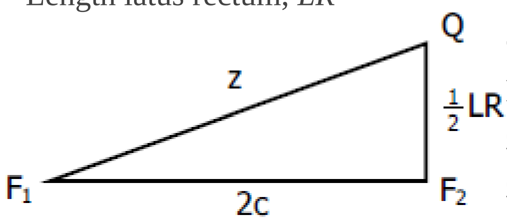
- Equation of Tangent and Normal
- Define hyperbola and its elements (i.e. center, foci vertices, directrices, transverse and conjugate axes, eccentricity focal chord and latera recta).
- Derive the standard form of equation of a hyperbola and identify its elements.
- Find the equation of a hyperbola with the following given elements:
  I. Transverse and conjugate axes with center at origin,
  II. Two points
  III. Eccentricity, latera recta and transverse axes,
  IV. Focus, eccentricity and center,
  V. Focus, center and directrix
- Convert a given equation to the standard form of equation of a hyperbola, find its element and sketch the graph.
- Identify tangent and normal to a hyperbola
- Find:
  I. Points of intersection of a hyperbola with a line including the condition of tangency
  II. The equation of a tangent in slope form
- Find the equation of a tangent and a normal to a hyperbola at a point

Elements of Ellipse

Elements of the ellipse are shown in the figure below.



1. Center (h, k). At the origin, (h, k) is (0, 0).
2. Semi-major axis = a and semi-minor axis = b.
3. Location of foci c, with respect to the center of ellipse. $c=\sqrt{a^2-b^2}$.
4. Length latus rectum, LR



Consider the right triangle F_1QF_2:
Based on the [definition of ellipse](#)
$z+\frac{1}{2}LR=2a$
$z=2a-\frac{1}{2}LR$
$z=\frac{4a-LR}{2}$

By Pythagorean Theorem
$(2c)^2+(\frac{1}{2}LR)^2=z^2$
$4c^2+\frac{1}{4}(LR)^2=(\frac{4a-LR}{2})^2$
$4c^2+\frac{1}{4}(LR)^2=\frac{(4a-LR)^2}{4}$
$16c^2+(LR)^2=(4a-LR)^2$
$16c^2+(LR)^2=16a^2-8a(LR)+(LR)^2$
$16c^2=16a^2-8a(LR)$
$8a(LR)=16a^2-16c^2$
$LR=\frac{16a^2-16c^2}{8a}$
$LR=\frac{16(a^2-c^2)}{8a}$

$$LR=\frac{2b^2}{a}$$

You can also find the same formula for the length of latus rectum of ellipse by using the definition of eccentricity.

5. Eccentricity, e
$e=\frac{\text{distance from focus to ellipse}}{\text{distance from ellipse to directrix}}$

From the figure of the ellipse above,
$e=\frac{d_3}{d_4}=\frac{a}{d}=\frac{a-c}{d-a}$
From
$\frac{a}{d}=\frac{a-c}{d-a}$

ad−a2=ad−cdad−a2=ad−cd
d=a2/cd=a2/c

Thus,
e=ad=aa2/ce=ad=aa2/c

e=ca<1.0e=ca<1.0

6. Location of directrix *d*, with respect to the center of ellipse.

   From the derivation of eccentricity,

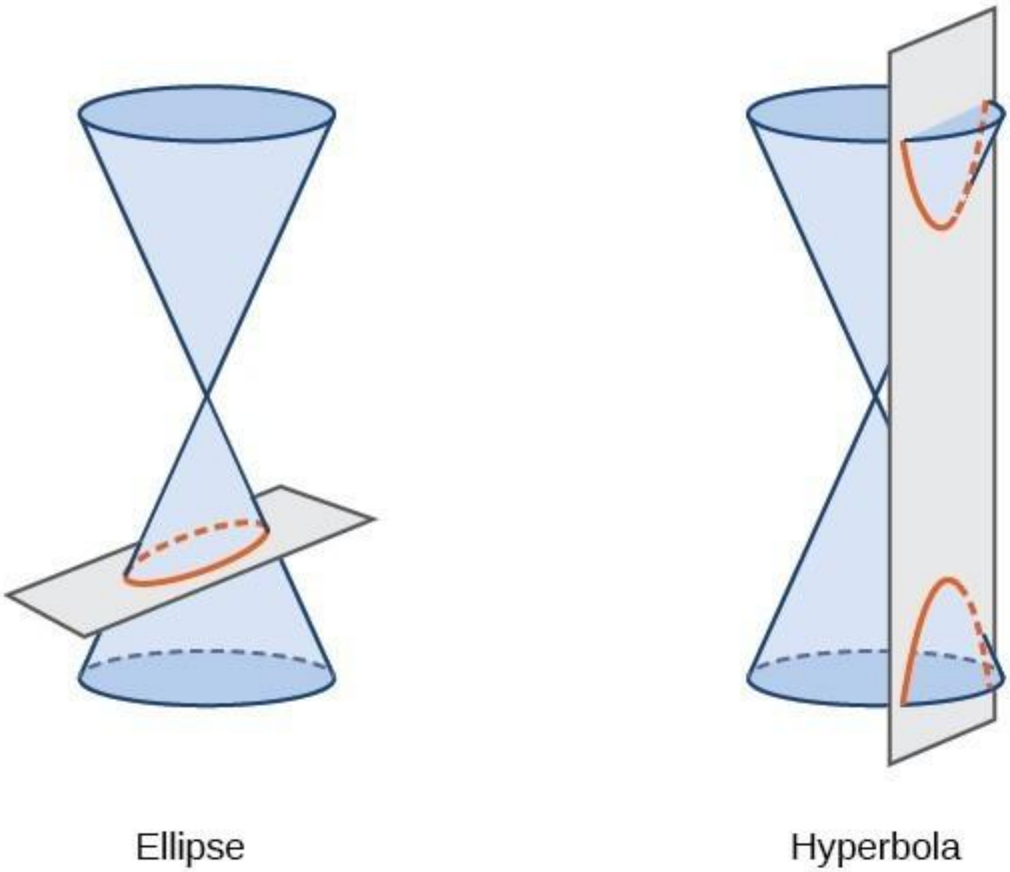
   d=ae or d=a2cd=ae or d=a2c

Equations of Ellipses

LEARNING OUTCOMES

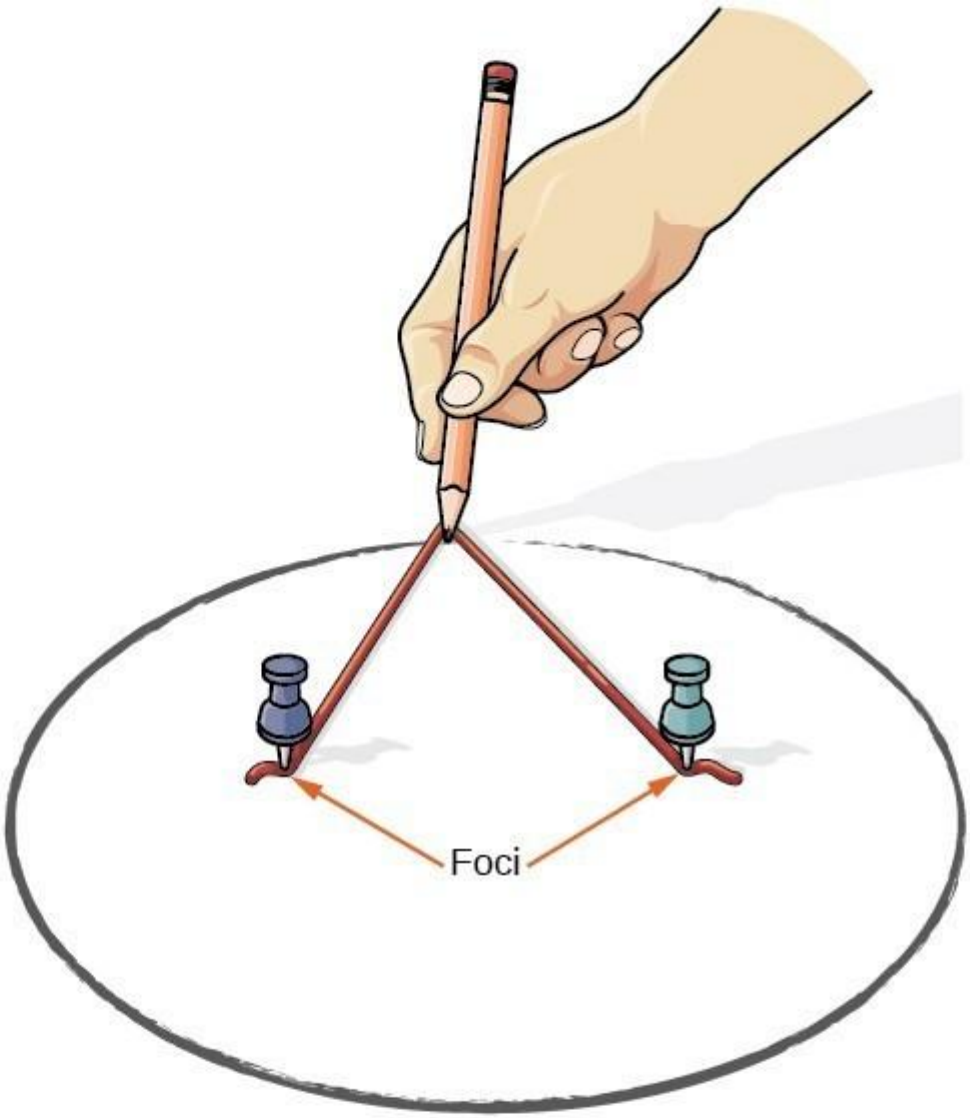
- Identify the foci, vertices, axes, and center of an ellipse.
- Write equations of ellipses centered at the origin.
- Write equations of ellipses not centered at the origin.

A conic section, or **conic**, is a shape resulting from intersecting a right circular cone with a plane. The angle at which the plane intersects the cone determines the shape.

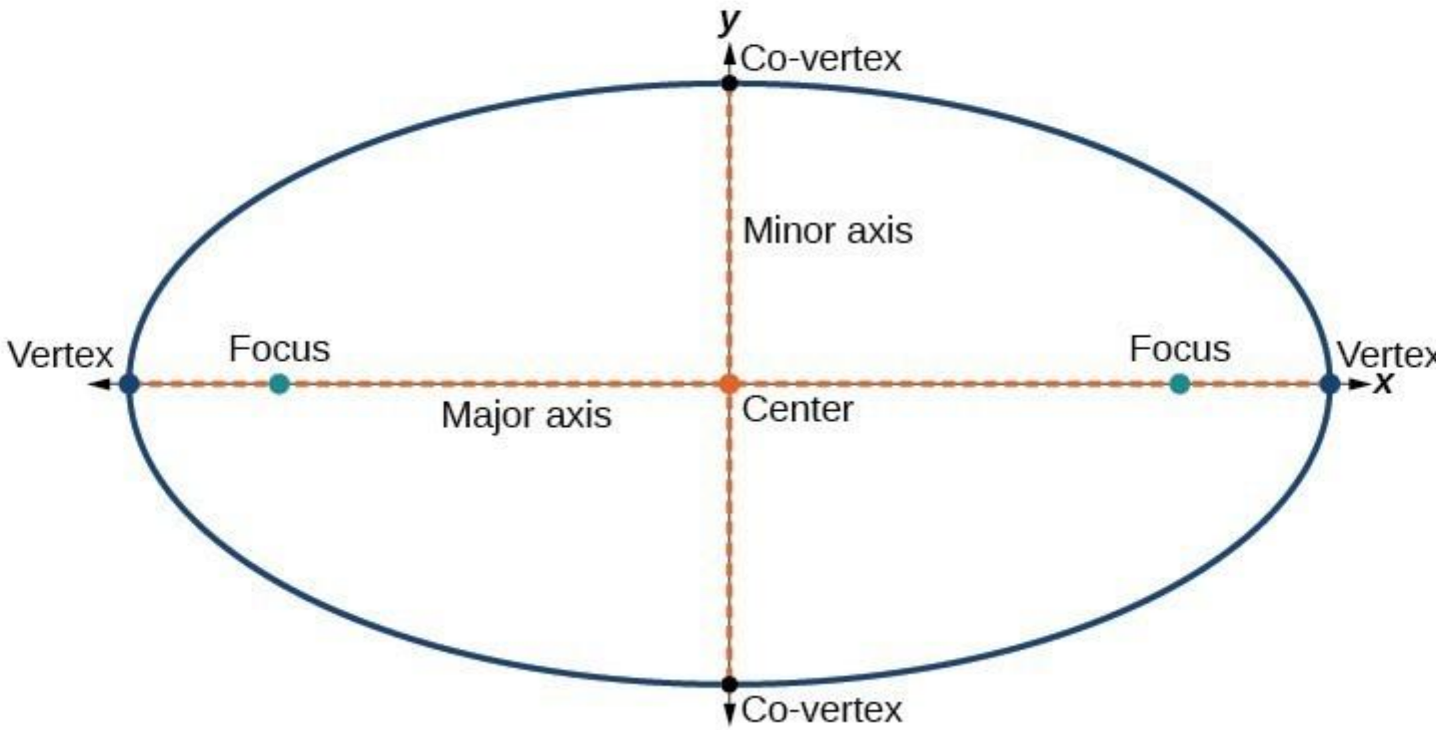


Conic sections can also be described by a set of points in the coordinate plane. Later in this chapter we will see that the graph of any quadratic equation in two variables is a conic section. The signs of the equations and the coefficients of the variable terms determine the shape. This section focuses on the four variations of the standard form of the equation for the ellipse. An **ellipse** is the set of all points (x,y)(x,y) in a plane such that the sum of their distances from two fixed points is a constant. Each fixed point is called a **focus** (plural: **foci**) of the ellipse.

We can draw an ellipse using a piece of cardboard, two thumbtacks, a pencil, and string. Place the thumbtacks in the cardboard to form the foci of the ellipse. Cut a piece of string longer than the distance between the two thumbtacks (the length of the string represents the constant in the definition). Tack each end of the string to the cardboard, and trace a curve with a pencil held taut against the string. The result is an ellipse.



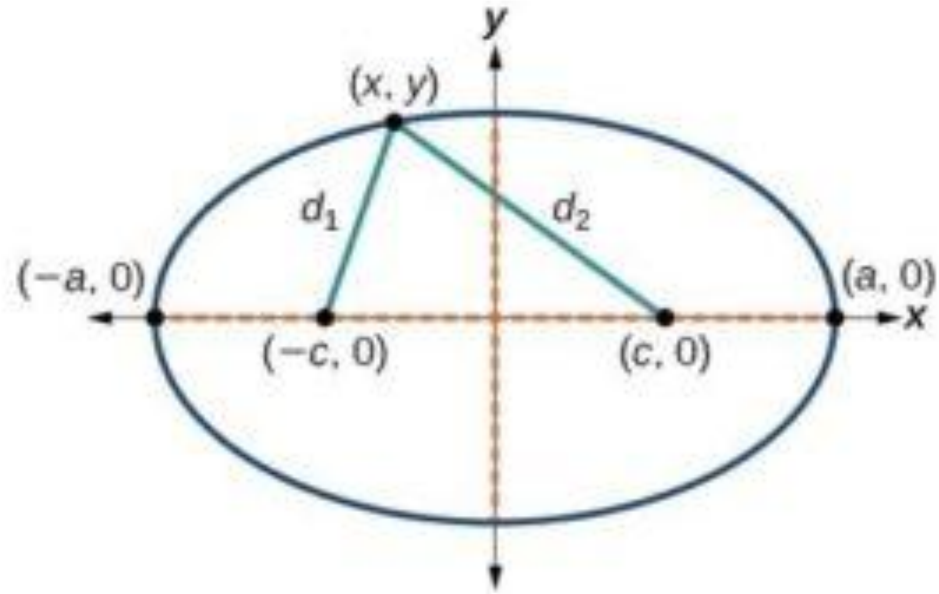
Every ellipse has two axes of symmetry. The longer axis is called the **major axis**, and the shorter axis is called the **minor axis**. Each endpoint of the major axis is the **vertex** of the ellipse (plural: **vertices**), and each endpoint of the minor axis is a **co-vertex** of the ellipse. The **center of an ellipse** is the midpoint of both the major and minor axes. The axes are perpendicular at the center. The foci always lie on the major axis, and the sum of the distances from the foci to any point on the ellipse (the constant sum) is greater than the distance between the foci.



In this section we restrict ellipses to those that are positioned vertically or horizontally in the coordinate plane. That is, the axes will either lie on or be parallel to the *x*– and *y*-axes. Later in the chapter, we will see ellipses that are rotated in the coordinate plane.

To work with horizontal and vertical ellipses in the coordinate plane, we consider two cases: those that are centered at the origin and those that are centered at a point other than the origin. First we will learn to derive the equations of ellipses, and then we will learn how to write the equations of ellipses in standard form. Later we will use what we learn to draw the graphs.

To derive the equation of an ellipse centered at the origin, we begin with the foci (−c,0)(−c,0) and (−c,0)(−c,0). The ellipse is the set of all points (x,y)(x,y) such that the sum of the distances from (x,y)(x,y) to the foci is constant, as shown in the figure below.



If $(a,0)(a,0)$ is a vertex of the ellipse, the distance from $(-c,0)(-c,0)$ to $(a,0)(a,0)$ is $a-(-c)=a+ca-(-c)=a+c$. The distance from $(c,0)(c,0)$ to $(a,0)(a,0)$ is $a-ca-c$. The sum of the distances from the foci to the vertex is

$$(a+c)+(a-c)=2a(a+c)+(a-c)=2a$$

If $(x,y)(x,y)$ is a point on the ellipse, then we can define the following variables:

d1=the distance from (−c,0) to (x,y)d2=the distance from (c,0) to (x,y)d1=the distance from (−c,0) to (x,y)d2=the distance from (c,0) to (x,y)

By the definition of an ellipse, $d1+d2d1+d2$ is constant for any point $(x,y)(x,y)$ on the ellipse. We know that the sum of these distances is $2a2a$ for the vertex $(a,0)(a,0)$. It follows that $d1+d2=2ad1+d2=2a$ for any point on the ellipse. The derivation of the standard form of the equation of an ellipse relies on this relationship and the distance formula. The derivation is beyond the scope of this course, but the equation is:

$$x2a2+y2b2=1x2a2+y2b2=1$$

for an ellipse centered at the origin with its major axis on the *X*-axis and

$$x2b2+y2a2=1x2b2+y2a2=1$$

for an ellipse centered at the origin with its major axis on the *Y*-axis.

Writing Equations of Ellipses Centered at the Origin in Standard Form

Standard forms of equations tell us about key features of graphs. Take a moment to recall some of the standard forms of equations we've worked with in the past: linear, quadratic, cubic, exponential, logarithmic, and so on. By learning to interpret standard forms of equations, we are bridging the relationship between algebraic and geometric representations of mathematical phenomena.

The key features of the ellipse are its center, vertices, co-vertices, foci, and lengths and positions of the major and minor axes. Just as with other equations, we can identify all of these features just by looking at the standard form of the equation. There are four variations of the standard form of the ellipse. These variations are categorized first by the location of the center (the origin or not the origin), and then by the position (horizontal or vertical). Each is presented along with a description of how the parts of the equation relate to the graph. Interpreting these parts allows us to form a mental picture of the ellipse.

A GENERAL NOTE: STANDARD FORMS OF THE EQUATION OF AN ELLIPSE WITH CENTER $(0,0)$

The standard form of the equation of an ellipse with center $(0,0)$ and **major axis** parallel to the *x*-axis is

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

where

- $a>b$
- the length of the major axis is $2a$
- the coordinates of the vertices are $(\pm a,0)$
- the length of the minor axis is $2b$
- the coordinates of the co-vertices are $(0,\pm b)$
- the coordinates of the foci are $(\pm c,0)$, where $c^2=a^2-b^2$.

The standard form of the equation of an ellipse with center $(0,0)$ and major axis parallel to the *y*-axis is

$$\frac{x^2}{b^2}+\frac{y^2}{a^2}=1$$

where

- $a>b$
- the length of the major axis is $2a$
- the coordinates of the vertices are $(0,\pm a)$
- the length of the minor axis is $2b$
- the coordinates of the co-vertices are $(\pm b,0)$
- the coordinates of the foci are $(0,\pm c)$, where $c^2=a^2-b^2$.

Note that the vertices, co-vertices, and foci are related by the equation $c^2=a^2-b^2$. When we are given the coordinates of the foci and vertices of an ellipse, we can use the relationship to find the equation of the ellipse in standard form.

