# The Best of Calculus

Tom Elliott

January 9, 2020

# Contents

# Part I

# Archimedes

# Chapter 1

# Introduction

This was supposed to be a short book, an exploration of problems like the volume of the cone and sphere, or even just the area of a circle, with some simple physics thrown in. These questions contain within them the heart of calculus: infinities both large and small. I imagine myself looking over Archimedes' shoulder as he explains it to me.

I wrote many of the early chapters originally as short explanations for my son Sean as he studied calculus in high school. It bothers me that so often the good stuff gets left out — the ideas which make you go ... wow. Now, years later, I still find a lot of pleasure in trying to understand what Kepler and Newton did. It took a genius to figure it out the first time, but it is within anyone's grasp to appreciate what they found.

Then I thought, why not include other favorite problems like the area of the ellipse, the "headlight" problem for the parabola, or the reflective property of the ellipse, and the length and area under the cycloid curve (the "light on a bicycle wheel"). These are problems where calculus easily produces answers that can be checked by more elaborate geometric arguments. In fact, this book might as well be titled .. *Best* *of Calculus and Geometry.*

So here we are, with a somewhat longer book.

In the introduction to his book *Calculus*, Morris Kline says

> Anyone who adds to the plethora of introductory calculus texts owes an explanation, if not an apology, to the mathematical community.

I think of this book as akin to ultralight backpacking. We shed weight so as to ascend peaks rapidly, skimming the best of calculus — focusing on geometry and physics, and slinging differentials with abandon. Epsilon is a bit player in the production. Starting with an intuitive notion of adding up many small pieces, we put integrals to work early solving problems.

Going fast allows time to get a view of sophisticated topics, among others, line integrals for work and flux, Newton's proof that a spherical mass acts as a point mass, and integration of a parametrized surface like the torus. Not to mention Kepler's Laws, and a derivation of the Gaussian distribution from first principles.

We do not disdain proof. Proof is central to the enterprise. We prove the Pythagorean Theorem, and the quotient rule for derivatives, as well as Green's Theorem. There is a fun chapter on induction. We prove that $\pi$ is a constant. In fact, the word "proof" appears nearly 200 times in the text and one of its most interesting features is the natural use of proofs that I have tried to make as simple and easy to follow as possible.

My favorite authors on calculus are Morris Kline, Richard Hamming, and Gil Strang. Sylvanus Thompson's simple book is my favorite first text, and it's even a Project Gutenberg project:

https://www.gutenberg.org/files/33283/33283-pdf.pdf

Having said what I like, briefly, here are some things I don't like.

The rigorous approach to calculus pioneered by Cauchy in the 1820's and exported to American schools by Richard Courant in the 1940's is a bad idea. We must motivate rigorous proof by demonstrating utility first. As Ian Stewart says, "proofs come *after* understanding." Courant's method is the way to teach the subject the second or even third time through.

Thompson:

> You don't forbid the use of a watch to every person who does not know how to make one. You don't object to the musician playing on a violin that he has not himself constructed. You don't teach the rules of syntax to children until they have already become fluent in the use of speech. It would be equally absurd to require general rigid demonstrations to be expounded to beginners in the calculus.

A second thing I dislike is calculus problems that are gratuitously arithmetic. Calculus consists of bright ideas, not complicated ones; if the computation is difficult, it's usually *not* a good problem. Also, a good problem often is one with a physical or practical foundation. Having said that, if a course could integrate elementary programming with calculus, I would be very happy.

Finally, a saying attributed to Manaechmus (speaking to Alexander the Great), "there is no royal road to geometry". Which means, practically, that learning mathematics requires that you follow the argument with pencil and paper and work out each step yourself, to your own satisfaction. That is the only way of really learning, and at heart, one of the reasons I wrote this book.

I express my sincere thanks to the authors of my favorite books, which are listed in the references and mentioned at various places in the text. Almost everything in here was appropriated from them, and styled to my taste. I offer my profound thanks also to Eugene Colosimo, S.J. He was, for me, the best of a bunch of very special teachers.

If I stole your figure off the internet, I'm sorry. I intended to redraw it but have not yet found the time.

We start with my favorite mathematician, Archimedes.

You can find the current version of the book on github here:

https://github.com/telliott99/calculus_book

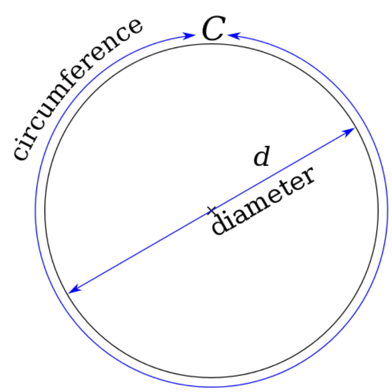
# Chapter 2

# Area of a circle

In this first unit we will develop the most famous of Archimedes geometrical contributions, a theorem on the volume of the sphere.