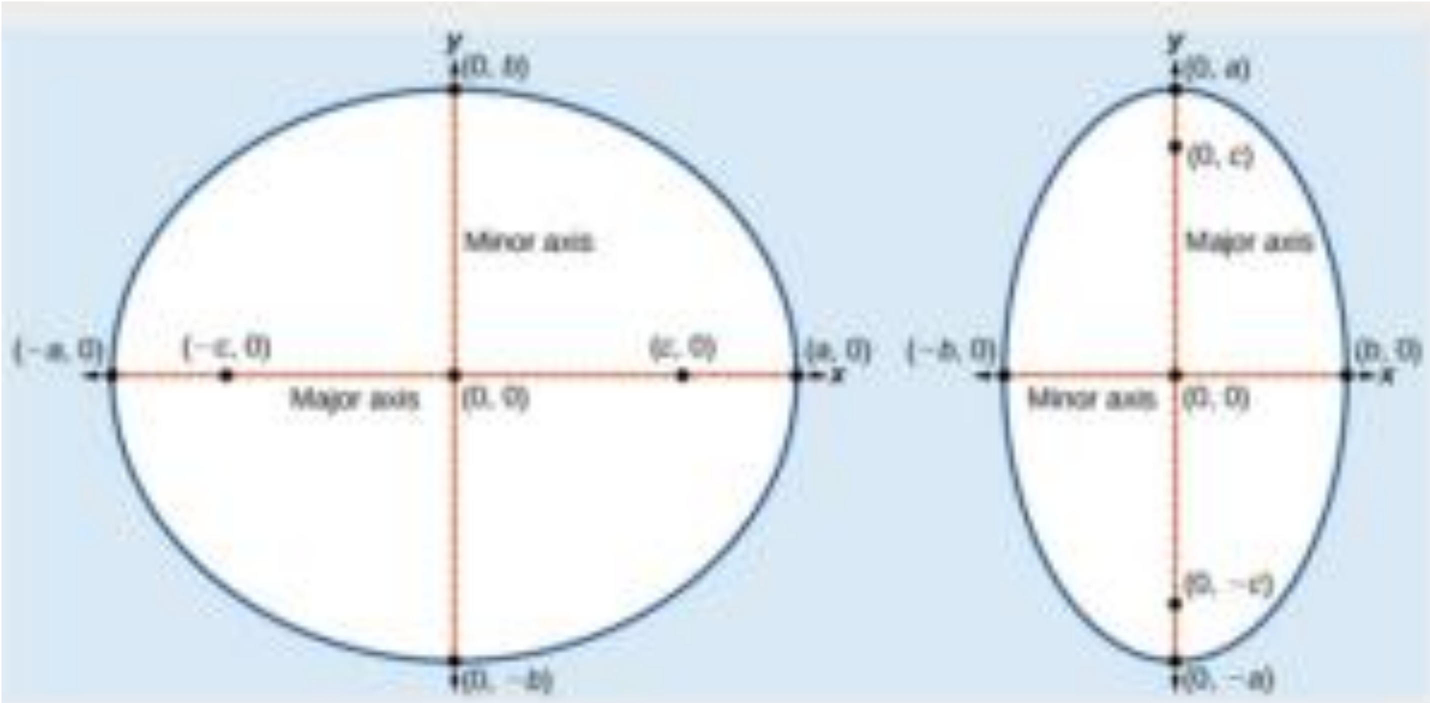


Figure: (a) Horizontal ellipse with center (0,0), (b) Vertical ellipse with center (0,0)

HOW TO: GIVEN THE VERTICES AND FOCI OF AN ELLIPSE CENTERED AT THE ORIGIN, WRITE ITS EQUATION IN STANDARD FORM.

1. Determine whether the major axis is on the *x*– or *y*-axis.
    1. If the given coordinates of the vertices and foci have the form $(\pm a,0)$ and $(\pm c,0)$ respectively, then the major axis is parallel to the *x*-axis. Use the standard form $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$.
    2. If the given coordinates of the vertices and foci have the form $(0,\pm a)$ and $(0,\pm c)$ respectively, then the major axis is parallel to the *y*-axis. Use the standard form $\frac{x^2}{b^2}+\frac{y^2}{a^2}=1$.
2. Use the equation $c^2=a^2-b^2$ along with the given coordinates of the vertices and foci, to solve for b^2.
3. Substitute the values for a^2 and b^2 into the standard form of the equation determined in Step 1.

EXAMPLE: WRITING THE EQUATION OF AN ELLIPSE CENTERED AT THE ORIGIN IN STANDARD FORM

What is the standard form equation of the ellipse that has vertices $(\pm 8,0)$ and foci $(\pm 5,0)$?

Show Show Solution

TRY IT

What is the standard form equation of the ellipse that has vertices $(0,\pm 8)$ and foci $(0,\pm\sqrt{5})$?
Show Show Solution

Writing Equations of Ellipses Not Centered at the Origin

Like the graphs of other equations, the graph of an **ellipse** can be translated. If an ellipse is translated h units horizontally and k units vertically, the center of the ellipse will be (h,k). This **translation** results in the standard form of the equation we saw previously, with x replaced by $(x-h)$ and *y* replaced by $(y-k)$.

A GENERAL NOTE: STANDARD FORMS OF THE EQUATION OF AN ELLIPSE WITH CENTER (*H, K*)

The standard form of the equation of an ellipse with center (h, k) and **major axis** parallel to the *x*-axis is

$$\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=1$$

where

- $a>b$
- the length of the major axis is $2a$
- the coordinates of the vertices are $(h\pm a,k)$
- the length of the minor axis is $2b$
- the coordinates of the co-vertices are $(h,k\pm b)$
- the coordinates of the foci are $(h\pm c,k)$, where $c^2=a^2-b^2$.

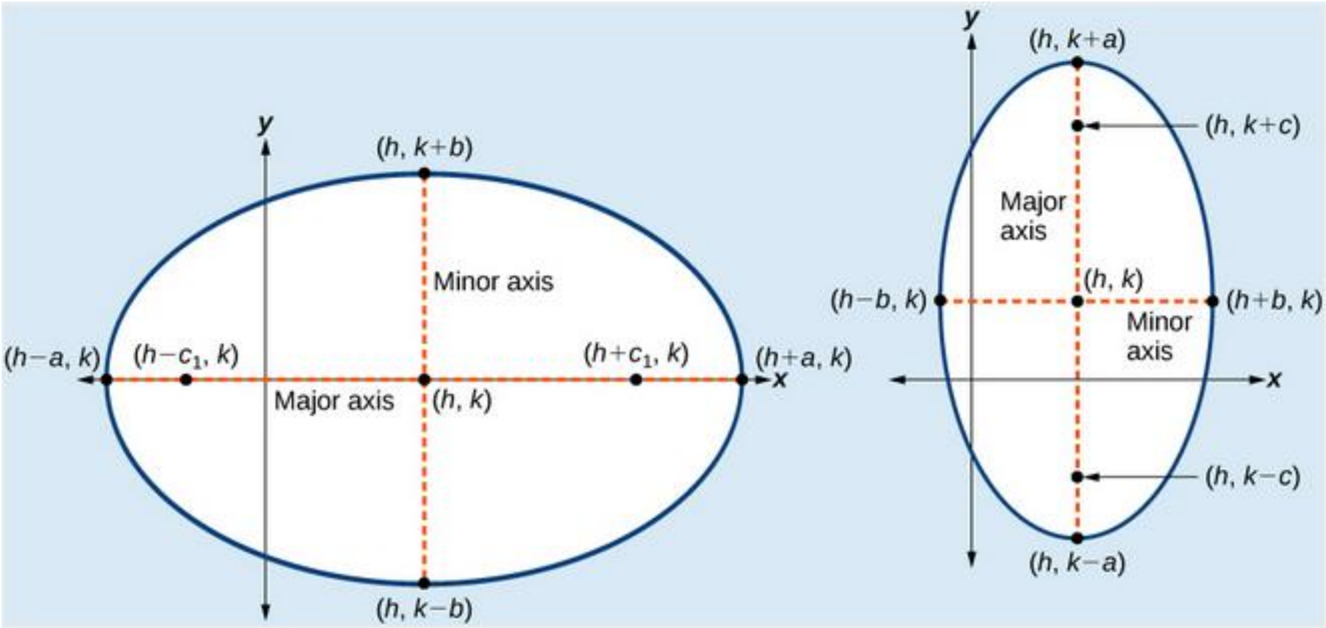
The standard form of the equation of an ellipse with center (h,k) and major axis parallel to the *y*-axis is

$$\frac{(x-h)^2}{b^2}+\frac{(y-k)^2}{a^2}=1$$

where

- $a>b$
- the length of the major axis is $2a$
- the coordinates of the vertices are $(h,k\pm a)$
- the length of the minor axis is $2b$
- the coordinates of the co-vertices are $(h\pm b,k)$
- the coordinates of the foci are $(h,k\pm c)$, where $c^2=a^2-b^2$.

Just as with ellipses centered at the origin, ellipses that are centered at a point (h,k) have vertices, co-vertices, and foci that are related by the equation $c^2=a^2-b^2$. We can use this relationship along with the midpoint and distance formulas to find the equation of the ellipse in standard form when the vertices and foci are given.



(a) Horizontal ellipse with center (h,k)(h,k) (b) Vertical ellipse with center (h,k)(h,k)

HOW TO: GIVEN THE VERTICES AND FOCI OF AN ELLIPSE NOT CENTERED AT THE ORIGIN, WRITE ITS EQUATION IN STANDARD FORM.

1. Determine whether the major axis is parallel to the *x*– or *y*-axis.
    1. If the *y*-coordinates of the given vertices and foci are the same, then the major axis is parallel to the *x*-axis. Use the standard form (x−h)2a2+(y−k)2b2=1(x−h)2a2+(y−k)2b2=1.
    2. If the *x*-coordinates of the given vertices and foci are the same, then the major axis is parallel to the *y*-axis. Use the standard form (x−h)2b2+(y−k)2a2=1(x−h)2b2+(y−k)2a2=1.
2. Identify the center of the ellipse (h,k)(h,k) using the midpoint formula and the given coordinates for the vertices.
3. Find a2a2 by solving for the length of the major axis, 2a2a, which is the distance between the given vertices.
4. Find c2c2 using hh and kk, found in Step 2, along with the given coordinates for the foci.
5. Solve for b2b2 using the equation c2=a2−b2c2=a2−b2.
6. Substitute the values for h,k,a2h,k,a2, and b2b2 into the standard form of the equation determined in Step 1.

EXAMPLE: WRITING THE EQUATION OF AN ELLIPSE CENTERED AT A POINT OTHER THAN THE ORIGIN

What is the standard form equation of the ellipse that has vertices (−2,−8)(−2,−8) and (−2,2)(−2,2) and foci (−2,−7)(−2,−7) and (−2,1)?(−2,1)?
Show Solution

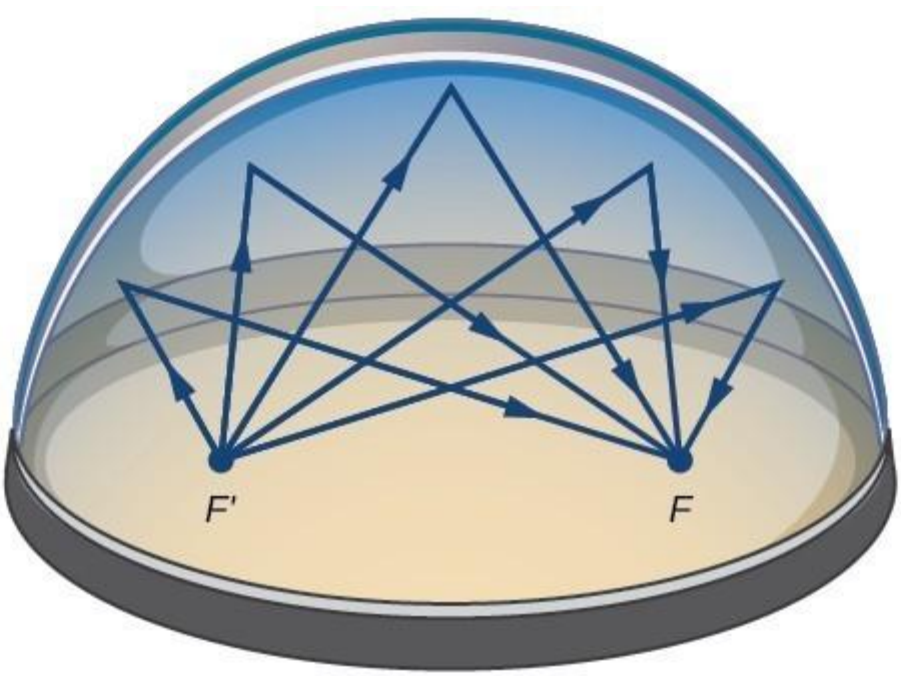
TRY IT

What is the standard form equation of the ellipse that has vertices (−3,3)(−3,3) and (5,3)(5,3) and foci (1−2√3,3)(1−23,3) and (1+2√3,3)?(1+23,3)?

Show Solution

Solving Applied Problems Involving Ellipses

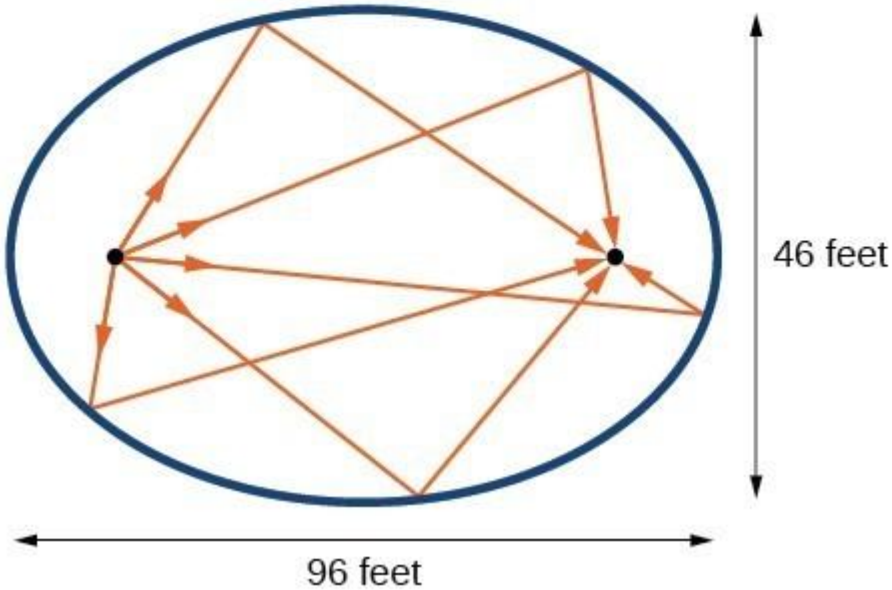
Many real-world situations can be represented by ellipses, including orbits of planets, satellites, moons and comets, and shapes of boat keels, rudders, and some airplane wings. A medical device called a lithotripter uses elliptical reflectors to break up kidney stones by generating sound waves. Some buildings, called whispering chambers, are designed with elliptical domes so that a person whispering at one focus can easily be heard by someone standing at the other focus. This occurs because of the acoustic properties of an ellipse. When a sound wave originates at one focus of a whispering chamber, the sound wave will be reflected off the elliptical dome and back to the other focus. In the whisper chamber at the Museum of Science and Industry in Chicago, two people standing at the foci—about 43 feet apart—can hear each other whisper.



EXAMPLE: LOCATING THE FOCI OF A WHISPERING CHAMBER

The Statuary Hall in the Capitol Building in Washington, D.C. is a whispering chamber. Its dimensions are 46 feet wide by 96 feet long.

a. What is the standard form of the equation of the ellipse representing the outline of the room? Hint: assume a horizontal ellipse, and let the center of the room be the point (0,0)(0,0).
b. If two senators standing at the foci of this room can hear each other whisper, how far apart are the senators? Round to the nearest foot.



Show Solution

TRY IT

Suppose a whispering chamber is 480 feet long and 320 feet wide.

a. What is the standard form of the equation of the ellipse representing the room? Hint: assume a horizontal ellipse, and let the center of the room be the point (0,0)(0,0).
b. If two people are standing at the foci of this room and can hear each other whisper, how far apart are the people? Round to the nearest foot.

Show Solution

Equation of Tangent and Normal to the Ellipse

The equations of tangent and normal to the ellipse x2a2+y2b2=1x2a2+y2b2=1 at the point (x1,y1)(x1,y1) are x1xa2+y1yb2=1x1xa2+y1yb2=1 and a2y1x–b2x1y–(a2–b2)x1y1=0a2y1x–b2x1y–(a2–b2)x1y1=0 respectively.
Consider that the standard equation of ellipse with vertex at origin (0,0)(0,0) can be written as

x2a2+y2b2=1 – – – (i)x2a2+y2b2=1 – – – (i)

Since the point (x1,y1)(x1,y1) lies on the given ellipse, it must satisfy equation (i). So we have x12a2+y12b2=1 – – – (ii)x12a2+y12b2=1 – – – (ii)

Now differentiating equation (i) on both sides with respect to xx, we have
2xa2+2yb2dydx=0⇒yb2dydx=–xa2⇒dydx=–b2xa2y2xa2+2yb2dydx=0⇒yb2dydx=–xa2⇒dydx=–b2xa2y

If mm represents the slope of the tangent at the given point (x1,y1)(x1,y1), then

m=dydx(x1,y1)=–b2x1a2y1m=dydx(x1,y1)=–b2x1a2y1

The equation of a tangent at the given point (x1,y1)(x1,y1) is
y–y1=–b2x1a2y1(x–x1)y1b2(y–y1)=–x1a2(x–x1)⇒y1yb2–y12b2=–x1xa2+x12a2⇒x1xa2+y1yb2=x12a2+y12b2⇒x1xa2+y1yb2=1y–y1=–b2x1a2y1(x–x1)y1b2(y–y1)=–x1a2(x–x1)⇒y1yb2–y12b2=–x1xa2+x12a2⇒x1xa2+y1yb2=x12a2+y12b2⇒x1xa2+y1yb2=1

This is the equation of the tangent to the given ellipse at (x1,y1)(x1,y1).
The slope of the normal at (x1,y1)(x1,y1) is –1m=–(–a2x1b2y1)=a2x1b2y1–1m=–(–a2x1b2y1)=a2x1b2y1
The equation of the normal at the point (x1,y1)(x1,y1) is y–y1=a2x1b2y1(x–x1)y–y1=a2x1b2y1(x–x1)

⇒a2y1x–b2x1y–a2x1y1+b2x1y1=0⇒a2y1x–b2x1y–(a2–b2)x1y1=0⇒a2y1x–b2x1y–a2x1y1+b2x1y1=0⇒a2y1x–b2x1y–(a2–b2)x1y1=0

This is the equation of the normal to the given ellipse at (x1,y1)(x1,y1).

Read more: https://www.emathzone.com/tutorials/geometry/equation-of-tangent-and-normal-to-ellipse.html#ixzz6COpkZ3zG

The Hyperbola

Introduction to Hyperbolas

Hyperbolas are one of the four conic sections, and are described by certain kinds of equations.

LEARNING OBJECTIVES

Connect the equation for a hyperbola to the shape of its graph

KEY TAKEAWAYS

Key Points

- A hyperbola is formed by the intersection of a plane perpendicular to the bases of a double cone.
- All hyperbolas have an eccentricity value greater than 11.
- All hyperbolas have two branches, each with a vertex and a focal point.
- All hyperbolas have asymptotes, which are straight lines that form an X that the hyperbola approaches but never touches.

Key Terms

- **hyperbola**: One of the conic sections.
- **ellipse**: One of the conic sections.
- **vertices**: A turning point in a curved function. Every hyperbola has two vertices.
- **focal point**: A point not on a hyperbola, around which the hyperbola curves.

A hyperbola can be defined in a number of ways. A hyperbola is:

1. The intersection of a right circular double cone with a plane at an angle greater than the slope of the cone (for example, perpendicular to the base of the cone)
2. The set of all points such that the difference between the distances to two focal points is constant
3. The set of all points such that the ratio of the distance to a single focal point divided by the distance to a line (the directrix) is greater than one

Let's see how that second definition gives us what is called the standard form of a hyperbola equation.

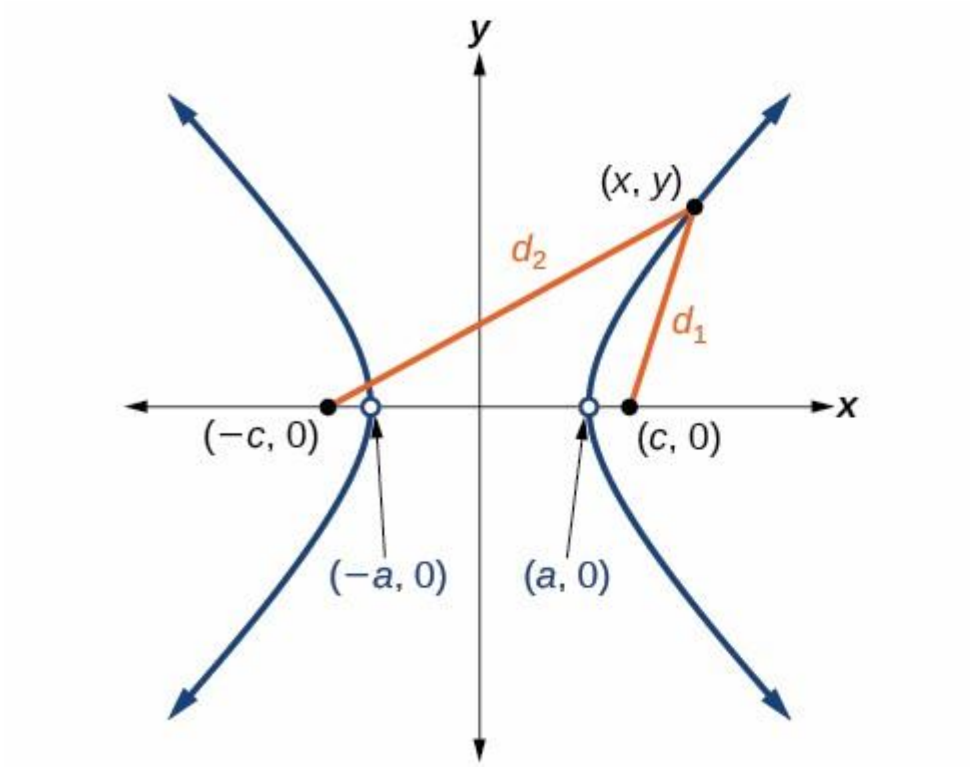


Diagram of a hyperbola: The hyperbola, shown in blue, has a center at the origin, two focal points at (−c,0)(−c,0) and (c,0)(c,0), and two vertices located at +a+a and −a−a on the xx-axis.

We begin with two focal points, F1F1 and F2F2, located on the xx-axis, so that they have coordinates
(c,0)(c,0) and (−c,0)(−c,0) (other arrangements are possible). We want the set of all points that have the same *difference* between the distances to these points. The center of this hyperbola is the origin (0,0)(0,0).