Before we get there, however, we need to spend some time with circles (also a topic to which he contributed) and look at the problem of the volume of cones and pyramids. These are topics in geometry that come even before the volume of the sphere.

We start with the circle. A fundamental result about circles is that the ratio of the circumference of a circle to its diameter is independent of the size of the circle.



The proportionality constant is

$$\pi = \frac{C}{d}$$

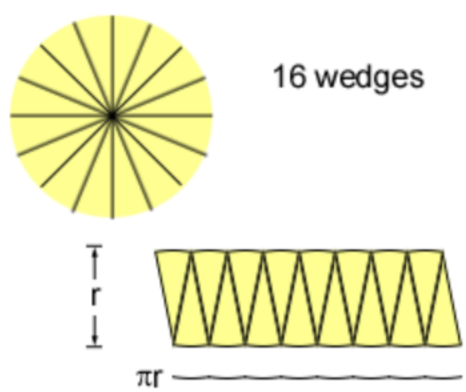
Since the radius is one-half the diameter, $2r = d$ and

$$2\pi r = C$$

This is usually stated as a self-evident fact, but it is actually a theorem to be proved. We will need some of the apparatus of calculus, so we defer the proof.

### area of a circle

Imagine dividing a circle into wedges, like you might do with a pizza. Here, the pie has been divided into 16 parts.




Since the pieces are triangular, it is easy to stack them next to each other with the bases and tips alternating, as shown. Of course the bases are not straight, but have the same curvature as the edge of the circle.

The length of the short side is the radius, $r$, and the length of the long side is approximately one-half the circumference so

$$A = r \cdot \frac{1}{2} \cdot 2\pi r = \pi r^2$$

The trick is to imagine that we subdivide the circle into many slices. If there are infinitely many slices, the edges will be straight and this calculation becomes exact.

According to wikipedia

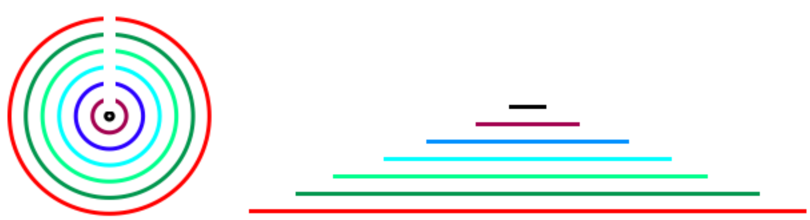
https://en.wikipedia.org/wiki/Area_of_a_circle

Eudoxus of Cnidus, born in the 5th century (408 BCE), proved that the area of a circle, like that of regular polygons, is proportional to both horizontal and vertical dimensions, and thus is proportional to the radius squared.

Somewhat later, it became clear that for a regular polygon, the area is equal to one-half the perimeter times the altitude from the center to each side (called the apothem). Allowing the polygon to achieve many, many sides, that formula gives $1/2 \cdot 2\pi r \cdot r = \pi r^2$.

The proof we gave above is very much like one attributed to Leonardo da Vinci, among others.

Another idea is to remove concentric strips from the edge and stack them.



We obtain a triangle of height $r$ and base $2\pi r$ so its area is

$$\frac{1}{2}\,2\pi r \cdot r = \pi r^2$$

A proof that this triangle has the same area as the circle was given by Archimedes and is found in his *Measurement of a Circle*, proposition 1. However, many sources, including

http://www.math.tamu.edu/~dallen/masters/Greek/eudoxus.pdf

attribute the proof to Eudoxus, who was perhaps the second most famous mathematician of antiquity, and a colleague of Plato in Athens.

We'll look at the proof here even though it's a bit sophisticated for so early in the book. The quote from Plutarch that follows is priceless.

◦ Let $A$ be the area of the circle

◦ Let $T$ be the area of the triangle formed with base $2\pi r$ and height $r$ (i.e. the area of $T$ is $\pi r^2$).

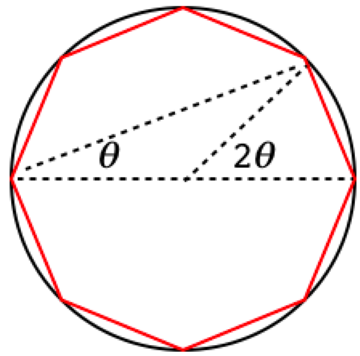
The method of proof is by finding a contradiction. We will assume something next, and then prove that a contradiction results, so the assumption must be incorrect

◦ Assume $A > T$.

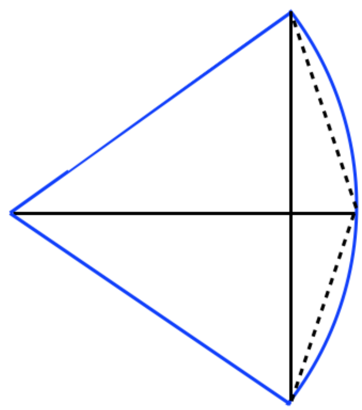
That is, the difference $A - T$ is non-zero and positive:

$$A - T > 0$$

Using the methods described **here**, we know that it is possible to construct an inscribed polygon whose area differs from $A$ by *as little as we please*




Here is a single sector:



In the figure, suppose that this is a sector of a blue circle, and the black vertical is one side of an inscribed polygon of $n$ sides,.

In the next step, we find a way to double the number of sides (dotted lines). This is a simple construction, just divide the secant between two adjacent vertices, and draw the radius through that point to the edge of the circle. By this means we can obtain a series, like $6, 12, 24, 48, 96 \ldots$ sides, as Archimedes did.

Clearly, the polygon with $2n$ sides is still contained within the circle, but its area more closely approximates that of the circle. This can be repeated forever.

Call the area of the inscribed polygon $P$.

So what we meant by as little as we please is that $P$ can be made closer to $A$ than $T$ is, simply on the assumption that $T < A$. Since this is an *inscribed* polygon, we have

$$A - P < A - T$$

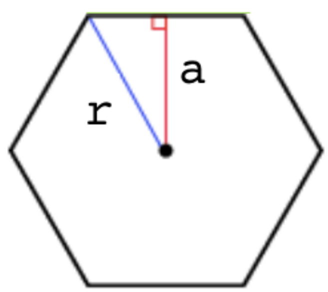
Add $-A$ to both sides:

$$-P < -T$$

Now, add $P + T$ to both sides:

$$T < P$$

However, for an inscribed polygon, the area is the number of sides $n$ times the length of the base of each side, which is the perimeter, times the apothem (the vertical to the sides of the polygon), labeled $a$, times 1/2.




In the figure, it must be that $a < r$.

But the perimeter is certainly less than the circumference of the circle ($2\pi r$) so:

$$P < \frac{1}{2} \cdot 2\pi r \cdot r$$

By this second argument, we have shown that the area of the regular polygon, $P < T$. However, we first showed that $T < P$. We have reached a contradiction.

Therefore, our assumption that $A > T$, is incorrect. $A$ is *not* greater than $T$:

$$A \ngtr T$$

A similar argument assuming $A < T$ also leads to a contradiction.

Since $A$ is neither greater than nor smaller than $T$ it must be equal to $T$.

$$A = T = \frac{1}{2} \cdot 2\pi r \cdot r = \pi r^2$$

The analysis is taken from Dunham's *Journey Through Genius.* Here is a quote he presents from Plutarch, talking about Archimedes:

> It is not possible to find in all geometry more difficult and intricate questions, or more simple and lucid explanations. Some ascribe this to his natural genius; while others think that incredible effort and toil produced these, to all appearances, easy and unlabored results. No amount of investigation of yours would succeed in attaining the proof, and yet, once seen, you immediately believe you would have discovered it; by so smooth and so rapid a path he leads you to the conclusion required.

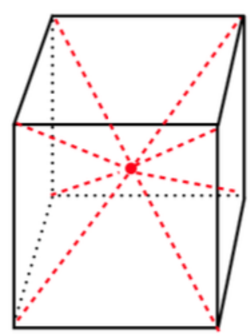
# Chapter 3

# Volume of a cone

We need a formula for the volume of a cone in order to find the volume of a sphere.

Let's start with something simpler, a pyramid with a square base. Consider a cube with all eight edges having length $s$. So each of the six faces is a square with sides of length $s$ and area $s^2$.

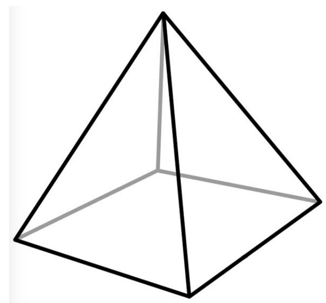
Label the central point inside the solid as $P$. Draw lines connecting each of the 8 external vertices to $P$, something like this.



Now we imagine slicing on planes that connect adjacent pairs of lines.

You can't do this in real life by slicing up a single cube or rectangular solid, because the cuts to form one surface would ruin some of the other pieces. The cuts must enter the solid at a corner and then pivot on a line ending at the exact center. (Perhaps you could do it with a *light saber* since the beam comes to a point).