Imagine that we take a point on the red hyperbola curve, called PP, and we let that point be the +a+a value on the xx-axis. Then the difference of distances between PP and the two focal points is:

(P→F2)−(P→F1)=(c+a)−(c−a)=2a(P→F2)−(P→F1)=(c+a)−(c−a)=2a

where aa is the distance from the center (origin) to the vertices of the hyperbola. With this value for the difference of distances, we can choose any point (x,y)(x,y) on the hyperbola and construct an equation by use of the distance formula:

√(x−c)2+(y−0)2−√(x−(−c))2+(y−0)2=2a√(x−c)2+y2−√(x+c)2+y2=2a(x−c)2+(y−0)2−(x−(−c))2+(y−0)2=2a(x−c)2+y2−(x+c)2+y2=2a

From here there is some straightforward, but messy, algebra. We need to square both sides of this equation multiple times if we want the variables to escape their square roots. When the dust settles, we have:

$$x^2(c^2-a^2)-a^2y^2=a^2(c^2-a^2)$$

At this point we introduce one more parameter, defined as $b^2=c^2-a^2$, which reduces the hyperbola even further:

$$b^2x^2-a^2y^2=a^2b^2$$

Lastly we divide both sides of the equation by a^2b^2:

$$\frac{b^2x^2}{a^2b^2}-\frac{a^2y^2}{a^2b^2}=\frac{a^2b^2}{a^2b^2}$$

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

Thus, the standard form of the equation for a hyperbola with focal points on the x axis is:

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

If the focal points are on the y-axis, the variables simply change places:

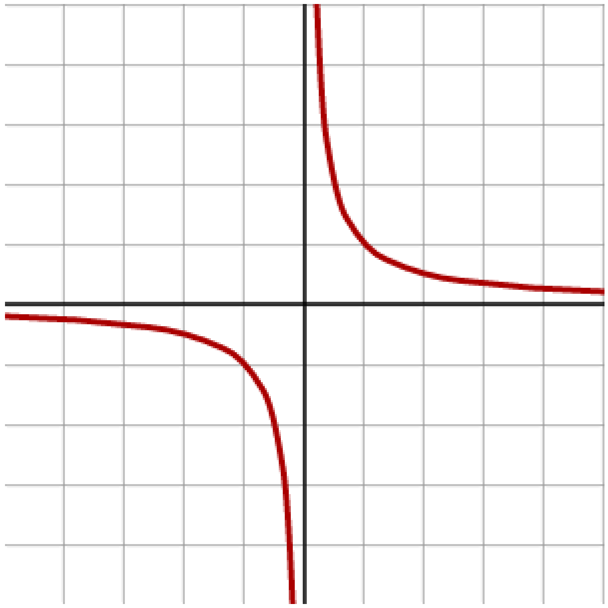
$$\frac{y^2}{a^2}-\frac{x^2}{b^2}=1$$

Note that the hyperbola standard form is very similar to the standard form of the ellipse:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

The similarity is not coincidental. The ellipse can be defined as all points that have a constant *sum* of distances to two focal points, and the hyperbola is defined as all points that have constant *difference* of distances to two focal points.

There is another common form of hyperbola equation that, at first glance, looks very different: $y=\frac{1}{x}$ or $xy=1$.



Reciprocal hyperbola: This hyperbola is defined by the equation $xy=1xy=1$.

From the graph, it can be seen that the hyperbola formed by the equation $xy=1xy=1$ is the same shape as the standard form hyperbola, but rotated by $45^{\circ}45^{\circ}$. To prove that it is the same as the standard hyperbola, you can check for yourself that it has two focal points and that all points have the same difference of distances. Another way to prove it algebraically is to construct a rotated xx–yy coordinate frame.

Parts of a Hyperbola

The features of a hyperbola can be determined from its equation.

LEARNING OBJECTIVES

Describe the parts of a hyperbola and the expressions for each

KEY TAKEAWAYS

Key Points

- Hyperbolas are conic sections, formed by the intersection of a plane perpendicular to the bases of a double cone.
- Hyperbolas can also be understood as the locus of all points with a common difference of distances to two focal points.
- All hyperbolas have two branches, each with a focal point and a vertex.

- Hyperbolas are related to inverse functions, of the family y=1xy=1x.

A hyperbola is one of the four conic sections. All hyperbolas share common features, and it is possible to determine the specifics of any hyperbola from the equation that defines it.

Standard Form

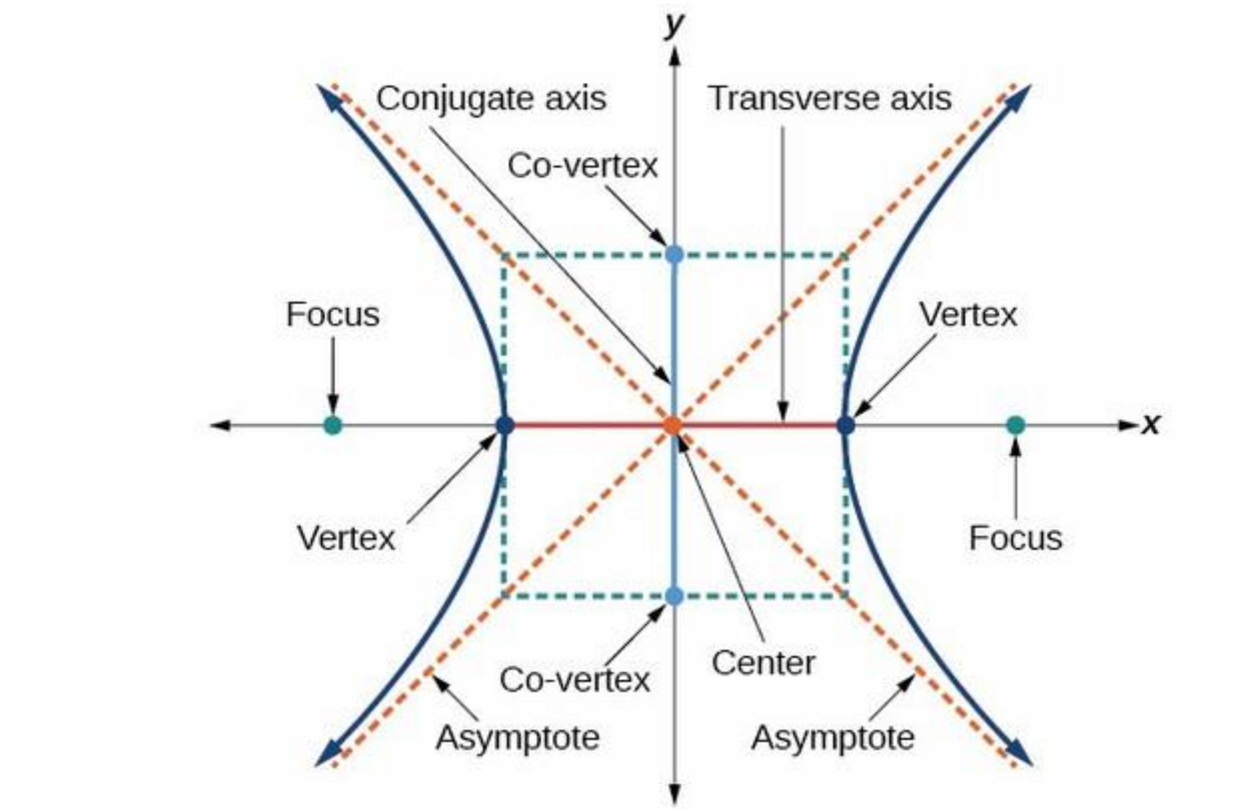


Diagram of a hyperbola: All hyperbolas share common features. A hyperbola consists of two curves, each with a vertex and a focus. The transverse axis is the axis that crosses through both vertices and foci, and the conjugate axis is perpendicular to it. A hyperbola also has asymptotes which cross in an "x". The two branches of the hyperbola are on opposite sides of the asymptotes' cross. The vertices and asymptotes can be used to form a rectangle, with the vertices at the centers of two opposite sides and the corners on the asymptotes. The centers of the other two sides, along the conjugate axis, are called the co-vertices. Where the asymptotes of the hyperbola cross is called the center.

If the foci lie on the xx-axis, the standard form of a hyperbola is:

(x−h)2a2−(y−k)2b2=1(x−h)2a2−(y−k)2b2=1

If the foci lie on the yy-axis, the standard form is:

$$\frac{(y-k)^2}{a^2}-\frac{(x-h)^2}{b^2}=1$$

We will use the x-axis hyperbola to demonstrate how to determine the features of a hyperbola, so that a is associated with x-coordinates and b is associated with y-coordinates. For a y-axis hyperbola, the associations are reversed.

Center

The center has coordinates (h,k).

Vertices

The vertices have coordinates $(h+a,k)$ and $(h-a,k)$. The line connecting the vertices is called the transverse axis.

Co-Vertices

The co-vertices correspond to b, the " minor semi-axis length", and have coordinates $(h,k+b)$ and $(h,k-b)$.

Asymptotes

The major and minor axes a and b, as the vertices and co-vertices, describe a rectangle that shares the same center as the hyperbola, and has dimensions $2a\times 2b$. The asymptotes of the hyperbola are straight lines that are the diagonals of this rectangle. We can therefore use the corners of the rectangle to define the equation of these lines:

$$y=\pm\frac{a}{b}(x-h)+k$$

The rectangle itself is also useful for drawing the hyperbola graph by hand, as it contains the vertices. When drawing the hyperbola, draw the rectangle first. Then draw in the asymptotes as extended lines that are also the diagonals of the rectangle. Finally, draw the curve of the hyperbola by following the asymptote inwards, curving in to touch the vertex on the rectangle, and then following the other asymptote out. Repeat for the other branch.

Focal Points

The foci have coordinates $(h+c,k)$ and $(h-c,k)$. The value of c is found with this relation:

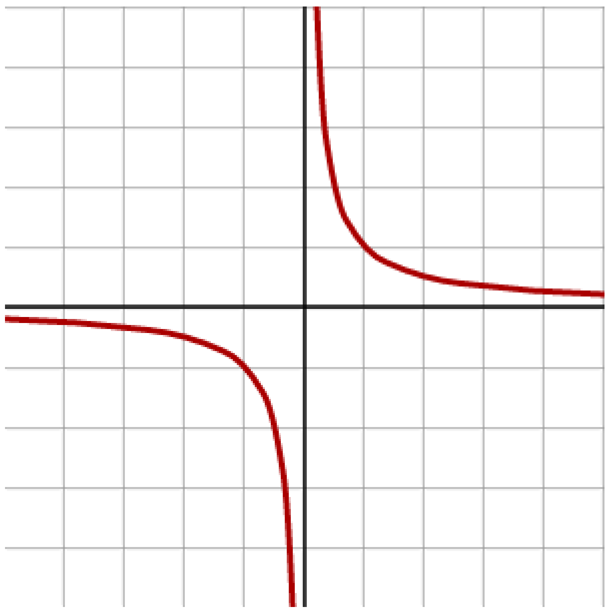
$$c^2=a^2+b^2$$

Rectangular Hyperbola

Rectangular hyperbolas, defined by

$(x-h)(y-k)=m(x-h)(y-k)=m$

for some constant mm, are much simpler to analyze than standard form hyperbolas.



Rectangular hyperbola: This rectangular hyperbola has its center at the origin, and is also the graph of the function $f(x)=1xf(x)=1x$.

Center

The center of a rectangular hyperbola has coordinates $(h,k)(h,k)$.

Vertices and Co-Vertices

The rectangular hyperbola is highly symmetric. Both its major and minor axis values are equal, so that $a=b=\sqrt{2m}a=b=2m$. The vertices have coordinates $(h+\sqrt{2m},k+\sqrt{2m})(h+2m,k+2m)$ and $(h-\sqrt{2m},k-\sqrt{2m})(h-2m,k-2m)$.

The co-vertices have coordinates $(h-\sqrt{2m},k+\sqrt{2m})(h-2m,k+2m)$ and $(h+\sqrt{2m},k-\sqrt{2m})(h+2m,k-2m)$.

Asymptotes

The asymptotes of a rectangular hyperbola are the xx– and yy-axes.

Focal Points

We can use c2=a2+b2c2=a2+b2 as before. With a=b=√2ma=b=2m, we find that c=±2√mc=±2m. Therefore the focal points are located at (h+2√m,k+2√m)(h+2m,k+2m) and (h−2√m,k−2√m)(h−2m,k−2m).

Applications of Hyperbolas

A hyperbola is an open curve with two branches and a cut through both halves of a double cone, which is not necessarily parallel to the cone's axis.

LEARNING OBJECTIVES

Discuss applications of the hyperbola to real world problems

KEY TAKEAWAYS

Key Points

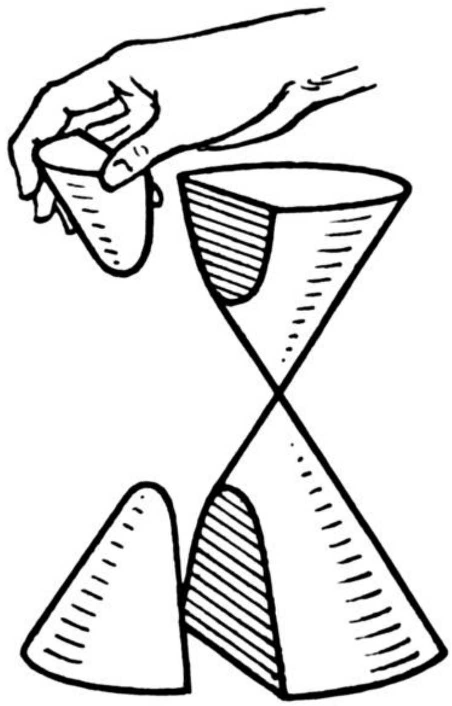
- Hyperbolas have applications to a number of different systems and problems including sundials and trilateration.
- Hyperbolas may be seen in many sundials. On any given day, the sun revolves in a circle on the celestial sphere, and its rays striking the point on a sundial trace out a cone of light. The intersection of this cone with the horizontal plane of the ground forms a conic section.
- A hyperbola is the basis for solving trilateration problems, the task of locating a point from the differences in its distances to given points—or, equivalently, the difference in arrival times of synchronized signals between the point and the given points.

Key Terms

- **trilateration**: The determination of the location of a point based on its distance from three other points.
- **hyperbola**: A conic section formed by the intersection of a cone with a plane that intersects the base of the cone and is not tangent to the cone.
- **conic section**: Any of the four distinct shapes that are the intersections of a cone with a plane, namely the circle, ellipse, parabola and hyperbola.

Applications and Problem Solving

As we should know by now, a hyperbola is an open curve with two branches, the intersection of a plane with both halves of a double cone. The plane may or may not be parallel to the axis of the cone.



Hyperbola: A hyperbola is an open curve with two branches, the intersection of a plane with both halves of a double cone. The plane may or may not be parallel to the axis of the cone.

Here are some examples of hyperbolas in the real world.

Sundials

Hyperbolas may be seen in many sundials. Every day, the sun revolves in a circle on the celestial sphere, and its rays striking the point on a sundial traces out a cone of light. The intersection of this cone with the horizontal plane of the ground forms a conic section. The angle between the ground plane and the sunlight cone depends on where you are and the axial tilt of Earth, which changes seasonally. At most populated latitudes and at most times of the year, this conic section is a hyperbola.

Sundials work by casting the shadow of a vertical marker, sometimes called a gnomon, over a clock face on the horizontal surface. The angle between the sunlight and the ground will be the same as the angle formed by the line connecting the tip of the gnomon with the end of its shadow.

If we mark where the end of the shadow falls over the course of the day, the line traced out by the shadow forms a hyperbola on the ground (this path is called the *declination line*). The shape of this hyperbola varies with the geographical latitude and with the time of the year, since those factors affect the angle of the cone of the sun's rays relative to the horizon.

The parameters of the traced hyperbola, such as its asymptotes and its eccentricity, are related to the specific physical conditions that produced it, namely the angle between the sunlight and the ground, and the latitude at which the sundial exists.

Hyperbolas and sundials: Hyperbolas as declination lines on a sundial.

Trilateration

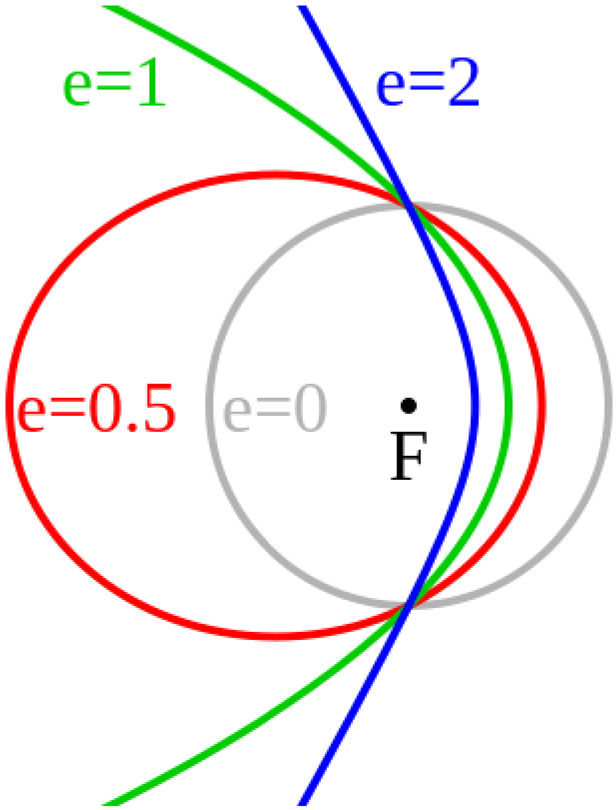
Trilateration is the a method of pinpointing an exact location, using its distances to a given points. The can also be characterized as the difference in arrival times of synchronized signals between the desired point and known points. These types of problems arise in navigation, mainly nautical. A ship can locate its position using the arrival times of signals from GPS transmitters. Alternatively, a homing beacon can be located by comparing the arrival times of its signals at two separate receiving stations. This can be used to track people, cell phones, internet signals, and many other things.

In the case in which a ship, or other object to be located, only knows the difference in distances between itself and two known points, the curve of possible locations is a hyperbola. One way of defining a hyperbola is as precisely this: the curve of points such that the absolute value of the difference between the distances to two focal points remains constant.

So if we call this difference in distances 2a2a, the hyperbola will have vertices separated by the same distance 2a2a, and the foci of the hyperbola will be the two known points.

The Kepler Orbit of Particles

The Kepler orbit is the path followed by any orbiting body. This can be applied to a particle of any size, as long as gravity is the only force causing the orbital trajectory. Depending on the orbital properties, including size and shape (eccentricity), this orbit can be any of the four conic sections. In particular, if the eccentricity *e* of the orbit is greater than 11, the path of such a particle is a hyperbola. In the figure, the blue line shows the hyperbolic Kepler orbit. In the common case of a gravitational orbit, the massive object is one of the foci of the hyperbola (or other conic section).



Kepler orbits: A diagram of the various forms of the Kepler Orbit and their eccentricities. Blue is a hyperbolic trajectory (e>1e>1). Green is a parabolic trajectory (e=1e=1). Red is an elliptical orbit (e<1e<1). Grey is a circular orbit (e=0e=0).

Physically, another way to understand hyperbolic orbits is in terms of the energy of the orbiting particle. Orbits which are circular or elliptical are bound orbits, which is to say the object never escapes its closed path around one of the focal points. This is associated with the particle's total energy EE being less than the minimum energy required to escape, and so EE is said to be negative in these cases.

A parabolic trajectory does have the particle escaping the system. However, this is the very special case when the total energy EE is exactly the minimum escape energy, so EE in this case is considered to be zero.

If there is any additional energy on top of the minimum (zero) value, the trajectory will become hyperbolic, and so EE is positive in the hyperbolic orbit case.

9.4 Translation and Rotation of Axes

- Define translation and rotation of axes and demonstrate through examples.

- Find the equations of transformation form
  I. Translation of axes,
  II. Rotation of axes.
- Find the transformed equation by using translation or rotation of axes
- Find new origin and new axes referred to old origin and old axes.
- Find the angle through which the axes be rotated about the origin so that the product term xy is removed from the transformed equations.

16. Translation / Rotation : Examples

This will be the last lesson in the Coordinate Geometry Basics series. I'll be closing with a few solved examples relating to translation and rotation of axes.

Example 1 Find the new coordinates of the point (3, 4)

(i) when the origin is shifted to the point (1, 3).

(ii) when the axes are rotated by an angle θ anticlockwise, where $\tan\theta = 4/3$

(iii) when the origin is shifted to (1, –2), and the axes are rotated by 90° in the clockwise direction.

Solution (i) We'll directly use the formula derived in the previous lesson: $x = X + h$, $y = Y + k$

We have, $3 = X + 1$ and $4 = Y + 3$, or $X = 2$ and $Y = 1$. Therefore, the coordinates with respect to the shifted origin are **(2, 1)**.

As simple as that! The next two are equally simple.

(ii) In this case, we need to calculate the values of $\sin\theta$ and $\cos\theta$ first. They'll come out to be 4/5 and 3/5 respectively. Now let's use our formulas $x = X\cos\theta – Y\sin\theta$ and $y = X\sin\theta + Y\cos\theta$.

That is, $3 = (3X – 4Y)/5$ and $4 = (4X + 3Y)/5$. Solving for X and Y, we get $X = 5$ and $Y = 0$. Therefore, the new coordinates will be **(5, 0)**.

(iii) We didn't talk about simultaneous rotation as well as translation. But turns out it is quite easy. We can find the new coordinates by first shifting the origin, followed by rotation, or the other way around.

We can also combine the two formulas straight away, i.e. $x = X\cos\theta - Y\sin\theta + h$ and $y = X\sin\theta + Y\cos\theta + k$, and solve for X and Y to obtain the new coordinates. (You may try doing it separately and compare the answers)

Note that the axes are rotated clockwise in this case, but our formulas consider anticlockwise direction. So we've to take $\theta = -90°$

Let's calculate now: $3 = X\cos(-90°) - Y\sin(-90°) + 1$ and $4 = X\sin(-90°) + Y\cos(-90°) - 2$.

This gives us $X = -6$ and $Y = 2$. Therefore, the final coordinates are **(–6, 2)**

The next few problems will talk about equations of curves with respect to the new coordinate systems.

Example 2 Find the new equation of the following curves after the coordinates are transformed as indicated:

(i) $x + 3y = 6$, when the origin is shifted to the point (–4, 1).

(ii) Find the equation of the curve $x^2 + y^2 = 4$, when the axes are rotated by an angle of 60° in the anticlockwise direction.

(iii) Find the equation of the curve $x^2 - y^2 = 10$, when the axes are rotated by an angle of 45° in the clockwise direction.

Solution (i) In this case we do not need to find the new coordinates. We only need to find the relation between them (that's what an equation is). So we'll simply replace the old coordinates with the new ones in the given equation.