The result is 6 identical pieces (square pyramids) looking something like this



This figure isn't quite accurate because our pyramids will have a height that is $s/2$, but just bear with me.

We started with a cube so that the six resulting solids would be identical. Unfortunately you can either have six pieces the same, or have some of the pieces with equal base and height, but you can't have both.

Let the six identical pyramid volumes each be $V$, and their sum is equal to the volume that we started with. We have that

$$6V = s^3$$
$$V = \frac{1}{6}s^3$$

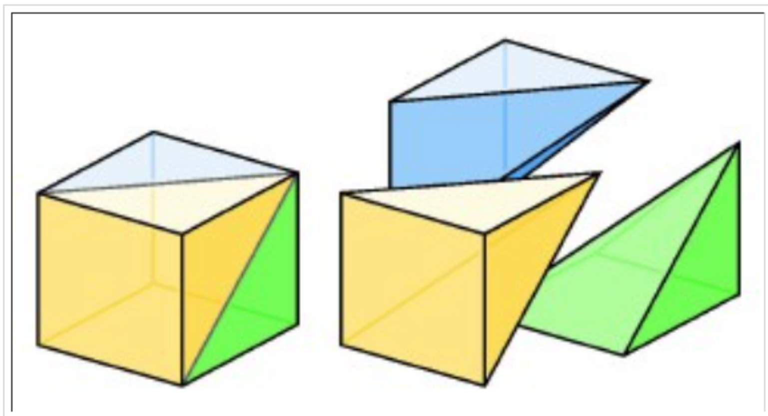
This is the volume for each pyramid with base area $s^2$ and height $s/2$.

The volume depends linearly on the height and the area of the base. The more general formula for a pyramid is really a linear function of $h = s/2$

$$V = \frac{1}{3}hs^2$$

and you can show this by starting with solids that are longer in one-dimension.

Here is an even better way to slice a cube



Three congruent pyramids meet along a diagonal of a cube.

When I first saw this, I thought it was a trick. But in fact, we have 3 identical right square pyramids.

The original cube has 12 edges. Each pyramid ends up getting three of those edges, all of them meeting at a vertex, plus it has two more edges along the base where there has been a cut, so the edge was shared.

In addition to those, there are two edges where a cut occurred along the diagonal of a face, and then finally the longest edge is (always the same) interior diagonal of the cube. The total number of edges is 8.

All three pyramids have a single one of the original external (square) bases, two faces that are one-half of an external face cut along the diagonal, and two faces that were originally internal. These latter two faces lie along the plane formed between the original interior diagonal axis and the diagonal cuts of the faces.

http://www.math.brown.edu/~banchoff/Beyond3d/chapter2/section02.

html

Of course, a pyramid is not a cone. But an argument identical to the one we will use for the sphere shows that the volume is independent of the shape of the base. It just depends on the area. So for a cone we finally obtain

$$V = \frac{1}{3}\pi r^2 h$$

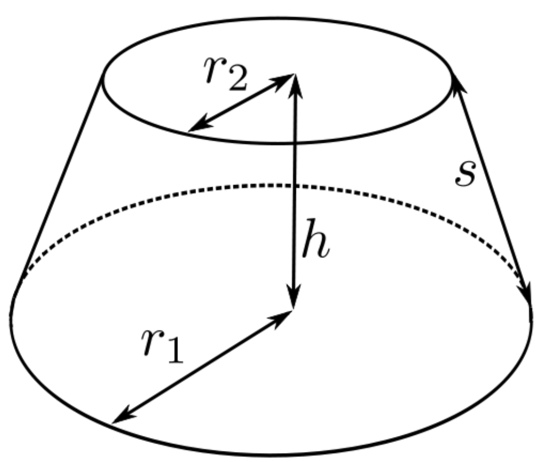
**algebraic derivation of the constant 1/3**

I found an algebraic argument on the web at

https://web.maths.unsw.edu.au/~mikeh/webpapers/paper47.pdf

Let us assume for this proof that the volume of a cone is proportional to both the area of the base and the height: $V = cAh$; our objective is to find the constant of proportionality.

Consider a conical frustum, a cone with the top lopped off.




Suppose the area of the base is $A$ and the height of the frustum is $h$.