Using the expressions for shifting of origin, we have $x = X - 4$ and $y = Y + 2$. Substituting the values in the given equation, we get $X - 4 + 3(Y + 1) = 6$. Or, $\mathbf{X + 3Y = 7}$.

And that's all. We've obtained the relation between the new coordinates, which is nothing but the equation of that curve with respect to the new origin.

(ii) This one is quite similar to the previous one, except that we're rotating the axes instead of translating them. Using the formulas, we have $x = X\cos60° - Y\sin60°$ and $y = X\sin60° + Y\cos60°$. Let's substitute in the given equation. We get, $(X\cos60° - Y\sin60°)^2 + (X\sin60° + Y\cos60°)^2 = 4$. This leads to $X^2\cos^2 60° + Y^2\sin^2 60° - 2XY\cos60°\sin60° + X^2\sin^2 60° + Y^2\cos^2 60° + 2XY\sin60°\cos60° = 4$ or $\mathbf{X^2 + Y^2 = 4}$.

Hmm. Nothing happened. Strange.

(iii) This one is quite similar to the previous one, except that we're rotating the axes instead of translating them. Using the formulas, we have $x = X\cos45° + Y\sin45°$ and $y = -X\sin45° + Y\cos45°$. On substituting in the given equation, we get $(X\cos45° + Y\sin45°)^2 - (-X\sin45° + Y\cos45°)^2 = 10$. This on simplification gives us $\mathbf{XY = 5}$.

Example 3 To what point should the origin be shifted so that the equation $x^2 + y^2 – 4x + 6y – 4 = 0$ becomes free of the first degree terms? (i.e. -4x and 6y)

Solution Let the origin be shifted to the point (h, k). Let's transform the equation, and see what happens. Replacing x by X + h and y by Y + k, we get $X^2 + Y^2 + X(-4 + 2h) + Y(6 + 2k) + h^2 + k^2 – 4h + 6k – 4 = 0$.

Now the first degree terms in this equation have the coefficients (–4 + 2h) and (6 + 2k). Since we want to get rid of them, we'll equate both to them to zero. By doing that, we get **h = 2** and **k = -3.**

The new equation will look like $X^2 + Y^2 – 17 = 0$. (No first degree terms, yay!)

Example 4 By what angle should the axes be rotated so that the equation $3x^2 + 2xy + y^2 = 1$ becomes free of the xy term?

Solution This one is similar to the previous one, except that now we've got to rotate the axes. Let's do the hard work. The transformed equation will become $3(X\cos\theta – Y\sin\theta)^2 + 2(X\cos\theta – Y\sin\theta)(X\sin\theta + Y\cos\theta) + (X\sin\theta + Y\cos\theta)^2 = 1$. (Whoa!)

$X^2(3\cos^2\theta + 2\sin\theta\cos\theta + \sin^2\theta) + XY(-4\sin\theta\cos\theta +2\cos^2\theta – 2\sin^2\theta) + Y^2(3\sin^2\theta + \cos^2\theta – 2\sin\theta\cos\theta) = 1$

We don't want the XY term, so we'll equate its coefficient to zero. We get $2\cos2\theta = 2\sin2\theta$ or $\tan2\theta = 1$, or θ = **45°**.

That'll be all for coordinate geometry basics. We now have all the tools we need to sail through the next set of lessons. Hope you enjoyed!

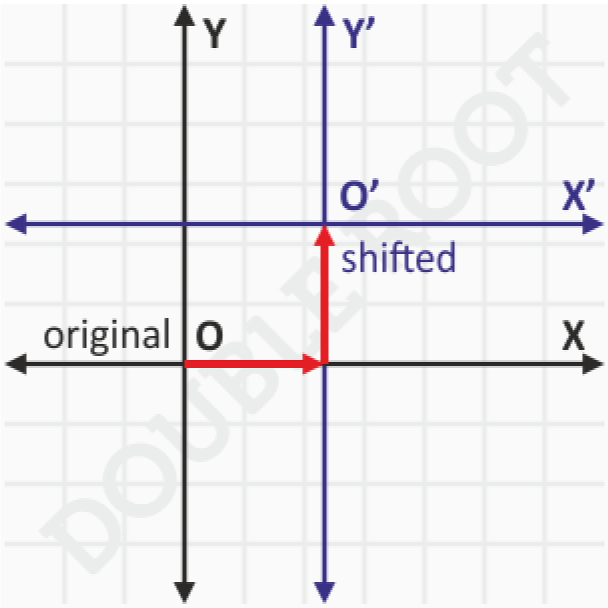
14. Translation of Axes

This lesson will talk about something known as 'translation of axes' or 'shifting of origin'.

Shifting of Origin / Translation of Axes

What we're trying to do here is shift the origin to a different point (without changing the orientation of the axes), and see what happens to the coordinates of a given point.

See the figure below to get an idea of what we'll be doing.

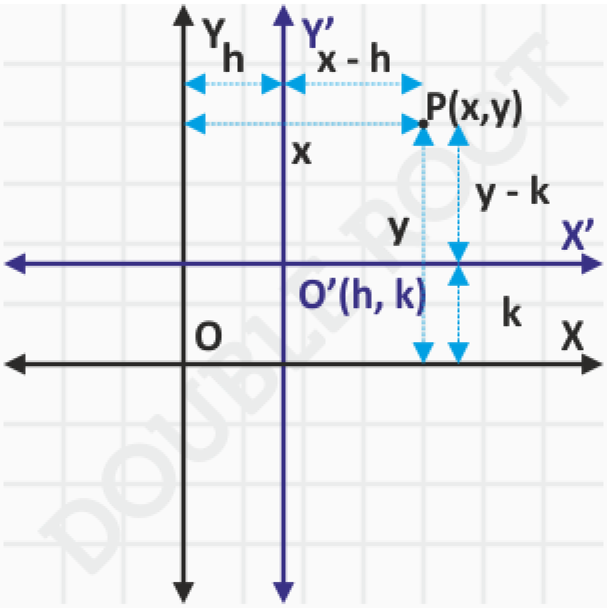


But why are we doing that?

Well, for the moment, you'll have to believe me that shifting of origin leads to simplification of many problems in coordinate geometry. But it'll make sense to you only when you see it in action in subsequent chapters, especially conic sections. For now, let's just focus on how shifting of origin works, and how to apply it to problems.

Consider a point P(x, y), and let's suppose the origin has been shifted to a new point, say (h, k). What will be the coordinates of the point P, with respect to this new origin?

Recall that the coordinates of a point are it's (signed) distances from the coordinate axes. So, to find the coordinates of the point P(x, y), we have to find its distances from the shifted coordinate axes. Turns out this is quite easy to find. Have a look at the figure below.



The shifted origin has the coordinates (h, k). That is, the shifted X and Y axes are at distances h and k from the original X and Y axes respectively.

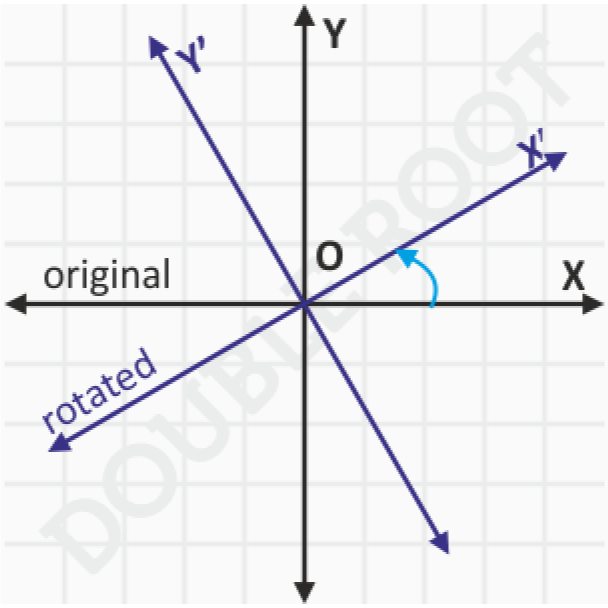
Therefore, the distance of the point P from the new X-axis will be x – h and from the shifted Y-axis will be y – k. This means that the coordinates of the point P will be (x – h, y – k). And that's all there is to it.

In other words, if (X, Y) be the coordinates of the point with respect to the new origin, then X = x – h and Y = y – k. Or **x = X + h and y = Y + k**.

These will be the very substitutions which you'll make when finding equations of a curve with respect to shifted axes. Don't worry about this now, we'll come to that later on.

15. Rotation of Axes

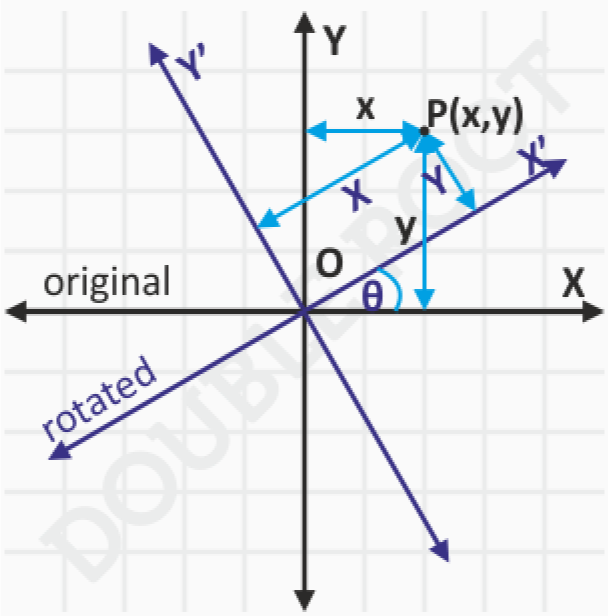
This lesson will discuss rotation of the coordinate axes about the origin. Something like this:



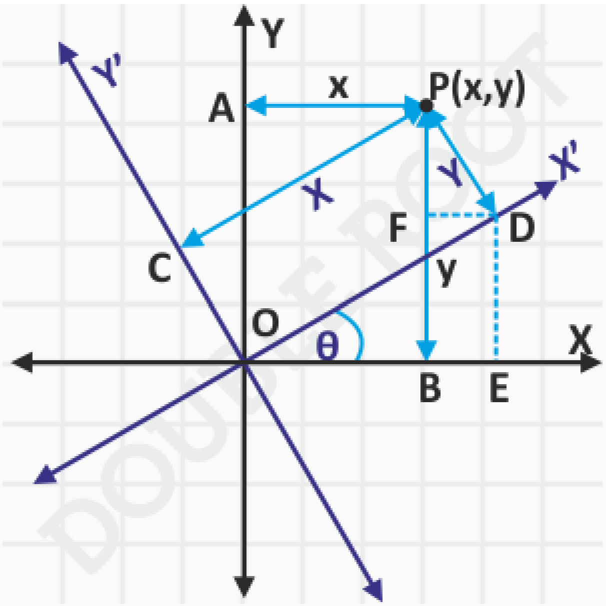
Rotation of Axes

Consider a point P(x, y), and let's suppose that the axes have been rotated about origin by an angle θ in the anticlockwise direction. What will be the coordinates of the point P, with respect to the new axes?

To find the new coordinates, we'll find the distances of the point from the rotated axes. Let's do the calculations. Let (x, y) be the original coordinates and (X, Y) be the final ones. The following figure well help us.



It'll get a little messy. Let's zoom in a bit. I've labeled the points nicely – have a close look.



Told you it'll get messy. Let's begin with the calculations: PA = x and PB = y (the old coordinates). Next, PC = X, PD = Y (the new coordinates). I've dropped perpendiculars from D to the old X axis, and to PB (DE and DF respectively).

Now PF = Ycosθ, and FD = BE = Ysinθ. Lastly, OE = ODcosθ = CPcosθ = Xcosθ. And, FB = DE = ODsinθ = Xsinθ.

And we're almost done.

Now AP = x = OB, which can be written as OE – BE, that is Xcosθ – Ysinθ. Therefore, **x = Xcosθ – Ysinθ**.

Next, OA = y = PB, which can be written as FB + PF = Xsinθ + Ycosθ. In other words, **y = Xsinθ + Ycosθ**.

And that's it! We've obtained the relation between the old coordinates (x and y) and the new ones (X and Y).

Lesson Summary

1. Let P(x, y) be a point in the X-Y plane. If the axes are rotated by an angle θ in the anticlockwise direction about the origin, then the coordinates of P with respect to the rotated axes will be given by the following relations:

$$\mathbf{x = X\cos\theta - Y\sin\theta}$$

$$\mathbf{y = X\sin\theta + Y\cos\theta}.$$

UNIT 10 DIFFERNTIAL EQUATIONS

10.1 Introduction

- Define ordinary differential equation (DE), order of a DE, degree of a DE, solution of a DE – general solution and particular solution

In contrary to what has been mentioned in the other two already existing answers,i would like to mention a few very **crucial points** regarding the **order** and **degree** of differential equations.

This topic,i.e. **order** and **degree** of the differential equation is not as trivial as it is often projected in most high school mathematics text books by avoiding certain finer mathematical aspect.

So let's first start by conventionally defining what an **order and** a **degree** is and then find out where do we stumble as per that definition.

Definition..

ORDER : The order of a differential equation is the order of the highest derivative involved in the equation.

DEGREE: The degree of a differential equation,**of which the differential coefficients are free from radicals and fractions,** is the positive integral index of the highest power of the highest order derivatives involved.

For example, in the following equation,

d2ydx2+3dydx+2y=0d2ydx2+3dydx+2y=0
Order:2, degree:1 .

Thus the above equation is of second order and first degree.

So where is the definition wrong or incomplete?

Let me tell you that although the above definition of order is complete, that of degree is not!

Let me illustrate that with the help of another example.

Can you as per the above definition of **degree** determine the degree of the following differential equation.

sin(d2ydx2)+e5dydx−3cos(x3d3ydx3)=0sin(d2ydx2)+e5dydx−3cos(x3d3ydx3)=0

You can't.

Strictly speaking, the definition of **degree**, as i have suggested or the one we are used to is **unimpressive and lacks mathematical precision.**

In fact, I would like to emphasize on this point that ***one can talk about the degree of a differential equation when it can be expressed as a polynomial in the derivatives,i.e. when the D.E. can be expressed as a sum of finite number of terms equated to 0, where each term is a finite product of the form*** f(x,y)(dydx)n1(d2ydx2)n2...f(x,y)(dydx)n1(d2ydx2)n2..., f(x,y)f(x,y) ***being the function of x, y(may be a constant or simply a function of x,y only, but not involving derivatives) and*** n1,n2,...n1,n2,... ***being non-negative integers(provided of course, that at least one derivative of some order in at least one such term survives).***
For example,

x(d2ydx2)3dydx−5ex(d2ydx2)+ylog(x)=0x(d2ydx2)3dydx−5ex(d2ydx2)+ylog(x)=0 and 2logx(dydx)2+7cosx(d2ydx2)4(dydx)7+xy=02logx(dydx)2+7cosx(d2ydx2)4(dydx)7+xy=0 are **D.E.'s expressed in polynomials in derivatives(which are evidently of degree 3 and 4 respectively)**
Thus, it's very important to note that the concept of degree cannot be attributed to all differential equations. For example,

edydx+sin(xd2ydx2)+cos(d3ydx3)+7xy=0edydx+sin(xd2ydx2)+cos(d3ydx3)+7xy=0 **has no degree.**
It is extremely difficult task,more so at that level,to conclude whether a given differential equation can be expressed as a polynomial in that derivatives involved.

Thus it is extremely important to note this point.

10.2 Formation of Differential Equations

- Explain the concept of formation of a differential equation.

Differential Equations

A differential equation is an equation which contains a derivative (such as dy/dx).

Example

Verify that $\frac{1}{x^3}$

is a solution of the differential equation:

$$\frac{dy}{dx} = \frac{y}{x} - x^2y^2 - \frac{3}{x^4}$$

To verify that something solves an equation, you need to substitute it into the equation and show that you get zero. So, here we need to work out dy/dx and show that this is equal to the right hand side when we substitute the x^{-3} into it.

$dy/dx = -3x^{-4}$

and check that when you substitute x^{-3} into the right hand side you also get $-3x^{-4}$.

Solving

When given a differential equation, you will often be asked to "solve" the differential equation or find the "general solution". This basically means find an expression which does not contain any derivatives. To do this you will need to integrate.

Example

Solve the differential equation: $\frac{dy}{dx} = 5x + 3$

This can be rewritten as:

$$\int dy = \int (5x+3)dx$$

(If you have difficulties with this, think dy/dx as a fraction. It is as if we have multiplied each side by dx)

Now simply integrate both sides:

$$y = \frac{5x^2}{2} + 3x + c$$

Initial or Boundary Conditions

You may be given information which allows you to work out the constant of integration.

In the above example, for example, you may have been told that y(0) = 0 [which means y = 0 when x = 0].

We can substitute these values into our answer to determine c:

0 = 0 + 0 + c, so c = 0

The answer in this case would therefore be $y = 5x^2/2 + 3x$.

Forming Differential Equations

Using what you now know, you should be able to form simple differential equations from a statement.

Example

The velocity of a body is proportional to its distance from O. The body starts at 1.

If x is the distance from O, then the velocity is the rate of change of distance = dx/dt

Hence $\frac{dx}{dt}$ = ?k x, where k is the constant of proportionality

Now, integrating gives us: ln x = kt + c
and we know that x(0) = 1, hence 0 = 0 + c

so ln x = kt
$x = e^{kt}$

10.3 Solution of Differential Equation

- Solve differential equations of first order and first degree of the form
  - I. Separable variables,
  - II. Homogeneous equations,
  - III. Equations reducible to homogeneous form.
- Simplify real life problems related to differential equations.

Differential Equations

A Differential Equation is an equation with a function and one or more of its derivatives:

Example: an equation with the function **y** and its derivative ***dy*dx**

Solving

We **solve** it when we discover **the function y** (or set of functions y).

There are many "tricks" to solving Differential Equations (*if* they can be solved!).

But first: why?

Why Are Differential Equations Useful?

In our world things change, and **describing how they change** often ends up as a Differential Equation:

Example: Rabbits!

The more rabbits we have the more baby rabbits we get. Then those rabbits grow up and have babies too! The population will grow faster and faster.

The important parts of this are:

- the population **N** at any time **t**
- the growth rate **r**
- the population's rate of change $\frac{d}{dt}$ **N**

Let us imagine some actual values:

- the population **N** is **1000**
- the growth rate **r** is **0.01** new rabbits per week **for every current rabbit**

The population's rate of change $\frac{d}{dt}$ **N** is then 1000×0.01 = **10 new rabbits** per week.

But that is only true at a **specific time**, and doesn't include that the population is constantly increasing.

Remember: the bigger the population, the more new rabbits we get!

So it is better to say the rate of change (at any instant) is the growth rate times the population at that instant:

$$\frac{dN}{dt} = rN$$

And that is a **Differential Equation**, because it has a function **N(t)** and its derivative.

And how powerful mathematics is! That short equation says "the rate of change of the population over time equals the growth rate times the population".

Differential Equations can describe how populations change, how heat moves, how springs vibrate, how radioactive material decays and much more. They are a very natural way to describe many things in the universe.

What To Do With Them?

On its own, a Differential Equation is a wonderful way to express something, but is hard to use.

So we try to **solve** them by turning the Differential Equation into a simpler equation without the differential bits, so we can do calculations, make graphs, predict the future, and so on.

Example: Compound Interest

Money earns interest. The interest can be calculated at fixed times, such as yearly, monthly, etc. and added to the original amount.

This is called compound interest .

But when it is compounded continuously then at any time the interest gets added in proportion to the current value of the loan (or investment).

And the bigger the loan the more interest it earns.

Using **t** for time, **r** for the interest rate and **V** for the current value of the loan:

$$\frac{dV}{dt} = rV$$

And here is a cool thing: it is the same as the equation we got with the Rabbits! It just has different letters. So mathematics shows us these two things behave the same.

Solving

The Differential Equation says it well, but is hard to use.

But don't worry, it can be solved (using a special method called Separation of Variables) and results in:

$$V = Pe^{rt}$$

Where **P** is the Principal (the original loan).

So a continuously compounded loan of $1,000 for 2 years at an interest rate of 10% becomes:

$$V = 1000 \times e^{(2\times 0.1)}$$

$$V = 1000 \times 1.22140...$$

V = $1,221.40 (to nearest cent)

So Differential Equations are great at describing things, but need to be solved to be useful.

More Examples of Differential Equations

The Verhulst Equation

Example: Rabbits Again!

Remember our growth Differential Equation:

$$\frac{dN}{dt} = rN$$

Well, that growth can't go on forever as they will soon run out of available food.

So let's improve it by including:

- the maximum population that the food can support **k**

A guy called Verhulst figured it all out and got this Differential Equation:

$$\frac{dN}{dt} = rN(1-N/k)$$

The Verhulst Equation

Simple harmonic motion

In Physics, Simple Harmonic Motion is a type of periodic motion where the restoring force is directly proportional to the displacement. An example of this is given by a mass on a spring.

Example: Spring and Weight

A spring gets a weight attached to it:

- the weight gets pulled down due to gravity,
- as the spring stretches its tension increases,
- the weight slows down,
- then the spring's tension pulls it back up,
- then it falls back down, up and down, again and again.

Describe this with mathematics!

The weight is pulled down by gravity, and we know from Newton's Second Law that force equals mass times acceleration:

$$\mathbf{F} = m\mathbf{a}$$

And acceleration is the second derivative of position with respect to time, so:

$$\mathbf{F} = m \frac{d^2x}{dt^2}$$

The spring pulls it back up based on how stretched it is (**k** is the spring's stiffness, and **x** is how stretched it is): **F = -kx**

The two forces are always equal:

$$m \frac{d^2x}{dt^2} = -kx$$

We have a differential equation!

It has a function **x(t)**, and it's second derivative $\frac{d^2x}{dt^2}$

*Note: we haven't included "damping" (the slowing down of the bounces due to friction), which is a little more complicated, but you can play with it here (press **play**):*

10.4 Orthogonal Trajectories Find orthogonal trajectories (rectangular coordinates) of the given family of curves.

- Apply MAPLE graphics commands to view the graphs of given family of curves and its orthogonal trajectories

Creating a differential equation is the first major step. But we also need to **solve** it to discover how, for example, the spring bounces up and down over time.

FIELD LINES, ORTHOGONAL LINES, DOUBLE ORTHOGONAL SYSTEM

Two families of curves are said to be *orthogonal* when at every point common to a curve of each family, the tangents are orthogonal, and one of the families is said to be composed of the *orthogonal trajectories* of the other. This constitutes a *double orthogonal system of curves*.

	If the first family of curves is defined by:	then the orthogonal trajectories are defined by:
Geometrical definition	$f(M)$ = constant	$g(M)$ = constant with $\overrightarrow{\mathbf{grad}}\, f\ \overrightarrow{\mathbf{grad}}\, g = 0$
Cartesian implicit equation	$P(x, y)$ = constant	$Q(x, y)$ = constant with $\frac{\partial P}{\partial x}\frac{\partial Q}{\partial x} + \frac{\partial P}{\partial y}\frac{\partial Q}{\partial y} = 0$
Harmonic Cartesian implicit equation	$P(x, y)$ = constant with P harmonic	$Q(x, y)$ = constant with $\frac{\partial Q}{\partial x} = -\frac{\partial P}{\partial y}$ **et** $\frac{\partial Q}{\partial y} = \frac{\partial P}{\partial x}$
Complex implicit equation	Re ($f(z)$) = constant with f holomorphic (hence	Im ($f(z)$) = constant

	conformal)	
Polar implicit equation	$P(\rho, \theta)$ = constant	$Q(\rho, \theta)$ = constant with $\rho^2 \frac{\partial P}{\partial \rho}\frac{\partial Q}{\partial \rho} + \frac{\partial P}{\partial \theta}\frac{\partial Q}{\partial \theta} = 0$
Cartesian differential equation	$y' = f(x, y)$	$y' = -1 / f(x, y)$
Polar differential equation	$\rho' = f(\rho, \theta)$	$\rho' = -\rho^2 / f(\rho, \theta)$
Field lines of the Cartesian field:	$(f(x, y), g(x, y))$	$(g(x, y), -f(x, y))$
Field lines of the polar field:	$(f(\rho, \theta), g(\rho, \theta))$	$(g(\rho, \theta), -f(\rho, \theta))$

If both the families are given by a unique parametric form: $\begin{cases} x = P(u,v) \\ y = Q(u,v) \end{cases}$, with fixed *u* and variable *v* for the first family, fixed *v* and variable *u* for the second one, then the families are orthogonal iff $\frac{\partial P}{\partial u}\frac{\partial P}{\partial v} + \frac{\partial Q}{\partial u}\frac{\partial Q}{\partial v} = 0$, which is always the case when *P* and *Q* are the real and imaginary part of a holomorphic function (inverse function of the one above).

The orthogonal trajectories of a family of lines are the involutes of the envelope of this family; therefore, they are parallel curves (see example 13 below).

Examples :

Definitions	Common parametric expression (red	Inverse images of the Cartesian coordinate lines by the	Physical interpretation of the red curves	Plot

curves: u = constant
blue curves: v = constant)

conformal map f defined by

1

initial curves (red circles)			orthogonal curves (blue lines)		
	Cartesian	polar		Cartesian	polar
implicit equation	$x^2 + y^2 =$ cte	$\rho = a$	implicit equation	$y = kx$	$\theta = \theta_0$
differential equation	$yy' + x = 0$	$\rho' = 0$	differential equation	$xy' - y = 0$	$d\theta/d\rho = 0$
field	$(y, -x)$	$(0, 1)$	field	(x, y)	$(1, 0)$

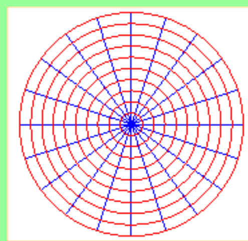
$\begin{cases} x = u\cos \\ y = u\sin \end{cases}$ so

$f(z) = \ln$

$f^{-1}(z) = e$

Magnetic field lines induced by a uniform linear current orthogonal at O to xOy.

Electrostatic equipotential induced by a charge placed at O or charges uniformly distributed on a line orthogonal at O to xOy



2
Family of homofocal parabolas

initial curves (red parabolas)	orthogonal curves (blue parabolas)

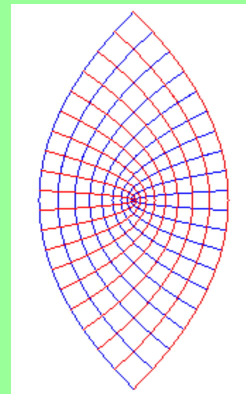
Initial curves:

	Cartesian	polar
implicit equation	$y^2 = 4u^2(u^2 - x)$	□ = $2u^2$ / (1+cos □)
differential equation	$yy'^2 + 2xy' - y = 0$	

Orthogonal curves:

	Cartesian	polar
implicit equation	$y^2 = 4v^2(v^2 + x)$	□ =-$2v^2$/ (1+cos θ)
differential equation	$yy'^2 - 2xy' - y = 0$	

$\begin{cases} x = u^2 - \\ y = 2uv \end{cases}$ so $f(z) = \sqrt{z}$ $f^{-1}(z) = z$



3

initial curves (red hyperbolas)	orthogonal curves (blue hyperbolas)

Initial curves:

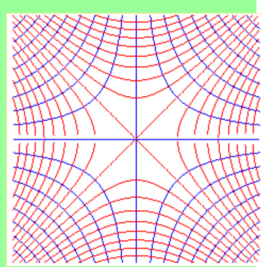
	Cartesian	polar
impl	x^2 -	□ ²c

Orthogonal curves:

	Cartesian	polar
impl	xy	□ ²

$\begin{cases} x = \pm\sqrt{\sqrt{}} \\ y = \pm\sqrt{\sqrt{}} \end{cases}$ so $f(z) = z^2$ $f^{-1}(z) = \sqrt{}$

Approximate view of the example n° 8 below in a neighbourhood of *O*.



icit equation	$y^2 = cte$	os2 θ = cte
differential equation	$yy' - x = 0$	□ ' = □ □ tan 2 θ
field	(y, x)	(sin 2 θ , cos 2 θ)

icit equation	= cte	sin 2 θ = cte
differential equation	$xy' + y = 0$	□ ' = - □ □ cot 2 θ
field	(- x, y)	(- cos2 θ , sin 2 θ)

They are the contour lines of the hyperbolic paraboloid

4

initial curves (red circles)	orthogonal curves (blue circles)

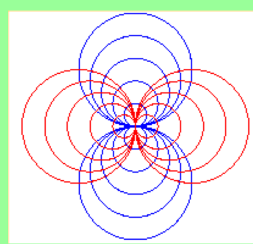
	Cartesian	polar
implicit equation	$x^2 + y^2 = ax$	□ = $a$cos

	Cartesian	polar
implicit equation	$(x^2 + y^2) = a$	□ = a sin θ

$$\begin{cases} x = \dfrac{u}{u^2 +} \\ y = \dfrac{v}{u^2 +} \end{cases} \quad f(z) = f^{-1}(z \quad = 1/z$$

limit case of the example n°7 below when the conductors are infinitely close.

See also the Smit



= two pencils of orthogonal singular circles

		θ
differential equation	$2xy\,y' = y^2 - x^2$	□ ' + □ □ □ tan θ □ □ □ □
field	$(2xy, y^2 - x^2)$	(sin θ, - cos θ)

	y	
differential equation	$(y^2 - x^2)\,y' + 2xy = 0$	□ ' = □ □ cot θ
field	$(x^2 - y^2, 2xy)$	(cos θ □ □ sin θ)

h chart.

5

initial curves (red cardioids)	orthogonal curves (blue cardioids)

	polar
implicit equation	□ □ $a$cos$^2\theta$ □ □
differential equation	□ ' = □ □ tan θ

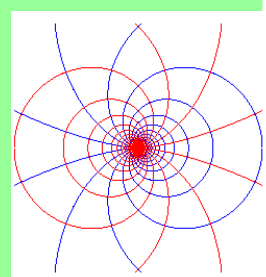
	polar
implicit equation	□ □ a sin^2 θ □ □
differential equation	□ ' = - □ □ cot θ

$$\begin{cases} x = \frac{u^2}{(u^2 \cdots} \\ y = \frac{2}{(u^2 \cdots} \end{cases}$$

$f(z) = \frac{1}{\sqrt{z}}$

$f'(z) = -\frac{1}{z^{\cdots}}$

Remark: figure obtained by inversion of that of the example n° 2.



	□ □
field	(sinθ □ □ , cosθ □ □)

	□ □
field	(-cos θ □ □, sinθ □ □)

6

initial curves (red sinusoidal spirals of index -*n*)	orthogonal curves (blue sinusoidal spirals of index -*n*)

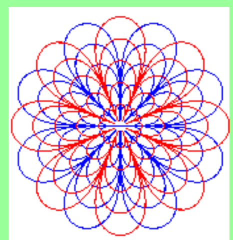
	polar
implicit equation	□^{n}cos $n\theta$ = cte
differential equation	□ ' = □ □tan $n\theta$
field	(sin$n\theta$, cos$n\theta$)

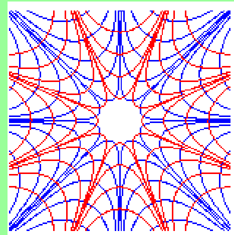
	polar
implicit equation	□^{n}sin$n\theta$ = cte
differential equation	□ ' = - □ □ cot$n\theta$
field	(-cos$n\theta$, sin$n\theta$)

The four previous cases correspond

$f(z)=z^n$ so $f^{-1}(z)=z$

Opposite, view for *n* = 4 and *n* = -4.





to n = 1/2, n = 2, n = -1 , n =-1/2

7

initial curves (red double eggs)	orthogonal curves (blue curves of the dipole)

Initial curves:

	Cartesian	polar
implicit equation	$(x^2 + y^2)^3 = a^2 x^4$	□ = $a\cos 2\theta$
differential equation	$3xy\, y' = 2y^2 - x^2$	□ ' + 2□ □ tan□ □ □ □ □
field	$(3xy, 2y^2 - x^2)$	$(2\sin\theta, -\cos\theta)$

Orthogonal curves:

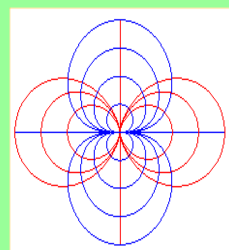
	Cartesian	polar
implicit equation	$(x^2 + y^2)^3 = a^4 y^2$	□ $^2 = a^2 \sin\theta$
differential equation	$(2y^2 - x^2)\, y' + 3xy = 0$	□ □ ' = □ □ $\cot\theta$
field	$(x^2 - 2y^2, 3xy)$	$(\cos\theta$ □ □ $2\sin\theta)$

Remark: the examples 4 and 7 are part of the more general example

???? ????

Field lines of a magnetic dipole

Field lines of an electrostatic dipole (limit case, inverting the red and blue curves, of the example 9 below).



of Clairaut's curves: □ = $a\cos^n\theta$ and □^{n}= $a^n\sin\theta$ □

8

initial curves (red circles)	orthogonal curves (blue circles)

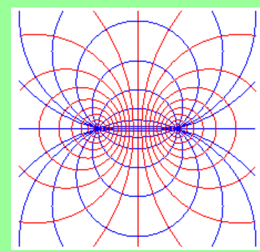
geometrical definition	MA/MB = constant with A(1,0) and B(-1,0)
implicit equation	$(x-1)^2+y^2$ = cte. $((x+1)^2+y^2)$

geometrical definition	(MA, MB) = constant
field	**MA**/MA² - **MB**/MB²

$$\begin{cases} x = \dfrac{s}{\operatorname{ch} u} \\ y = \dfrac{s}{\operatorname{ch} u} \end{cases}$$

$f(z) = \arg \operatorname{th}($ $= \dfrac{1}{2}\ln\dfrac{1+z}{1-z}$ so $f^{-1}(z) = \operatorname{th}(z)$ $= \dfrac{e^{2z}-1}{e^{2z}+1}$

Magnetic field lines induced by a uniform linear current orthogonal at *A* to *xOy* and a parallel current, in the opposite direction, passing by *B*. Electrostatic equipotential lines induced by charges uniformly distributed on a line orthogonal



= two pencils of orthogonal circles

at *A* to *x Oy* and opposite charges uniformly distributed on a line orthogonal at *B* to *x Oy*.

9

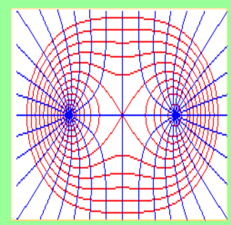
initial curves (red Cassinian ovals)	orthogonal curves (blue rectangular hyperbolas)
geometrical definition: MA.MB = constant	geometrical definition: (*Ox*, AM) + (*Ox*, BM) = constant
	field: **MA**/MA² - **MB**/MB²
polar implicit equation: $\rho^4 - 2\rho^2 \cos 2\theta$ = cte	polar implicit equation: $\rho^2 = \cos 2\theta$ □ □
differential: □ '(□ □ □ □ c	

$x = \pm\sqrt{\sqrt{e^{2u} \ldots} + e^{u} \cos \ldots}$, $y = \pm\sqrt{\sqrt{e^{2u} \ldots} - e^{u} \cos \ldots}$

$f(z) = \ln(z^2 - \ldots$ so $f^{-1}(z) = \sqrt{e^z + \ldots}$

Magnetic field lines induced by a uniform linear current orthogonal at *A* to *x Oy* and a parallel current, in the same direction, passing by *B*.

Electrostatic equipotential



See a generalisation at Cassinian curve for the red curves, and at stelloid for the blue curves: case where $f(z) = \sum_i \alpha_i \ln(z - a_i \ldots$.

equation	os2θ □ □ □ sin 2θ =0
field	(sin2θ □ □ cos 2θ □ □ □ □ ²)

ion	cos(2(θ □ □ □ □
differential equation	sin2θ □ '=□³ □ □ □ □ cos2□
field	(□ ²□ □ □ cos2θ , sin2θ)

lines induced by charges uniformly distributed on a line orthogonal at *A* to *x Oy* and equal charges uniformly distributed on a line orthogonal at *B* to *x Oy*.

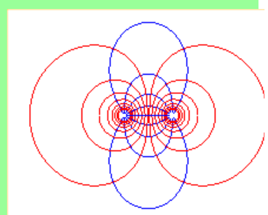
10

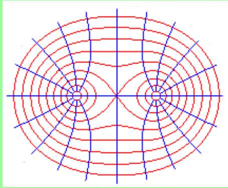
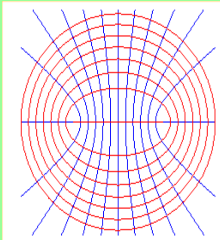
initial curves (red Cayley equipotential lines)	orthogonal curves (in blue)

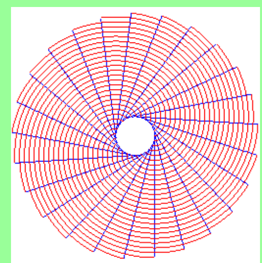
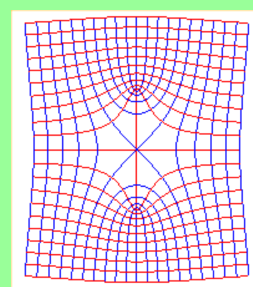
geometrical definition	1/MA-1/MB = cte
field	$\left(\frac{1}{MA}-\frac{1}{MB}\right)\overrightarrow{AB}-\frac{\overrightarrow{MA}\,AB}{MA^3}$

geometrical definition	$\left(\frac{\overrightarrow{MA}}{MA}-\frac{\overrightarrow{MB}}{MB}\right)$
field	**MA**/MA³ - **MB**/M

Electrostatic equipotential lines induced by two opposite charges placed at *A* and *B*, in other words, an electrostatic



B^3			dipole.	
11 initial curves (red [Cayley ovals](#)): geometrical definition: 1/MA + 1/MB = constant; field: $\left(\frac{1}{MA}+\frac{1}{MB}\right)\overrightarrow{AB}-\frac{\overrightarrow{MA}\,\overrightarrow{AB}}{MA^3}$ orthogonal curves (in blue): geometrical definition: $\left(\frac{\overrightarrow{MA}}{MA}+\frac{\overrightarrow{MB}}{MB}\right)$; field: **MA**/MA^3 + **MB**/MB^3			Electrostatic equipotential lines induced by two equal charges placed at *A* and *B*.	
12 Lattice of homofocal conics initial curves (red ellipses): geometrical definition: MA + MB = constant; field: **MA**/MA - **MB**/MB; Cartesian: x²/(1+cte) +y²/cte=1 orthogonal curves (blue hyperbolas): geometrical definition: MA - MB = constant; field: **MA**/MA + **M**	$\begin{cases} x = \mathrm{ch}\, u \\ y = \mathrm{sh}\, u \end{cases}$	*f*(*z*) = argcosh(*z*) *f*$^{-1}$(*z*) = cosh (*z*) (image of the first lattice by the [Joukovski](#) transformation: *j*(*z*) = (*z* + 1/*z*)/2)	Electrostatic equipotential lines induced by charges uniformly distributed on the segment line [AB]??? Interference	

equation | | **B**/MB Cartesian equation | $x^2/(1-cte)-y^2/cte=1$			pattern	
13 Involute of circles and their generatrices initial curves (red half-involutes of a fixed circle) | orthogonal curves (blue half-tangents to the circle) complex parametrization | $z = e^{i\alpha}($ complex parametrization | $z = e^{i\alpha}($	????	????		
14 initial curves (red quartics) | orthogonal curves (blue quartics) | Cartesian implicit equation | $y^2 = a^2(1$ | Cartesian implicit equation | $x^2 = a^2(1$		$f(z) = f^{-1} = i\sqrt{z^2+1}$	Streamlines of a uniform flow perturbed by an obstacle (the segment line [AB] with	 The blue curve passing by *O* (obtained for *a* = 1) is a bullet nose

	$+ 1 / (a^2+x^2))$
parametrization	$x = a \tan t$ $y^2 = a^2 + \cos^2 t$

	$- 1 / (a^2+y^2))$
parametrization	$x^2 = a^2 - \cos^2 t$ $y = a \tan t$

A(0, 1) and B(0, -1))

curve

15

initial curves (red cubic hyperbolas)	orthogonal curves (blue cubic hyperbolas)

	Cartesian
implicit equation	$(y - \text{constant}) (x^2+y^2) = y$

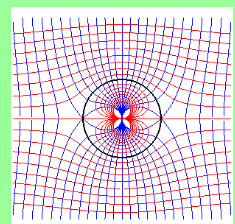
	Cartesian
implicit equation	$(x - \text{constant}) (x^2+y^2) + x = 0$

$f(z) = (z+1$
$= j(z)/i$

where j is the Joukovski transformation

$f^{-1}(z) = iz + i$

Streamlines of a uniform flow perturbed by the disk with centre O and radius 1.



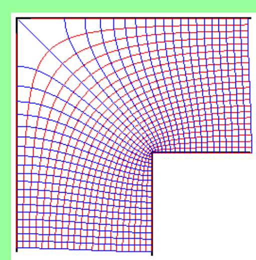
16

$$\begin{cases} x = \ln\left(\dfrac{\sqrt{A}+\sqrt{B}+\sqrt{e^{2u}+\sqrt{AB}-1}}{2}\right) - \text{arcc} \\ y = -\ln\left(\dfrac{\sqrt{C}+\sqrt{D}+\sqrt{e^{-2u}+\sqrt{CD}-1}}{2}\right) + \text{a} \end{cases}$$

where

$f^{-1}(z) =$
$\text{argch}\, e^z - \text{ar}$

Streamlines of a uniform flow in a bent tube.



$A = \left(e^{u} + \cos v\right)^{2} + \sin^{2} v, B = \left(e^{u} - \text{co}\right.$

$C = \left(e^{-u} + \cos v\right)^{2} + \sin^{2} v, D = \left(e^{-u} - \right.$

.

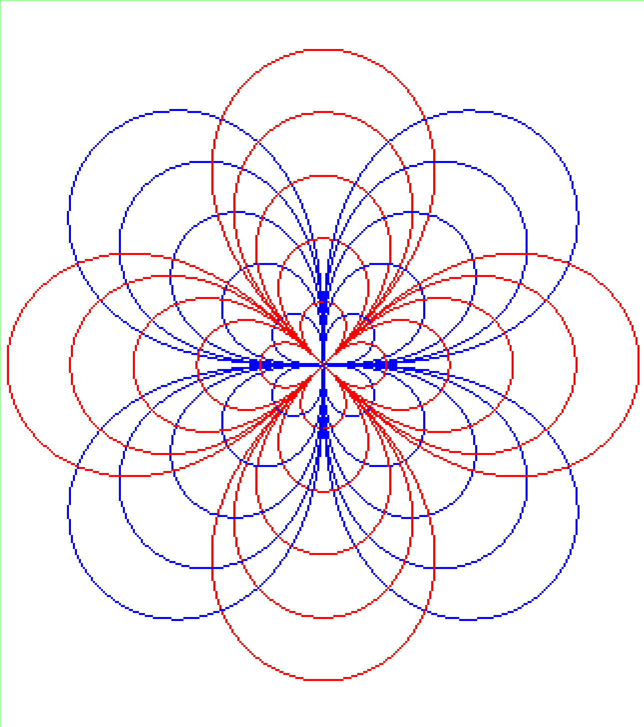
For example, when *v* = pi/2, we get:

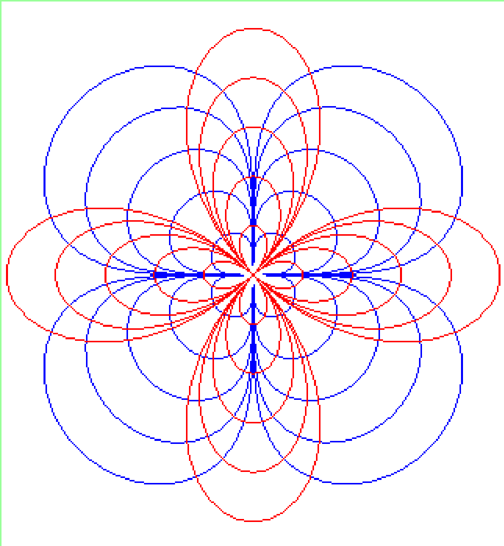
$$\begin{cases} x = \operatorname{argsh} e^{u} - \dfrac{\pi}{2} \\ y = \operatorname{argsh} e^{-u} + \dfrac{\pi}{2} \end{cases}$$

Other examples:

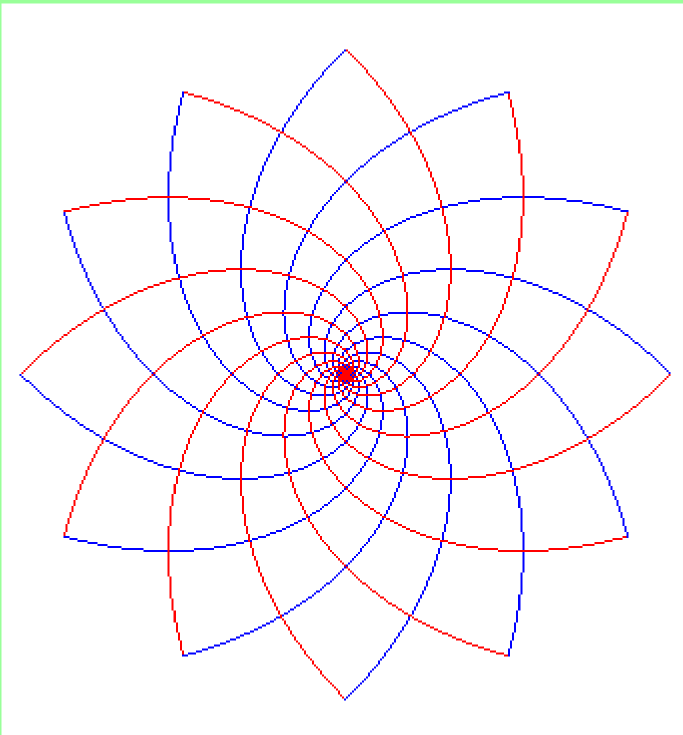
Lemniscates of Bernoulli $\rho^{2} = \alpha \cos 2\theta$ and $\rho^{2} = \alpha \sin 2\theta$ (case *n* = –2 of the example 6 above)

Quatrefoils $\rho = \alpha \cos 2\theta$ and their orthogonal trajectories $\rho^{4} = \alpha \sin 2\theta$.