Calculate the volume of the frustum as the difference between that of a larger cone with base $A$ and height $h + e$ (e for extra), and that of a small cone with base area $a$ and height $e$.

$$V = cA(e+h) - cae$$

Now, the area of the base of a cone is $\pi$ times the radius squared, and the radius is proportional to the height (depending on the slant angle). Hence

$$a = ke^2$$

The area of the small base is proportional to the height squared, and the same for the large one:

$$A = k(e+h)^2$$

so

$$k = \frac{a}{e^2} = \frac{A}{(e+h)^2}$$

Hence

$$\frac{\sqrt{a}}{e} = \frac{\sqrt{A}}{e+h}$$

Let us manipulate this expression to find $e$ in terms of $h$:

$$\frac{\sqrt{A}}{\sqrt{a}} = 1 + \frac{h}{e}$$

$$\frac{h}{e} = \frac{\sqrt{A} - \sqrt{a}}{\sqrt{a}}$$

$$e = \frac{\sqrt{a}}{\sqrt{A} - \sqrt{a}} \cdot h$$

And then

$$e + h = \frac{\sqrt{A}}{\sqrt{A} - \sqrt{a}} \cdot h$$

Substituting into what we had above for the volume:

$$V = cA(e+h) - cae$$

$$= cA\ [\frac{\sqrt{A}}{\sqrt{A}-\sqrt{a}} \cdot h\ ] - ca\ [\frac{\sqrt{a}}{\sqrt{A}-\sqrt{a}} \cdot h\ ]$$

$$= c\ [\ \frac{A\sqrt{A} - a\sqrt{a}}{\sqrt{A}-\sqrt{a}}\ ]h$$

This looks like a mess. But it is really $(m^3 - n^3)/(m-n)$ Factoring the numerator we get $m^2 + mn + n^2$. That is:

$$V = c(A + \sqrt{A}\sqrt{a} + a)h$$

Let $a \to A$. That is, consider what happens as $a$ gets larger and closer to $A$. The expression in parentheses becomes $3A$. Hence:

$$V = c(3A)h$$

But as $a \to A$ the frustum becomes a cylinder whose volume we know is equal to $Ah$.

$$V = c(3A)h = Ah$$

Therefore $c = 1/3$.

$\square$

We will revisit this problem, to use our first bit of calculus.

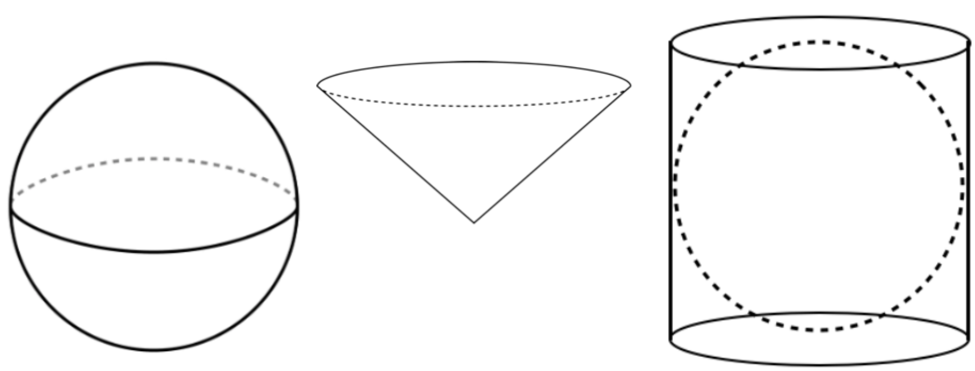
# Chapter 4

# Archimedes and the sphere

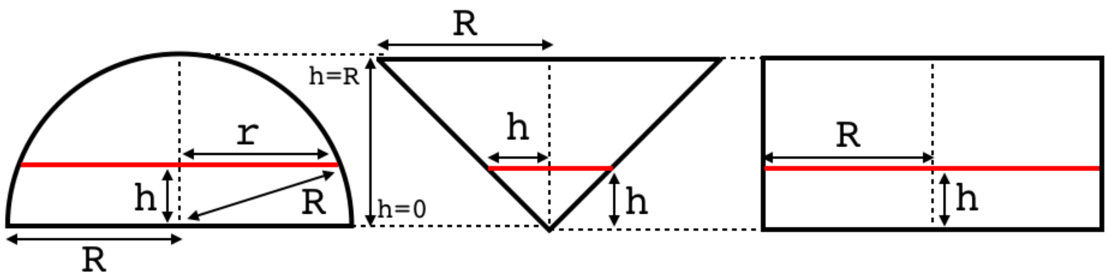
### volume of the sphere: geometry

The very first derivation of the volume of a sphere was discovered by Archimedes. The following is his "simple" but subtle argument.

We compare a half-sphere and an inverted cone to a cylinder.



Below is a diagram showing a vertical cross-section through the center of each solid so we can visualize the geometry. The radius $R$ is the same for all three. In addition, the cone and cylinder have overall height equal to $R$.




Now, imagine making a horizontal slice through each solid at a height $h$, shown by the red lines. We will choose different values of $h$ later and compare the results, the one shown here is arbitrary.