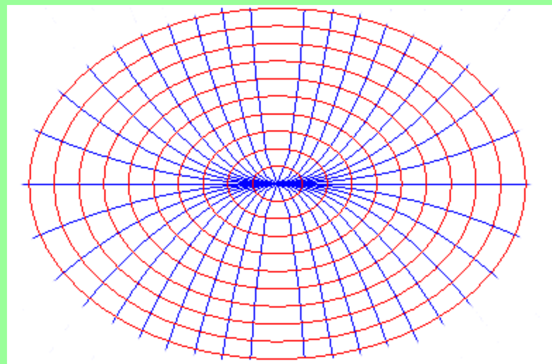




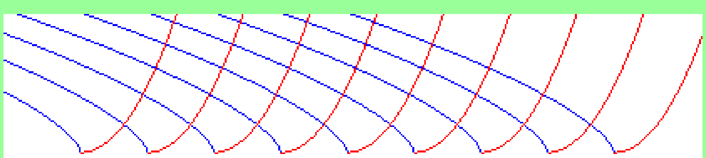
Logarithmic spirals $\rho = \alpha e^{\theta}$ and $\rho = \alpha e^{-\theta}$.



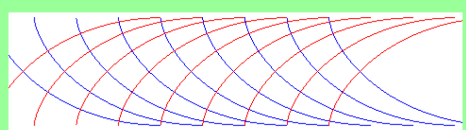
Parabolas $y = \alpha x^2$ and ellipses $2x^2 + y^2 = \alpha$



Red parabolas $y = (x - \alpha)^2$ and semicubical parabolas $4y^3 = 9(x - \alpha)^2$.



cycloids $\begin{cases} x = t - \sin t + \alpha \\ y = 1 - \cos t \end{cases}$ and symmetric cycloids $\begin{cases} x = t - \sin t + \alpha \\ y = 1 + \cos t \end{cases}$

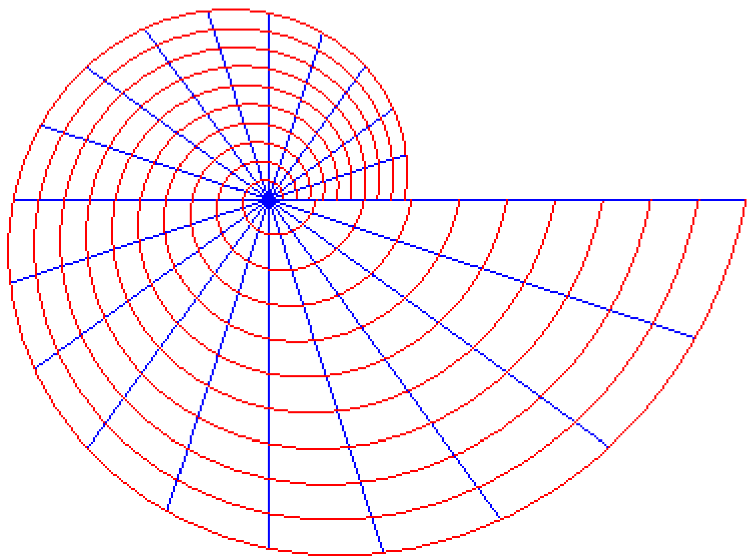


See also tractrix, as well as this article.

The projection on a horizontal plane of the slope and contour lines of a surface form two orthogonal lattices; see for example the egg box.

This notion of orthogonal curves can be generalised to any angle; two families of curves intersect under the angle V when, at each point common to the families, the tangents form an angle V, and one of the families is composed of the *trajectories at angle V* of the other one.

For example, the trajectories under the angle V of the pencil of lines passing by O are the logarithmic spirals $\rho = ae^{\frac{\theta}{\tan V}}$:



The 3D generalisation of the double orthogonal systems is the notion of triple orthogonal system of surfaces.

UNIT 11 PARTIAL DIFFERENTIATION

11.1 Differentiation of Function of Two Variables

- Define a function of two variables.
- Define partial derivative
- Find partial derivatives of a function of two variables.

Functions of Two Variables

Aim

To demonstrate how to differentiate a function of two variables.

Learning Outcomes

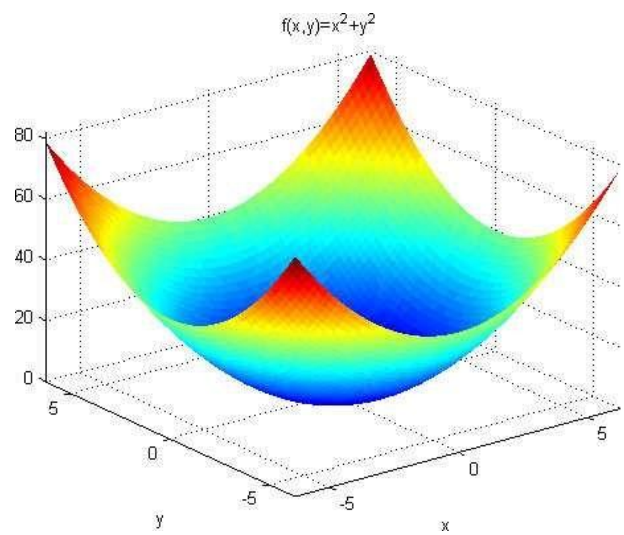
At the end of this section you will:

- Know how to recognise a function of two variables,
- Know how to differentiate functions of two variables.

We have already studied functions of one variable, which we often wrote as $f(x)$. We will now look at functions of two variables, $f(x,y)$. For example,

$$z = f(x,y) = x^2 + y^2.$$

We know that the graph of a function of one variable is a curve. The graph of a function of two variables is represented by a surface as can be seen below. The graph of a function of two variables will always be drawn in three dimensions.



Similar to the definition of a function that we have previously seen, a function of two variables can be defined as a rule that assigns to each incoming *pair* of numbers, (x,y), a uniquely defined outgoing number, z. Therefore, in order to be able to evaluate the function we have to specify the numerical values of both x and y.

Partial Differentiation

Given a function of two variables, $z = f(x,y)$ we can determine two first-order derivatives. The **partial derivative** of f with respect to x is written

$$\frac{\partial z}{\partial x} \quad \text{or} \quad \frac{\partial f}{\partial x} \quad \text{or} \quad f_x$$

and is found by differentiating f with respect to x, with y held constant. Similarly the partial derivative of f with respect to y is written

$$\frac{\partial z}{\partial y} \quad \text{or} \quad \frac{\partial f}{\partial y} \quad \text{or} \quad f_y$$

and is found by differentiating f with respect to y, with x held constant. We use the partial symbol, ∂, to distinguish partial differentiation of functions of several variables from ordinary differentiation of functions of one variable.

Example 1

Differentiate the following function with respect to x,

$$f(x,y) = x^2 + y^3$$

By the sum rule we know that we can differentiate each part separately and then add the solutions together. When we differentiate x^2 with respect to x we get $2x$. When we differentiate y^3 with respect to x we get 0. To see this, note that y is treated like a constant when differentiating with respect to x. Therefore any function of y, e.g. y^3 is also treated like a constant when differentiating with respect to x and, as we already know, the differential of a constant is 0. Therefore

$$\frac{\partial f}{\partial x} = 2x + 0 = 2x.$$

Example 2

Find both first-derivatives of the following function,

$$f(x,y) = x^2y$$

Care must be taken in this case because here we have a term consisting of both x and y. To find f_x we differentiate as normal taking x as the variable and y as the constant.

Remember that when we differentiate a constant times a function of x we differentiate the function of x as normal and then multiply it by the constant. For example,

$$3x^2 \text{ differentiates to give } \quad 3(2x) = 6x.$$

In our situation, y plays the role of a constant, so

$$x^2y \quad \text{differentiates to give} \quad (2x)y = 2xy.$$

Hence $f_x = 2xy$.

Similarly, to find f_y we treat y as the variable and x as the constant. When we differentiate a constant times y we just get the constant. In our case x^2 plays the role of the constant, so x^2y differentiates to give x^2. Hence,

$$f_y = x^2.$$

It is possible to find second-order derivatives of function of two variables. There are four second-order partial derivatives. The four derivatives are

$$fxx, \quad fyy, \quad fxy \quad \text{and} \quad fyx.$$

In general it is true that $f_{xy} \equiv f_{yx}$.

Note: f_{xy} means that we differentiate the function f first with respect to x and then we different the resulting answer, f_x, with respect to y.

Small Increments Formula

To provide an interpretation of a partial derivative let us take one step back for a moment and recall the corresponding situation for functions of one variable of the form

$$y = f(x).$$

The derivative, $\frac{dy}{dx}$, gives the rate of change of y with respect to x. In other words, if x changes by a small amount Δx then the corresponding changes in y satisfies

$$\Delta y \simeq \frac{dy}{dx}\Delta x.$$

The accuracy of the approximation improves as Δx becomes smaller and smaller. Given the way in which a partial derivative is found we can deduce that for a function of two variables

$$z = f(x,y)$$

if x changes by a small amount Δx and y is held fixed then the corresponding change in z satisfies

$$\Delta z \simeq \frac{\partial z}{\partial x}\Delta x.$$

Similarly, if *y* changes by Δy and *x* is fixed then *z* changes by

$$\Delta z \simeq \frac{\partial z}{\partial y}\Delta y.$$

In practice, of course, *x* and *y* may both change simultaneously. If this is the case then the net change in *z* will be the sum of the individual changes brought about by changes in *x* and *y* separately, so that

$$\Delta z \simeq \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y.$$

This is referred to as the small increments formula or the **total derivative**. If Δx and Δy are allowed to tend to zero then the above formula (which was only an approximation) can be rewritten as

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

where the symbols *dx*, *dy* and *dz* are called differentials and represent limiting values of Δx, Δy and Δz, respectively.

Related Reading

Jacques, I. 1999. *Mathematics for Economics and Business.* 3rd Edition. Prentice Hall.

11.2 Euler's Theorem

- Define a homogeneous function of degree n.
- Prove Euler's theorem on homogeneous functions.
- Prove (by Euler's theorem) homogeneous functions of different degrees (simple cases).
- Apply MAPLE command diff to find partial derivatives.

Euler's Theorem

The generalization of Fermat's theorem is known as Euler's theorem. In general, Euler's theorem states that, "if *p* and *q* are relatively prime, then $p^{\varphi(q)} = 1(\bmod q)$", where φ is Euler's totient function for integers. That is, $\varphi(q)$ is the number of non-negative numbers that are less than q and relatively prime to q.

Proof of Euler's theorem:

Consider the set of non-negative numbers,
$P=\{n_1,n_2,n_3,\cdots,n_{\varphi(q)}\}(\bmod q)$
These elements are relatively (co-prime) to q.

Consider another set of non-negative numbers,

$P_1=\{pn_1,pn_2,pn_3,\cdots,pn_{\varphi(q)}\}(\bmod q)$ where $(p,q)=1$

Since the sets are congruent to each other,
$$n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\equiv pn_1\cdot pn_2\cdot pn_3\cdots pn_{\varphi(q)}(\bmod q)$$
$$n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\equiv p\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)(\bmod q)$$
$$p^{\varphi(q)}\cdot\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)\equiv\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)(\bmod q)$$
Since the set of numbers are relatively prime to q, dividing by the term $\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)$ is permissible.
$$\frac{p^{\varphi(q)}\cdot\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}{\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}\equiv\frac{\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}{\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}(\bmod q)$$
$$p^{\varphi(q)}\equiv\frac{\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}{\left(n_1\cdot n_2\cdot n_3\cdots n_{\varphi(q)}\right)}(\bmod q)$$
$$p^{\varphi(q)}\equiv 1(\bmod q)$$

diff or Diff
differentiation or partial differentiation

[Calling Sequence](#)
[Parameters](#)
[Description](#)
[Examples](#)

Calling Sequence

diff(**f**, **x1**, ..., **xj**)	d*j*d*xj*...d*x1f*
diff(**f**, [**x1$n**])	d*n*d*xn*1*f*
diff(**f**, **x1$n**, [**x2$n**, **x3**], ..., **xj**, [**xk$m**])	d*r*d*xmk*d*xj*...d*x*3d*xn*2d*xn*1*f*

Remark: these calling sequences are also valid with the inert Diff command
Parameters

f -algebraic expression or an equation
x1, x2, ..., xj -names representing differentiation variables
n -algebraic expression entering constructions like **x$n**, representing **n**th order derivative, assumed to be integer order differentiation

Description

- The **diff** command computes the partial derivative of the expression **a** with respect to **x1, x2, ..., xn**, respectivel most frequent use is **diff(f(x),x)**, which computes the derivative of the function **f(x)** with respect to **x**.
- You can enter the command **diff** using either the 1-D or 2-D calling sequence. For example, **diff(x, x)** is equiva to dd*xx*.
- Note that where **j** in **xj** is greater than **1**, the call to **diff** is the same as diff called recursively. Thus **diff(f(x1,x2) x2);** is equivalent to the call **diff(diff (f(x1,x2), x1), x2)**. The sequence operator $ is useful for forming higher-o derivatives. **diff(f(x),x$4)**, for example, is equivalent to **diff(f(x),x,x,x,x)** and **diff(g(x,y),x$2,y$3)** is equivalent to **diff(g(x,y),x,x,y,y,y)**
- The names with respect to which the differentiation is to be done can also be given as a list of names. This forn allows for the special case of differentiation with respect to no variables, in the form of an empty list, so the *zeroth* order derivative is handled through **diff(f,[x$0]) = diff(f,[])**. In this case, the result is simply the orig expression, **f**. This format is especially useful when used together with the sequence operator and sequences wi potentially zero variables.
- Derivatives of **n**th order, where **n** is not specified as a number, can be constructed as in **diff(f(x),[x$n])** and are interpreted as *integer order* derivatives, that is, computed assuming **n** is an integer. The routines for computing these symbolic nth order derivatives can handle most functions of the mathematical language and an increasing number of expressions formed by composing them with other functions or algebraic expressions. The results ar returned in closed form or as finite sums - see the **Examples** section.
- **diff** has a user interface that will call the user's own differentiation functions. If the procedure **`diff/f`** is define the function call **diff(f(x, y, z), y)** will invoke **`diff/f`(x,y,z,y)** to compute the derivative. See example below.
- If the derivative cannot be expressed (if the expression is an undefined function), the diff function call itself is returned. (The prettyprinter displays the diff function in a two-dimensional dd*x* format.)
- The **diff** command assumes that partial derivatives commute.

- The capitalized function name **Diff** is the inert **diff** function, which simply returns unevaluated. It appears gray it is easily distinguished from a returned **diff** calling sequence.
- The differential operator **D** is also defined in Maple; see D. For a comparison of **D** and **diff** see operators[D].

Examples

Compute first order derivatives.

> dd*xx*

1**(1)**

> dd*xex*

e*x***(2)**

> dd*x*sin(*x*)

cos(x)(3)

> ddxtan(x)

1+tan(x)2(4)

> ddxx2−−−√

xx2−−−√(5)

> ddx(xsin(cos(x)))

sin(cos(x))−xsin(x)cos(cos(x))(6)

Find higher order derivatives.
> d3dx3sin(x)

−cos(x)(7)

> d2dx2(3x3+2x2+23x+2342)

18x+4(8)

Compute partial derivatives.
> ∂∂ysin(x)

0(9)

> ∂∂x(x2+xy2)

y2+2x(10)

> ∂∂y(2x+y2)

2y(11)

> ∂2∂y∂x(x2+xy2)

2y(12)

> h≔5x2+2x2y+3xy2+12yx+3y3x

h≔5x2+2x2y+3xy2+12yx+3y3x(13)

> ∂2∂y∂xh

4x+6y+12−9y2x2(14)

> ∂2∂x∂yh

4x+6y+12−9y2x2(15)

> diff(h,y$3)

18*x*(16)

The **Diff** command is inert, it returns unevaluated.
Note: To enter the 2-D calling sequence of the **Diff** command, type **Diff** at the input, press **Esc**, and then select **Diff(inline)**.

> ddxtan(*x*)

ddxtan(*x*)**(17)**

> ddxtan(*x*)=ddxtan(*x*)

ddxtan(*x*)=1+tan(*x*)2**(18)**

> ddx*f*(*x*)

ddx*f*(*x*)**(19)**

> ∂2∂*y*∂*xf*(*x*,*y*)

∂2∂*x*∂*yf*(*x*,*y*)**(20)**

> ∂2∂*y*∂*xf*(*x*,*y*)−(∂2∂*x*∂*yf*(*x*,*y*))

0**(21)**

An empty list specifies no derivatives:

> diff(*g*(*x*,*y*,*z*),[])

g(*x*,*y*,*z*)**(22)**

Teach Maple how to differentiate *f*(*g*(*x*)) = ddx*g*(*x*)*f*(*x*)2

> **`diff/f` := proc(g,x) diff(g,x)/f(x)^2 end proc:**

> ddx*f*(sin(*x*))

cos(*x*)*f*(*x*)2**(23)**

Symbolic order differentiation is also handled. For example, for arbitrary integer values of **n**,

> Diff(sin(*x*),*x*$*n*)

d*n*d*xn*sin(*x*)**(24)**

Inert objects can be evaluated with the value command.

> value()

sin(*x*+*n*π2)**(25)**

Note that in the context of a call to **diff** (or **Diff**), **n** entering d*n*d*xnf*(*x*) is understood to be an *integer*; that is: **diff** computes *integer order* derivatives. To compute fractional derivatives see fracdiff.
A more involved example

> Diff(exp(*x*2),*x*$*n*)

dndxnex2**(26)**

\>   value()

x−n2nMeijerG([[0,12],[]],[[0],[12+n2,n2]],−x2)**(27)**

The Leibniz rule for the **n**th derivative of a product
\>   Diff(*f*(*x*)*g*(*x*),*x*$*n*)

dndxn(f(x)g(x))**(28)**

\>   value()

∑n_k1=0(n_k1)(d_k1dx_k1f(x))(dn−_k1dxn−_k1g(x))**(29)**

UNIT 12 INTRODUCTION TO NUMERICAL METHODS

12.1 Numerical Solution of Non –linear Equations

- Explain importance of numerical methods.
- Explain the basic principles of solving a non – linear equation in one variable.
- Evaluate real roots of non – linear equation in one variable by
    - I. Bisection method,
    - II. Regula – falsi method,
    - III. Newton – Raphson method
- Apply MAPLE command fsolve to find numerical solution of an equation and demonstrate through examples

Appendix 8
Numerical Methods for Solving
Nonlinear Equations
1

An equation is said to be nonlinear when it involves terms of degree higher than 1 in the unknown quantity. These terms may be polynomial or capable of being broken down into Taylor series of degrees higher than 1.

Nonlinear equations cannot in general be solved analytically. In this case, therefore, the solutions of the equations must be approached using iterative methods. The principle of these methods of solving consists in starting from an arbitrary point – the closest possible point to the solution sought – and involves arriving at the solution gradually through successive tests.

The two criteria to take into account when choosing a method for solving nonlinear equations are:

- Method convergence (conditions of convergence, speed of convergence etc.).

- The cost of calculating of the method.

8.1 GENERAL PRINCIPLES FOR ITERATIVE METHODS

8.1.1Convergence

Any nonlinear equation $f(x) = 0$ can be expressed as $x = g(x)$.

If x_0 constitutes the arbitrary starting point for the method, it will be seen that the solution x^* for this equation, $x^* = g(x^*)$, can be reached by the numerical sequence:

$$x_{n+1} = g(x_n) \quad n = 0,1,2,...$$

[1]

This iteration is termed a Picard process and x^*, the limit of the sequence, is termed the fixed iterative point.

In order for the sequence set out below to tend towards the solution of the equation, it has to be guaranteed that this sequence will converge. A sufficient condition for convergence is supplied by the following theorem: if $x = g(x)$ has a solution a within the interval $I = [a - b;a + b] = \{x : |x - a| \leq b\}$ and if $g(x)$ satisfies Lipschitz's condition:

$$\exists L \in [0;1[: \forall x \in I, \quad |g(x) - g(a)| \leq L|x - a|$$

Then, for every $x_0 \in I$:

- all the iterated values x_n will belong to I;
- the iterated values x_n will converge towards a; • the solution a

  will be unique within interval I.