If you visualize this you should be able to see that each of these red slices is actually a circle. Any cross-section of a sphere is a circle. For the cylinder and cone, cross-sections perpendicular to the axis are circles as well.

The question we ask is: **what is the area for each horizontal slice**?

To answer that, we need to determine the radius for each red circle. Moving right-to-left, the radius of the cylinder is just $R$. For the cone, the radius at each height $h$ is equal to $h$, since $R = H$. And for the sphere, we use the Pythagorean theorem to find that

$$r^2 + h^2 = R^2$$

$$r^2 = R^2 - h^2$$

For more on this theorem see **here**.

The first insight of the proof is to recognize that the radius squared for the sphere's slice ($r^2$), plus the radius squared for the cone ($h^2$) is equal to $R^2$, the radius squared for the cylinder.

Since the areas are proportional to to the radius squared (namely $A =$

$\pi r^2$ and so on) and

$$\pi r^2 + \pi h^2 = \pi R^2$$

so the areas add too: **sphere plus cone equals cylinder**.

The second insight of the proof is to recognize that this property is invariant, it does not depend on which height we choose to make the slice. The three slices obtained at any height $h$ add up like this. So if we imagine making a bunch of slices for each solid and adding them all up to find the volume, the volumes will add too.

This idea is now called Cavalieri's principle, though it was called the "method of indivisibles" before that.

The volume of the cylinder is simply $\pi R^3$. The volume of the cone is known to be one-third the area of the base times the height, or $1/3\ \pi R^3$.

Subtract to find that the area of the half-sphere is $2/3\ \pi R^3$, and therefore the volume of the whole sphere is

$$V_{\text{sphere}} = \frac{4}{3}\pi R^3$$

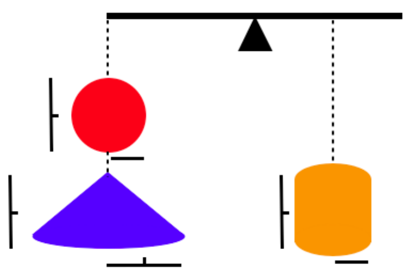
There is a bit of a trick here to hide the idea introduced in calculus, which makes this thinking rigorous. The sphere and cone have variable widths, which means that the radius will be different on the top of a slice compared to the bottom. Therefore, the slices have to be made very thin. In calculus they become infinitely thin, but we add up infinitely many of them.

Archimedes said that he discovered the correct result by balancing the three objects on a fulcrum.

According to Archimedes (in the Method, translation by Heath)

> For certain things which first became clear to me by a mechanical method had afterward to be demonstrated by geometry...it is of course easier, when we have previously acquired by the method some knowledge of questions, to supply the proof than it is to find the proof without any previous knowledge. This is a reason why, in the case of the theorems the proof of which Eudoxus was the first to discover, namely, that the cone is a third part of the cylinder, and the pyramid a third part of the prism, having the same base and equal height, we should give no small share of the credit to Democritus, who was the first to assert this truth...though he did not prove it.

I read somewhere that what Archimedes actually balanced is a set-up like that shown here



There are three factors that complicate our calculation: (i) we now have a single cone with radius $2r$ and height $2r$ (because it's doubled in both radius and height the cone's volume is increased by a factor of $2^3$), (ii) the sphere and cone are twice as far from the fulcrum as the cylinder, and (iii) the cylinder is made out of something denser than the other objects (four times more dense).

Let $\pi r^3$ be one unit of volume, then the volumes are

$$
\begin{aligned}
\text{sphere} &= \frac{4}{3} \\
\text{cone} &= \frac{1}{3} \times 8 = \frac{8}{3} \\
\text{cylinder} &= 2
\end{aligned}
$$

That's $12/3 = 4$ for the sphere plus cone, and furthermore they count double since they are twice the distance from the fulcrum, giving 8 in our volume units. So the left side is $4\times$ the weight on the right side.

However, we are told that the density of the material for the cylinder was four times that of the objects on the left. Hence, it should all balance.

I looked up some densities to try to guess what Archimedes used:

| | |
|---|---|
| marble | 2.56 |
| sand | 2.80 |
| copper | 8.63 |
| silver | 10.40 |
| gold | 19.30 |

How about marble and silver?

# Part II

# Numbers and proof

# Chapter 5

# Integers

## Integers

The *natural* or counting numbers which everyone learns very early in life are $1, 2, 3$ and so on.

One can get hung up on the question of whether the natural numbers would exist without the problem of counting three sheep or all ten of our fingers. Leopold Kronecker said "God made the integers; all else is the work of man".

We will not worry about they come from.

Mathematicians refer to the *set* of natural numbers and give this set a special symbol, $\mathbb{N}$. We write

$$\mathbb{N} = \{1, 2, 3 \ldots\}$$

The brackets contain between them the *elements* of the set. The symbol $\in$ means "in the set" or "is a member of the set".