We should also show a case in which Lipschitz's condition is satisfied: it is sufficient that for every $x \in I$, g'(x) exists and is such that $|g'(x)| \leq m$with m <1.

[1] This appendix is mostly based on Litt F. X., *Analyse nume´rique, premi`ere partie*, ULG, 1999. Interested readers should also read: Burden R. L. and Faires D. J., *Numerical Analysis*, Prindle, Weber & Schmidt, 1981; and Nougier J. P., *Me´thodes de calcul nume´rique*, Masson, 1993.

8.1.2 Order of convergence

It is important to choose the most suitable of the methods that converge. At this level, one of the most important criteria to take into account is the speed or order of convergence.

Thus the sequence x_n, defined above, and the error $e_n = x_n - a$. If there is a number p and a constant C >0 so that

$$\lim_{n \to \infty} \frac{|e_{n+1}|}{|e_n|^p} = C$$

p will then be termed the order of convergence for the sequence and C is the asymptotic error constant.

When the speed of convergence is unsatisfactory, it can be improved by the Aitken extrapolation,[2] which is a convergence acceleration process. The speed of convergence of this extrapolation is governed by the following result:

- If Picard's iterative method is of the order p, the Aitken extrapolation will be of the order $2p - 1$.
- If Picard's iterative method is of the first order, Aitken's extrapolation will be of the second order in the case of a simple solution and of the first order in the case of a multiple solution. In this last case, the asymptotic error constant is equal to $1 - 1/m$ where m is the multiplicity of the solution.

8.1.3 Stop criteria

As stated above, the iterative methods for solving nonlinear equations supply an approached solution to the solution of the equation. It is therefore essential to be able to estimate the error in the solution.

Working on the mean theorem:

$$f(x_n) = (x_n - a) f'(\xi), \text{ with } \xi \in [x_n; a]$$

we can deduce the following estimation for the error:

$$|x_n - a| \le \frac{|f(x_n)|}{M}, \quad |f'(x_n)| \ge M, \quad x \in [x_n; a]$$

[2] We refer to Litt F. X., *Analyse nume´rique, premi`ere partie*, ULG 1999, for further details.

In addition, the rounding error inherent in every numerical method limits the accuracy of the iterative methods to:

$$\varepsilon_a = \frac{\delta}{f'} \quad \text{(a)}$$

in which δ represents an upper boundary for the rounding error in iteration n:

$$\bar{\delta} \geq |\delta_n| = f(x_n) - \bar{f}(x_n)$$

$\bar{f}(x_n)$ represents the calculated value for the function.

Let us now assume that we wish to determine a solution a with a degree of precision ε. We could stop the iterative process on the basis of the error estimation formula.

These formulae, however, require a certain level of information on the derivative $f'(x)$, information that is not easy to obtain. On the other hand, the limit specification ε_a will not generally be known beforehand.[3] Consequently, we are running the risk of ε, the accuracy level sought, never being reached, as it is better than the limit precision $\varepsilon_a(\varepsilon < \varepsilon_a)$. In this case, the iterative process will carry on indefinitely. This leads us to accept the following stop criterion:

$$\begin{cases} |x_n - x_{n-1}| < \varepsilon \\ |x_{n+1} - x_n| \geq |x_n - x_{n-1}| \end{cases}$$

This means that the iteration process will be stopped when the iteration n produces a variation in value less than that of the iteration n + 1. The value of ε will be chosen in a way that prevents the iteration from stopping too soon.

8.2 PRINCIPAL METHODS

Defining an iterative method is based ultimately on defining the function h(x) of the equation $x = g(x) \equiv x - h(x)f(x)$.

The choice of this function will determine the order of the method.

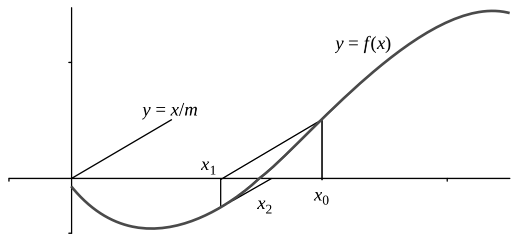
8.2.1First order methods

The simplest choice consists of taking h(x) = m = constant = 0.

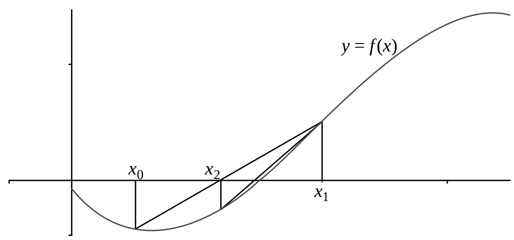
[3] This will in effect require knowledge of $f'(a)$, when a is exactly what is being sought.

8.2.1.1Chord method

This defines the *chord method* (Figure A8.1), for which the iteration is $x_{n+1} = x_n - mf(x_n)$.



FigureA8.1 Chordmethod



FigureA8.2 Classicchordmethod

The sufficient convergence condition (see Section A8.1.1) for this method is $0 < mf'(x) < 2$, in the neighbourhood of the solution. In addition, it can be shown that

$\lim_{n\to\infty} \frac{n}{} \quad g'(a) \quad 0.$|e +|1| = | |=

|en

The chord method is therefore clearly a first-order method (see Section A8.1.2).

8.2.1.2Classic chord method

It is possible to improve the order of convergence by making m change at each iteration:

$$x_{n+1} = x_n - m_n f(x_n)$$

The *classic chord method* (Figure A8.2) takes as the value for m_n the inverse of the slope for the straight line defined by the points $(x_{n-1};f(x_{n-1}))$ and $(x_n;f(x_n))$:

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n)$$

This method will converge if $f'(a)$= 0 and f''(x) is continuous in the neighbourhood of a. In addition, it can be shown that

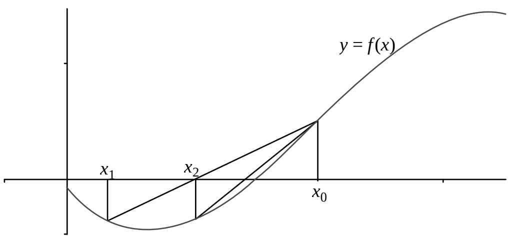
$$\frac{e_{n+1}|}{e_n{}^p} = \frac{f(a)}{2f(a)}^{1/p} \quad \lim || \quad | \quad = 0 \; {}^{n\to\infty}$$

for p $= \frac{1}{2}(1+\sqrt{5}) =$ 1.618... >1, which greatly improves the order of convergence for the method.

8.2.1.3Regula falsi method

The *regula falsi* method (Figure A8.3) takes as the value for m_n the inverse of the slope for the straight line defined by the points ($x_{n'}$; $f(x_{n'})$) and (x_n;f(x_n)) where nis the highest index for which $f(x_{n'}).f(x_n) < 0$:

$$x_{n+1} = x_n - \frac{x_n - x_{n'}}{f(x_n) - f(x_{n'})} f(x_n)$$



FigureA8.3 *Regulafalsi* method

This method always converges when f(x) is continuous. On the other hand, the convergence of this method is linear and therefore less effective than the convergence of the classic chord method.

$$g'(x_n) = \qquad f'(x_n) = 1/m_n$$

8.2.2Newton–Raphson method

If, in the classic chord method, we choose m_n so that 0, that is,, we will obtain a second-order iteration. The method thus defined,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

is known as the Newton-Raphson method (Figure A8.4).

It is clearly a second-order method, as

$$\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|^2} = \frac{1}{2}\left|\frac{f''(a)}{f'(a)}\right| \neq 0$$

The Newton–Raphson method is therefore rapid insofar as the initial iterated value is not too far from the solution sought, as global convergence is not assured at all.

Aconvergencecriterionisthereforegivenforthefollowingtheorem.Assumethat f'(x) = 0 and that f''(x) does not change its sign within the interval [a; b] and f(a).f(b) <0. If, furthermore, $\left|\frac{f(a)}{f'(a)}\right| < b - a$ and $\left|\frac{f(b)}{f'(b)}\right| < b - a$

the Newton–Raphson method will converge at every initial arbitrary point x_0 that belongs to [a;b].

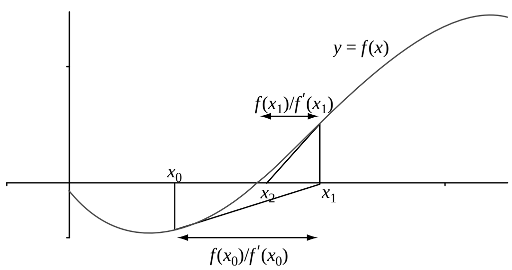


Figure 8.4 Newton–Raphson method

The classic chord method, unlike the Newton–Raphson method, requires two initial approximations but only involves one new function evaluation at each subsequent stage. The choice between the classic chord method and the Newton–Raphson method will therefore depend on the effort of calculation required for evaluation $f'(x)$.

Let us assume that the effort of calculation required for evaluation of $f'(x)$ is θ times the prior effort of calculation for f(x).

Given what has been said above, we can establish that the effort of calculation will be the same for the two methods if:

$$\frac{1+\theta}{\log 2} = \frac{1}{\log p} \text{ in which } p = \frac{1+\sqrt{5}}{2}$$

is the order of convergence in the classic chord method. In consequence:

- If θ > (log2/logp) − 1 ~0.44 → the classic chord method will be used.
- If θ ≤ (log 2/logp) − 1 ~0.44 → the Newton–Raphson method will be used.

8.2.3 Bisection method

The *bisection method* is a linear convergence method and is therefore slow. Use of the method is, however, justified by the fact that it converges overall, unlike the usual methods (especially the Newton–Raphson and classic chord methods). This method will therefore be used to bring the initial iterated value of the Newton–Raphson or classic chord method to a point sufficiently close to the solution to ensure that the methods in question converge.

Let us assume therefore that f(x) is continuous in the interval $[a_0;b_0]$ and such that[4] $f(a_0).f(b_0) < 0$. The principle of the method consists of putting together a converging sequence of bracketed intervals, $[a_1;b_1] \supset [a_2;b_2] \supset [a_3;b_3] \supset \ldots$, all of which contain a solution of the equation f(x) = 0.

If it is assumed that[5] $f(a_0) < 0$ and $f(b_0) > 0$, the intervals $I_k = [a_k;b_k]$ will be put together by recurrence on the basis of I_{k-1}.

$$[a_k; b_k] = \begin{cases} [m_k; b_{k-1}] & \text{if } f(m_k) < 0 \\ [a_{k-1}; m_k] & \text{if } f(m_k) > 0 \end{cases}$$

Here, $m_k = (a_{k-1} + b_{k-1})/2$. One is thus assured that $f(a_k) < 0$ and $f(b_k) > 0$, which guarantees convergence.

The bisection method is not a Picard iteration, but the order of convergence can be determined, as $\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|} = \frac{1}{2}$. The bisection method is therefore a first-order method.

8.3 NONLINEAR EQUATION SYSTEMS

We have a system of n nonlinear equations of n unknowns: $f_i(x_1,x_2,\ldots,x_n) = 0$ i = 1,2,...,n. Here, in vectorial notation, f(x) = 0. The solution to the system is an n-dimensional vector a.

[4] This implies that f(x) has a root within this interval.

[5] This is not restrictive in any way, as it corresponds to f(x) = 0 or −f(x) = 0, $x \in [a_0;b_0]$, depending on the case.

8.3.1General theory of *n*-dimensional iteration

n-dimensional iteration general theory is similar to the one-dimensional theory. The above equation can thus be expressed in the form:

$$x = g(x) \equiv x - \mathbf{A}(x)f(x)$$

where **A** is a square matrix of n^{th} order.

Picard's iteration is always defined as

$$x_{k+}1 = g(x_k) \quad k = 0,1,2 \text{ etc.}$$

and the convergence theorem for Picard's iteration remains valid in n dimensions.

In addition, if the Jacobian matrix **J**(x), defined by $[\mathbf{J}(x)]_{ij} = (\quad (x))g_{j\ x}i$ is such that for every x $\in$ I, $\|\mathbf{J}(x)\| \leq m$ for a norm compatible with m <1, Lipschitz's condition is satisfied.

The order of convergence is defined by

$$\lim \frac{\| \quad _{+1}\|}{\|e_k\|^p} = C \qquad e_k \ ^k\!\!\rightarrow\infty$$

where C is the constant for the asymptotic error.

8.3.2Principal methods

If one chooses a constant matrix **A** as the value for **A**(x), the iterative process is the generalisation in n dimensions of the chord method.

If the inverse of the Jacobian matrix of f is chosen as the value of **A**(x), we will obtain the generalisation in n dimensions of the Newton–Raphson method.

Another approach to solving the equation f(x) = 0 involves using the i^{th} equation to determine the $(i + 1)^{th}$ component. Therefore, for i = 1,2,...,n, the following equations will be solved in succession:

$$f_i(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \ldots, x_n^{(k)}) = 0$$

with respect to x_i. This is known as the *nonlinear Gauss–Seidel method*.

Algorithm

Newton Raphson method Steps (Rule)

Step-1:	Find points a and b such that $a<b$ and $f(a)\cdot f(b)<0$.
Step-2:	Take the interval $[a,b]$ and find next value $x_0=\frac{a+b}{2}$
Step-3:	Find $f(x_0)$ and $f'(x_0)$ $x_1=x_0-\frac{f(x_0)}{f'(x_0)}$
Step-4:	If $f(x_1)=0$ then x_1 is an exact root, else $x_0=x_1$
Step-5:	Repeat steps 2 to 4 until $f(x_i)=0$ or $\lvert f(x_i)\rvert\leq$ Accuracy

---

Example-1
Find a root of an equation $f(x)=x^3-x-1$ using Newton Raphson method

Solution:
Here $x^3-x-1=0$

Let $f(x)=x^3-x-1$

$\therefore f'(x)=3x^2-1$

Here

x	0	1	2
$f(x)$	-1	-1	5

Here $f(1)=-1<0$ and $f(2)=5>0$

$\therefore$ Root lies between 1 and 2

$x_0=\frac{1+2}{2}=1.5$

1^{st} iteration :

$f(x_0)=f(1.5)=0.875$

$f'(x_0)=f'(1.5)=5.75$

$x_1=x_0-\frac{f(x_0)}{f'(x_0)}$

$x_1=1.5-0.8755.75$

$x_1=1.34783$

2*nd* iteration :

$f(x_1)=f(1.34783)=0.10068$

$f'(x_1)=f'(1.34783)=4.44991$

$x_2=x_1-f(x_1)f'(x_1)$

$x_2=1.34783-0.100684.44991$

$x_2=1.3252$

3*rd* iteration :

$f(x_2)=f(1.3252)=0.00206$

$f'(x_2)=f'(1.3252)=4.26847$

$x_3=x_2-f(x_2)f'(x_2)$

$x_3=1.3252-0.002064.26847$

$x_3=1.32472$

4*th* iteration :

$f(x_3)=f(1.32472)=0$

$f'(x_3)=f'(1.32472)=4.26463$

$x_4=x_3-f(x_3)f'(x_3)$

$x_4=1.32472-04.26463$

$x_4=1.32472$

Approximate root of the equation $x^3-x-1=0$ using Newton Raphson mehtod is 1.32472

n	***x*0**	***f*(*x*0)**	***f′*(*x*0)**	***x*1**	**Update**
1	1.5	0.875	5.75	1.34783	$x_0=x_1$
2	1.34783	0.10068	4.44991	1.3252	$x_0=x_1$
3	1.3252	0.00206	4.26847	1.32472	$x_0=x_1$
4	1.32472	0	4.26463	1.32472	$x_0=x_1$

---

2. Find a root of an equation $f(x)=2x^3-2x-5$ using False Position method (regula falsi method)

Solution:

Here $2x^3-2x-5=0$

Let $f(x)=2x^3-2x-5$

Here

x	0	1	2
***f*(*x*)**	-5	-5	7

1^{st} iteration :

Here $f(1)=-5<0$ and $f(2)=7>0$

$\therefore$ Now, Root lies between $x_0=1$ and $x_1=2$

$x_2=x_0-f(x_0)\cdot\frac{x_1-x_0}{f(x_1)-f(x_0)}$

$x_2=1-(-5)\cdot\frac{2-1}{7-(-5)}$

$x_2=1.41667$

$f(x_2)=f(1.41667)=2\cdot1.41667^3-2\cdot1.41667-5=-2.14699<0$

2^{nd} iteration :

Here $f(1.41667)=-2.14699<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.41667$ and $x1=2$

$x3=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x3=1.41667-(-2.14699)\cdot 2-1.416677-(-2.14699)$

$x3=1.55359$

$f(x3)=f(1.55359)=2\cdot 1.553593-2\cdot 1.55359-5=-0.60759<0$

$3rd$ iteration :

Here $f(1.55359)=-0.60759<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.55359$ and $x1=2$

$x4=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x4=1.55359-(-0.60759)\cdot 2-1.553597-(-0.60759)$

$x4=1.58924$

$f(x4)=f(1.58924)=2\cdot 1.589243-2\cdot 1.58924-5=-0.15063<0$

$4th$ iteration :

Here $f(1.58924)=-0.15063<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.58924$ and $x1=2$

$x5=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x5=1.58924-(-0.15063)\cdot 2-1.589247-(-0.15063)$

$x5=1.59789$

$f(x5)=f(1.59789)=2\cdot 1.597893-2\cdot 1.59789-5=-0.0361<0$

$5th$ iteration :

Here $f(1.59789)=-0.0361<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.59789$ and $x1=2$

$x6=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x6=1.59789-(-0.0361)\cdot 2-1.597897-(-0.0361)$

$x6=1.59996$

$f(x6)=f(1.59996)=2\cdot 1.599963-2\cdot 1.59996-5=-0.00858<0$

6*th* iteration :

Here $f(1.59996)=-0.00858<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.59996$ and $x1=2$

$x7=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x7=1.59996-(-0.00858)\cdot 2-1.599967-(-0.00858)$

$x7=1.60045$

$f(x7)=f(1.60045)=2\cdot 1.600453-2\cdot 1.60045-5=-0.00203<0$

7*th* iteration :

Here $f(1.60045)=-0.00203<0$ and $f(2)=7>0$

∴ Now, Root lies between $x0=1.60045$ and $x1=2$

$x8=x0-f(x0)\cdot x1-x0f(x1)-f(x0)$

$x8=1.60045-(-0.00203)\cdot 2-1.600457-(-0.00203)$

$x8=1.60056$

$f(x8)=f(1.60056)=2\cdot 1.600563-2\cdot 1.60056-5=-0.00048<0$

Approximate root of the equation 2*x*3-2*x*-5=0 using False Position mehtod is 1.60056

n	*x*0	*f*(*x*0)	*x*1	*f*(*x*1)	*x*2	*f*(*x*2)	Update
1	1	-5	2	7	1.41667	-2.14699	*x*0=*x*2
2	1.41667	-2.14699	2	7	1.55359	-0.60759	*x*0=*x*2
3	1.55359	-0.60759	2	7	1.58924	-0.15063	*x*0=*x*2
4	1.58924	-0.15063	2	7	1.59789	-0.0361	*x*0=*x*2
5	1.59789	-0.0361	2	7	1.59996	-0.00858	*x*0=*x*2
6	1.59996	-0.00858	2	7	1.60045	-0.00203	*x*0=*x*2
7	1.60045	-0.00203	2	7	1.60056	-0.00048	*x*0=*x*2

12.2 Numerical Quadrature

- Define numerical quadrature.
- Use Trapezoidal and Simpson's Rules,
- Compute the approximate value of definite integrals without error terms.
- Apply MAPLE command trapezoid for trapezoidal rule and Simpsons for Simpsons rule and demonstrate trough examples.

Numerical Quadrature

- *Quadrature* refers to any method for numerically approximating the value of a definite integral $\int_a^b f(x)dx$. The goal is to attain a given level of precision with the fewest possible function evaluations.

  The crucial factors that control the difficulty of a numerical integration problem are the dimension of the argument *x* and the smoothness of the integrand *f*.

- Any quadrature method relies on evaluating the integrand *f* on a finite set of points (called the *abscissas* or *quadrature points*), then processing these evaluations somehow to produce an approximation to the value of the integral. Usually this involves taking a weighted average.

  The goal is to determine which points to evaluate and what weights to use so as to maximize performance over a broad class of integrands.

- A natural strategy is to approximate f using a spline g with knots at a certain set of quadrature points. The integral $\int_a^b g(x)dx$ is easy to evaluate since it is a piecewise polynomial, and since g approximates f it makes sense to use $\int_a^b g(x)dx$ as an approximation to $\int_a^b f(x)dx$. When the quadrature points are evenly spaced the resulting quadrature rules are called *Newton-Cotes formulas*.

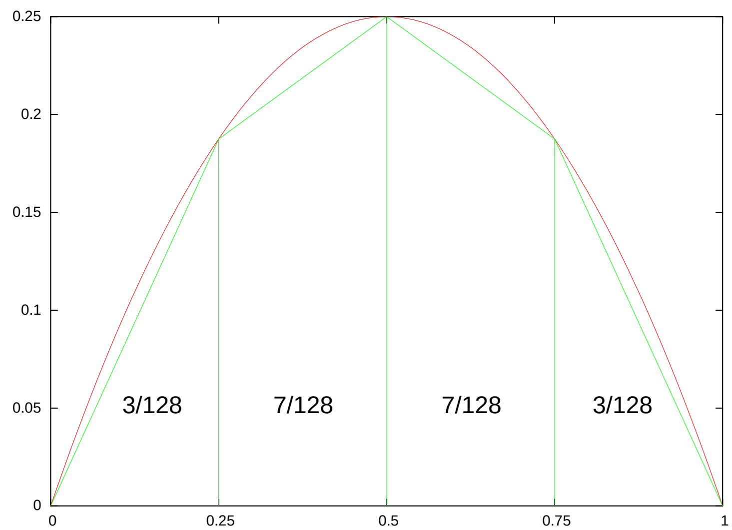
- Suppose we construct a grid on $[a,b]$, using mesh $h = (b - a)/m$, where m is an integer (the mesh is the distance between adjacent grid points). The grid points are $x_k = a+hk$. Let $f_k = f(x_k)$. Two straightforward rules are derived by approximating the integrand with linear or quadratic splines:

- *Trapezoidal rule:*

  The integral from x_k to x_{k+1} is $h(f_{k+1} + f_k)/2$. Therefore:

$$\int_a^b f(x)dx \approx h(\frac{1}{2}f_0 + f_1 + \ldots, +f_{m-1} + \frac{1}{2}f_m).$$

  The following shows the trapezoidal rule with $m = 4$ applied to $f(x) = x(1 - x)$ on [0,1]. The approximate integral is 5/32 while the exact value is 1/6.



- *Simpson's rule:*

  Using Lagrange's formula, the quadratic interpolant through x_k,x_{k+1},x_{k+2} can be shown to have integral $h(f_k + 4f_{k+1} + f_{k+2})/3$. Therefore:

$$\int_a^b f(x)dx \approx h(\frac{1}{3}f_0 + \frac{4}{3}f_1 + \frac{2}{3}f_2 + \ldots + \frac{4}{3}f_{m-1} + \frac{1}{3}f_m)$$

  Note that when applying Simpson's rule there must be an even number of intervals.

- Higher order (>2) rules can be formulated, but are rarely used.
- In order to characterize the accuracy of these rules, we need to determine the accuracy of polynomial interpolation:

Theorem: Let f be $n + 1$ times continuously differentiable. Let $\tilde{f}$ be the degree n polynomial interpolation of f at $x_0,\ldots,x_n$. Then the error of interpolation at x can be written:

$$f(x) - \tilde{f}(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!} \prod_j (x - x_j)$$

for some $\min(x_j) \le \eta \le \max(x_j)$.

Proof: Let $q(x) = {}^{Q}(x - x_j)$, and define g by:

$$g(y) \equiv f(y) - \tilde{f}(y) - q(y)\frac{f(x) - \tilde{f}(x)}{q(x)}.$$

Note that g is $n + 1$ times continuously differentiable, and g vanishes at x and at $x_0,\ldots,x_n$. Therefore by Rolle's theorem g^0 has at least $n + 1$ zeros. Repeated application of Rolle's theorem gives that $g^{(n+1)}$ has at least one zero, which we denote η. Evaluating $g^{(n+1)}$ at η gives the result.

Now we can derive the error estimate for the trapezoidal rule.

Theorem: Let f be twice continuously differentiable. Then for some $a \le \eta \le b$, the error for the trapezoidal rule is:

$$\int_a^b f(x)dx - \frac{b-a}{2}\left(f(a) + f(b)\right) = -\frac{(b-a)^3}{12} f''(\eta).$$

Proof: First observe that if $\tilde{f}$ denotes the linear interpolation of f on $[a,b]$, then the error can be represented as

$$\int_a^b f(x)dx - \frac{b-a}{2}\left(f(a) + f(b)\right) = \int_a^b \left(f(x) - \tilde{f}(x)\right) dx = \int_a^b (x-a)(x-b)\frac{f(x) - \tilde{f}(x)}{(x-a)(x-b)} dx.$$

The term $(x-a)(x-b)$ is non-positive, and the second factor is continuous. Therefore by the mean value theorem there exists η such that the above becomes equal to

$$\frac{f(\eta) - \tilde{f}(\eta)}{(\eta - a)(\eta - b)} \int_a^b (x-a)(x-b)dx.$$

The second term is equal to $-(b-a)^3/6$, and the first term is equal to $f^{00}(\theta)/2$ for some $\theta \in [a,b]$, which gives the result.

- Applying this result, if we use mesh h then the error of the trapezoidal rule between successive abscissas decreases as h^3. Since there are L/h intervals, the total error decreases as h^2, or as $1/m^2$ if there are m abscissas. Thus in order to reduce the error by half, the number of quadrature points must be increased by approximately a factor of $\sqrt{2} \approx 1.4$.

- A direct extension of the theorem gives that the error of approximation for a degree n interpolant $\tilde{f}$ is bounded in magnitude by

$$\max_{a \le u \le b} \left| \frac{f^{(n+1)}(u)}{(n+1)!} \right| (b-a)^{n+2}.$$

  Applying this to Simpson's rule ($n = 2$) gives that the error for a single quadratic interpolating polynomial on (a,b) decreases at order $(b-a)^4$. Thus the error for the interpolating spline with m abscissas decreases at order h^3, or as $1/m^3$ for m abscissas. Thus in order to reduce the error by half using Simpson's rule, the number of quadrature points must be increased by approximately a factor of $2^{1/3} \approx 1.26$.

- Surprisingly, Simpson's rule actually has error of magnitude $h^5 f^{(4)}(\eta)/90$ per interval, or $h^4 f^{(4)}(\eta)/90$ overall – an order better than calculated above.

  It is a general property that the order of the Newton-Cotes formula increases by two when moving from odd-order to even-order interpolations, and doesn't increase at all the other way around.

- Integration rules can be characterized in terms of the highest degree polynomial for which the error is zero. The trapezoidal rule is exact for degree 1 polynomials and Simpson's rule is exact for degree 2 polynomials – these statements are true because the interpolating polynomial is exactly equal to the integrand, $f \equiv \tilde{f}$.

  In fact, Simpson's rule is exact for cubics. To see this it is adequate to check that it holds for $f(x) = x^3$ (since Simpson's rule is additive and we already know that it is exact for quadratics). By direct evaluation $\int_0^1 x^3 dx = 1/4$, while Simpson's rule gives $(1 \cdot 0 + 4/8 + 1 \cdot 1)/6 = 1/4$ (in applying Simpson's rule the abscissas are $0, 1/2, 1$ and the mesh is $h = 1/2$).

  The reason for this is that the difference $f - \tilde{f}$, while not identically zero, has zero integral. The quadratic interpolant through $(0,0)$, $(1/2, 1/8)$, $(1,1)$ is

$$\tilde{f}(x) = 3x^2/2 - x/2$$

and the pointwise error of interpolation is

$$\tilde{f}(x) - f(x) = 3x^2/2 - x/2 - x^3 = -(x - 1/2)^3 + (x - 1/2)/4 + 1/2$$

which is an odd function with respect to $x = 1/2$, hence intergates to zero over any interval centered at $1/2$.

- For the trapezoidal rule, it is possible to descend through a sequence of meshes, each half the size of the previous mesh. This gives a sequence of approximations $A_n \to \int_a^b f(x)dx$ with mesh $(b - a)/2^n$.

  Note that this can be done efficiently without recomputing abscissas from the previous grid or even needing to save the individual function values, using

$$A_{n+1} = A_n/2 + (b - a)(\text{sum of new ordinates})/2^{n+1}.$$

  No such simple updating is available for Simpson's rule.

Unevenly spaced abscissas

- We can gain a lot of efficiency by using unevenly spaced abscissas. In particular, we can get exact results for polynomials of degree up to $2n-1$ with only n evaluations of f.

Lemma 1: Let $w(x) > 0$ be integrable on $[a,b]$. There exists an orthogonal basis $\{q_0(x),...,q_n(x)\}$ of monic polynomials with respect to $w(x)$, where $q_n(x)$ has degree n:

$$\int_a^b w(x)q_n(x)q_{n'}(x)dx \propto \mathcal{I}(n = n')$$

Proof: Use the Gram-Schmidt procedure on the canonical basis $\{1, x, x^2, ...\}$.

Lemma 2: Each $q_n(x)$ with $n > 0$ has n simple, real zeros.

Proof: First observe that $q_0(x) \propto 1$, so for $n > 0$, $\int_a^b w(x)q_n(x)dx = 0$ by orthogonality. Therefore $q_n(x)$ has at least one real root. Let $x_1,...,x_m$ denote the real roots of $q_n(x)$, so that $q_n(x) = g(x)\prod(x - x_j)^{d_j}$ where $g(x)$ has no real roots and the x_j are distinct. Let $r(x) = \prod_{\ell : d_\ell \text{ odd}}(x - x_\ell)$. If $r(x) = q_n(x)$ then we are done. Otherwise, $q_n(x) = g(x)r(x)h(x)$, where $h(x)$ has no real roots of odd multiplicity, and hence no sign changes. Since $r(x)$ has lower degree than $q_n(x)$, $\int_a^b w(x)r(x)q_n(x) = 0$ by orthogonality. But the integrand can be written $w(x)g(x)r(x)^2h(x)$ which has no sign changes, which requires $g(x)r(x)h(x) = q_n(x)$ to be identically zero, which is a contradiction.

Lemma 3: Let $x_0,\dots,x_n$ be the roots of $q_{n+1}(x)$. Then there exists a unique set of numbers $a_0,\dots,a_n$ such that:

$$\int_a^b w(x)q(x)dx = \sum_{j=0}^{n} a_j q(x_j)$$

for all polynomials q of degree at most n.

Proof: By linearity, and since $q_0,\dots,q_n$ form a basis for the set of all polynomials of degree n, it is sufficient that the rule hold for this basis, that is,

$$\int_a^b w(x)q_k(x)dx = \sum_{j=0}^{n} a_j q_k(x_j) \qquad k = 0,\dots,n.$$

Since $\int_a^b w(x)q_k(x)dx = 0$ for $k > 0$ by orthogonality, we can determine the a_j by solving the linear system

$$\begin{pmatrix} q_0(x_0) & \dots & q_0(x_n) \\ q_1(x_0) & \dots & q_1(x_n) \\ \vdots & \dots & \vdots \\ q_n(x_0) & \dots & q_n(x_n) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \int_a^b w(x)q_0(x)dx \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It is a simple exercise to show that the coefficient matrix is non-singular, hence the solution exists.

Theorem: The values a_j and x_j defined in lemma 3 satisfy

$$\int_a^b w(x)p(x)dx = \sum_{j=0}^{n} a_j p(x_j),$$

for all polynomials $p(x)$ of degree at most $2n + 1$.

Proof: Let $\tilde{p}(x)$ denote the degree n polynomial that agrees with $p(x)$ at $x_0,\dots,x_n$ (the roots of q_{n+1}). Write $p(x) = \tilde{p}(x)+\tilde{q}(x)$, and observe that since $\tilde{q}(x) = 0$ at $x_0,\dots,x_n$, we can write $\tilde{q}(x) = q_{n+1}(x)q(x)$ for $q(x)$ of degree n. Therefore

$$\int_a^b w(x)p(x)dx = \int_a^b w(x)\tilde{p}(x)dx + \int_a^b w(x)q_{n+1}(x)q(x)dx$$

a a a

$$= \int_a^b w(x)\tilde{p}(x)dx.$$

By lemma 3,

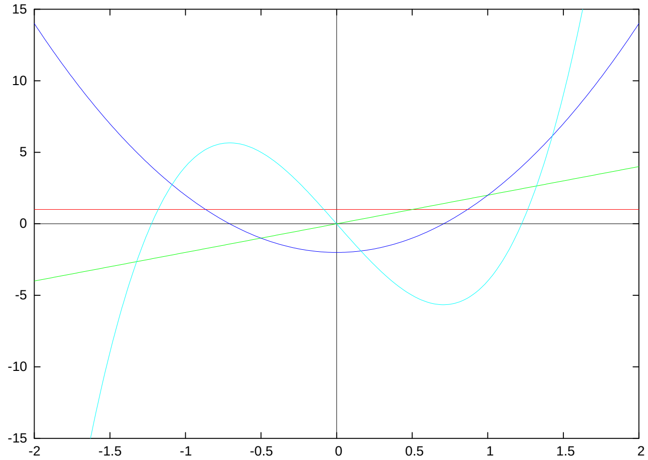
$$\int_a^b w(x)\tilde{p}(x)dx = \sum_{j=0}^{n} a_j \tilde{p}(x_j) = \sum_{j=0}^{n} a_j p(x_j),$$

which completes the proof.

Quadrature weight functions

- $w(x) = 1$ on $[a,b]$ (*Gauss-Legendre*)
- $w(x) = \sqrt{1 - x^2}$ on $[-1,1]$.
- $w(x) = 1/\sqrt{1 - x^2}$ on $[-1,1]$ (*Gauss-Chebyshev*)
- $w(x) = \exp(-x^2)$ on $(-\infty,\infty)$ (*Gauss-Hermite*)

For example, the first four Guass-Hermite basis functions are shown below.



The procedure is used in two ways:

1. Integration of an arbitrary integrand:
$$\int_a^b g(x)dx = \int_a^b w(x)(g(x)/w(x))dx \approx \sum_j a_j g(x_j)/w(x_j)$$

2. Integration with respect to an explicit weight function, often a probability densityfunction, e.g.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(x)\exp(-x^2/2)dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(\sqrt{2}x)\exp(-x^2)dx \approx \frac{1}{\sqrt{\pi}}\sum_j a_j g(\sqrt{2}x_j).$$

- In the former case, the best results will be achieved if $g(x)/w(x)$ is approximately a low order polynomial. In the latter case, the best results will be achieved if $g(x)$ is approximately a low order polynomial.

- The error estimate for Gaussian quadrature is given by the following, for some $a \leq \eta \leq b$:
$$\int_a^b w(x)f(x)dx - \sum_{j=0}^{n} a_k f(x_k) = \frac{f^{(2n+2)}(\eta)}{(2n+2)!}\int_a^b w(x)q_{n+1}(x)^2dx$$
In general, the quadrature points x_j and weights w_j can be obtained by (i) searching for the roots, for instance using Newton's method, and (ii) solving a linear system for the weights. For the standard weights listed above, there are usually recurrence formulas that eliminate the second part. For instance, in the Gauss-Hermite case:

$$w_j = \frac{2^{n-1}n!\sqrt{\pi}}{n^2 H_{n-1}(x_j)^2},$$

where H_j is the j^{th} Gauss-Hermite polynomial (i.e. the j^{th} element of the basis described in Lemma 1 with $w(x) = \exp(-x^2)$). The following recurrences are useful when implementing the Newton search for the x_j:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x),$$

$$H_n'(x) = 2nH_{n-1}(x).$$

The squared norm of H_n (using the Gauss-Hermite weight function to determine the inner product) is $2^n n! \sqrt{\pi}$. Therefore the polynomials get large very quickly, so for numerical reasons it is essential to work with the normalized Gauss-Hermite polynomials. Letting $\tilde{H}_n(x)$ denote the normalized n^{th} Gauss-Hermite polynomial, we get the following recurrences:

$$\tilde{H}_{n+1}(x) = \sqrt{2/(n+1)} \cdot x\tilde{H}_n(x) - \sqrt{n/(n+1)}\tilde{H}_{n-1}(x).$$

$$\tilde{H}_n'(x) = \sqrt{2n}\tilde{H}_{n-1}(x).$$

- The one-dimensional quadrature rules discussed above can be directly generalized to higher dimensions. However as the dimension increases, the number of quadrature points grows geometrically. Consider the two-dimensional integral $\int f(x,y)dxdy$. We can integrate out one variable as follows:

$$\int f(x,y)dy \approx \sum_{j=0}^{n} w_j f(x,x_j).$$

Next we approximate the integral as

$$\int f(x,y)dxdy \approx \sum_{i=0}^{n}\sum_{j=0}^{n} w_i w_j f(x_i,x_j).$$

There is a lot of work on *adaptive quadrature* that attempts to place the quadrature points efficiently in high dimensions. One strategy is to use more quadrature points in the coordinate directions where f is more variable. Another strategy is to reparametrize f to try to make it uniformly smooth in all directions.

Trapezoidal Rule

Calling Sequence

ApproximateInt(f(x), x = a..b, method = trapezoid, opts)
ApproximateInt(f(x), a..b, method = trapezoid, opts)
ApproximateInt(Int(f(x), x = a..b), method = trapezoid, opts)

Parameters

f(x) -algebraic expression in variable **'x'**

x -name; specify the independent variable

a, b -algebraic expressions; specify the interval

opts-equation(s) of the form **option=value** where **option** is one of **boxoptions**, **functionoptions**, **iterations**, **method**, **outline**, **output**, **partition**, **pointoptions**, **refinemen** or Student plot options; specify output options

Description

- The **ApproximateInt(f(x), x = a..b, method = trapezoid)** command approximates the integral of **f(x)** from **a** t (function expression and range) can be replaced by a definite integral.
- If the independent variable can be uniquely determined from the expression, the parameter **x** need not be inclu
- Given a partition $P=(a=x0,x1,...,xN=b)$ of the interval (a,b), the trapezoidal rule approximates the integral on ea interpolates the endpoints $(xi-1,f(xi-1))$ and $(xi,f(xi))$. This value is

$$(xi-xi-1)(f(xi-1)+f(xi))2$$

- In the case that the widths of the subintervals are equal, the approximation can be written as

$$(b-a)(f(x0)+2f(x1)+2f(x2)+...+2f(xN-1)+f(xN))2N$$

- By default, the interval is divided into 10 equal-sized subintervals.
- For the options **opts**, see the ApproximateInt help page.
- This rule can be applied interactively, through the ApproximateInt Tutor.

Examples

> polynomial≔CurveFittingPolynomialInterpolation([x0,x1],[f(0),f(1)],z):

> integrated≔∫x1x0polynomialdz:

> factor(integrated)

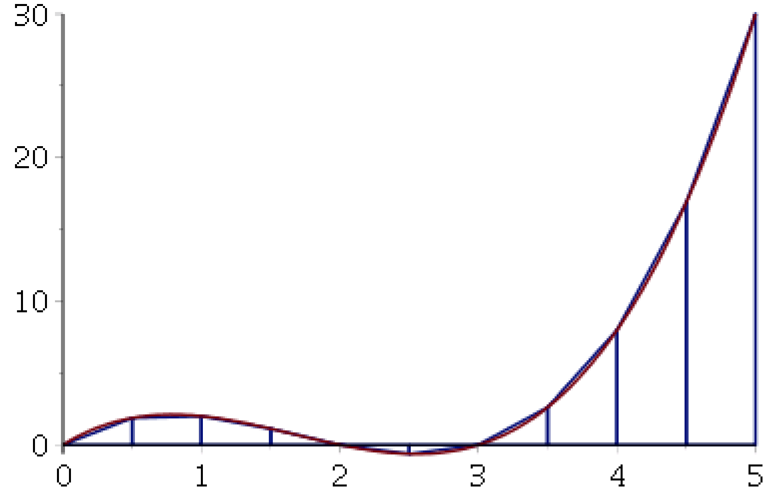
$-(-x1+x0)(f(1)+f(0))2$ **(1)**

> with(Student[Calculus1]):

> ApproximateInt(sin(x),x=0..5,method=trapezoid)

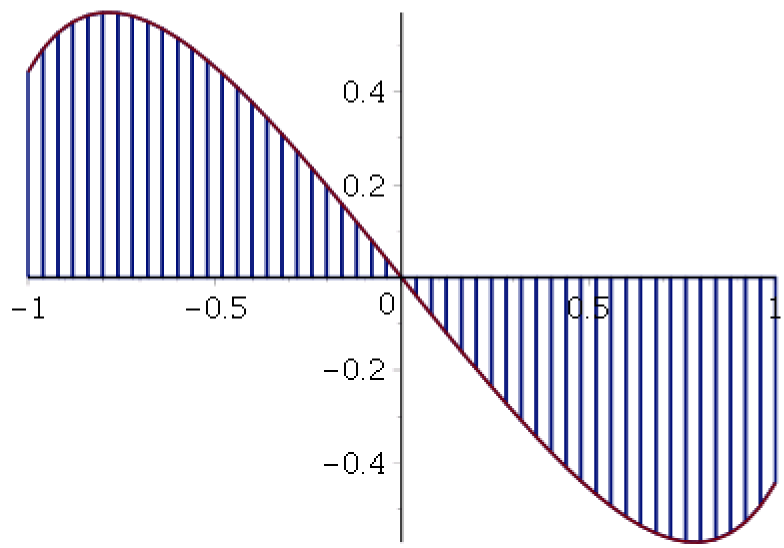
sin(12)2+sin(1)2+sin(32)2+sin(2)2+sin(52)2+sin(3)2+sin(72)2+sin(4)2+sin(92)2+sin(5)4 **(2)**

> ApproximateInt(x(x−2)(x−3),x=0..5,method=trapezoid,output=plot)



An approximation of $\int_0^5 f(x)\,dx$ using trapezoid rule, where $f(x) = x\,(x-2)\,(x-3)$ and the partition is uniform. The approximate value of the integral is 23.43750000. Number of subintervals used: 10.

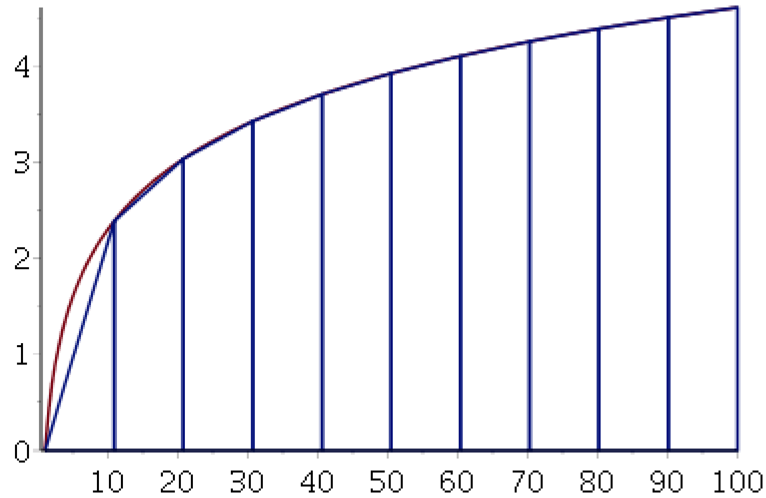
> ApproximateInt(tan(x)−2x,x=−1..1,method=trapezoid,output=plot,partition=50)



An approximation of $\int_{-1}^{1} f(x)\,dx$ using trapezoid rule, where $f(x) = \tan(x) - 2\,x$ and the partition is uniform. The approximate value of the integral is 0.. Number of subintervals used: 50.

To play the following animation in this help page, right-click (**Control**-click, on Macintosh) the plot to display t

> ApproximateInt(ln(x),1..100,method=trapezoid,output=animation)



An animated approximation of $\int_{1}^{100} f(x)\,dx$ using trapezoid rule, where $f(x) = \ln(x)$ and the partition is uniform. The approximate value of the integral is 356.5535441. Number of subintervals used: 10.