The dots mean that this sequence of numbers continue forever.

It seems hard to know how we can decide whether a particular $n$ is in the set if we can't enumerate all the members of the set. But we can decide by its form whether $n$ is a natural number or not. If this seems problematic, you might call $\mathbb{N}$ a *class* instead (Hamming); we carry out classification to decide whether $n$ is a natural number.

The notion of an unending sequence can be unnerving.

**construction of N**

To construct the set $\mathbb{N}$, start with the smallest element, 1. Then add successive elements by forming $a_{n+1} = a_n + 1$.

Sometimes people say that

$$0 \in \mathbb{N}$$

(0 is a part of the set), but most do not, and we will follow the definition given above. If you wanted to be explicit about this you could write

$$0 \notin \mathbb{N}$$

$\mathbb{N}$ is an infinite set, meaning that there is no largest number in $\mathbb{N}$, no largest $n \in \mathbb{N}$.

Proof: the proof is by contradiction.

We assume that the symbols and operations $>$ and $<$ are defined.

Suppose $\mathbb{N}$ does have a largest number, $a$. Well, what about $a + 1$? By the definition we can construct it, and it is clearly also a member of the set, but $a_{n+1} > a_n$ so $a_n$ is not the largest number in the set.

$\square$

What do we mean by infinity? We mean a kind of bound on the numbers. All numbers $n \in \mathbb{N}$ have the property that $n$ is contained in the interval $[1..\infty)$. The right parenthesis means $\infty$ is *not* part of the interval. $\infty$ is *not a number*!

### least element

A second important property of $\mathbb{N}$, as mentioned, is that there is a least number in the set. If pairwise comparisons are carried out, a single element, the number 1, has the property that $1 \leq n$ for all numbers $n \in \mathbb{N}$.

### well-ordered property

Since we can also find the least member of the set excluding 1, written $\mathbb{N} \setminus 1$, we can order every number in $\mathbb{N}$.

This property is called the **well-ordered** property.

### infinity is not a number

One way of approaching this is to say, what goes wrong when we attempt to divide by zero?

$$\frac{a}{0} = ?$$

Suppose

$$\frac{a}{0} = c$$

$$a = 0 \cdot c$$

Equivalently, what number when multiplied by 0 would result in $a$?

If there were such a number (say $\infty$), then what about

$$\frac{b}{0} = ?$$

$$\frac{c}{0} = ?$$

By definition $0 \cdot a = 0$. By definition, we do not allow division by zero. And, by definition, *infinity is not a number.*

**limits**

Some people say that calculus is all about limits. We will keep the discussion of limits and $\epsilon$-$\delta$ formalism to a minimum. But let us try to establish an intuitive idea about what we mean when we say "in the limit as $N \to \infty$".

Above we had that there is no greatest integer.

A corollary of that is that in the limit

$$\lim_{n\to\infty} \frac{(n+1)-n}{n} = ?$$

As $n$ increases without bound, the difference between successive numbers, as a fraction of $n$, tends to zero.

To get an idea about this, first simplify by multiplying by $1/n$ on top and bottom. Then

$$\lim_{n\to\infty} \frac{(1+1/n-1)}{1} = \frac{1}{n} = 0$$

We say that $1/n$ *tends* to zero as $n$ approaches $\infty$, and so does $[(n+1)-n]/n$.

**the Integers**

The set $\mathbb{Z}$ contains all the members of $\mathbb{N}$ plus their negatives, as well as the special number 0, often called the additive identity.

$$\mathbb{Z} = \{\cdots - 2, -1, 0, 1, 2, \dots\}$$

$\mathbb{Z}$ stands for the German word *Zahlen*, Number. The set $\mathbb{Z}$ are usually referred to as the integers.

$\mathbb{Z}$ is also an infinite set and also has the well-ordered property. To show this simply order all numbers $p > 0$ with respect to zero using $<$, and all the numbers $n < 0$ using $>$.

**inequality**

In the section above we used the symbols $>$ and $<$, greater than and less than, without introducing them. I'm sure you've seen and used them before. Among the axioms of the number systems is the collection of *order axioms*. As an example:

- $x < y$ means that $y - x$ is positive
- $y > x$ means that $x < y$

For arbitrary numbers $a$ and $b$ one of three statements is true:

$a < b$, $a = b$ or $a > b$.

There is no intent to be systematic here. Let us just mention that these properties (and their kin) are true not just for natural numbers, but also for the rational numbers and the real numbers, which we will talk about soon. Here are just a few more important theorems in this class:

- If $a < b$, and $c$ is any number, then $a + c < b + c$
- If $a < b$, then $-b < -a$
- If $a < b$ and $c > 0$, then $ac < bc$

# Chapter 6

# Induction

We can visualize an inductive proof as a kind of chain. We show that the "base case" is true, for some value of $n$. Then we show that if the formula works for $n$ , it must work for $n + 1$.

> Mathematical induction proves that we can climb as high as we like on a ladder, by proving that we can climb onto the bottom rung (the basis) and that from each rung we can climb up to the next one (the step).

- Graham, Knuth and Patashnik

[ There is a variant called *strong* induction where we assume some statement is true for *all* $0 < k \leq n$. ]

**the problem**

Suppose we have some theorem that we *think* might apply to all $k \leq n$. A classic example (Courant and Robbins?) is:

$$f(n) = n^2 - n + 41$$

The function $f(n)$ produces a prime number for $n \in [1..40]$, which is apparent by inspection (because 41 is prime). But for $n = 41$ all terms are divisible by 41, thus the result cannot be prime.

Hamming has some other examples. Here is one:

$$f(n) = n(n-1)(n-2)\ldots(n-k)$$

$f(n) = 0$ for all $0 \leq n \leq k$, but will never be zero for any other $n > k$! By choosing $k$ large, we can make the number of true cases as large as you like.

Furthermore, for any function $g(n)$, $f(n) + g(n)$ will have the same property.

### proof of induction

According to Hamming, if you are not convinced by the ladder analogy, here is another proof that induction works:

> Suppose the statement is not true for every positive (non-negative) integer. Then there are some false cases. Consider the set for which the statement is false. *If* this is a non-empty set, then it would have a least integer, which is $m$. Now consider the preceeding case, which is $m-1$. This $(m-1)$th case must be true by definition, and we know that there is such a case because as a basis for the induction we showed that there was at least one true case. We now apply the step forward, starting from this true case $m-1$, and conclude that the next case, case $m$, must be true. But we assumed that it was *false*! A contradiction.

Therefore, there are no false cases.

**sums of integers**

Later in the book, we will compute Riemann sums, and to do that we need to find formulas for the sum of integers, the sum of square integers, and so on.

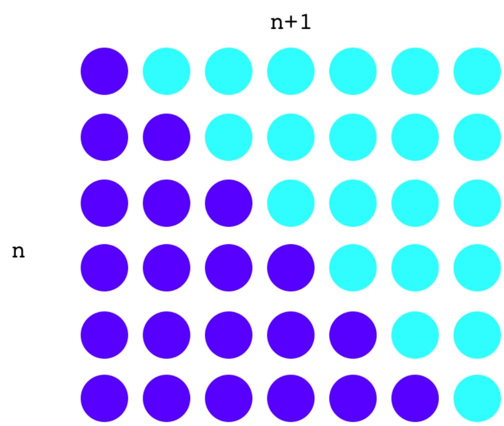
To keep it simple, let's start with finite sums like the integers from 1 to $n$

$$1 + 2 + 3 + \cdots + n$$

The numbers we seek are called the triangular numbers. These are

$$1, 3, 6, 10, 15 \cdots$$

Here is a striking "visual proof" of the formula to obtain $\mathrm{T}_n$, the $n^{th}$ such number. The total number of circles in the figure below is $n \times (n+1)$ and this is exactly two times the sum of the integers from 1 to $n$.




$$2S = n(n+1)$$

There is a famous story about Gauss that, as a schoolboy, he "saw" how to add the integers from 1 to 100 as two parallel sums

$$1 + \quad 2 + \quad 3 + \cdots + 99 + 100$$

$$100 + 99 + 98 + \cdots + 2 + 1$$

Added together horizontally, these two series must equal twice the sum of 1 to 100. But in the vertical, we notice that we have $n$ sums, each of which is equal to $n + 1$. So, again

$$2S = n(n + 1)$$

$$S = \frac{1}{2}\, n(n + 1)$$

For $n = 100$ the value of the sum is 5050. Another way of looking at this result is that between 1 and 100 there are 100 representatives of the "average" value in the sequence, which (because of the monotonic steps) is $(100 + 1)/2 = 50.5$.

Or alternatively, view the sum as ranging from 0 to 100 (with the same answer). Now there are 101 examples of the average value $(100{+}0)/2 = 50$).

**Proof of the formula $n(n + 1)/2$ by induction**

Returning to the sum of integers, one proof follows the method of induction. In this approach, however, one must first guess the correct formula. We guess $n(n + 1)/2$, of course.

For this example of induction, we will assume the formula is true for $n - 1$ and show that if so, it is true for $n$. Hamming says this may be easier sometimes than starting with $n$.

So, we *assume* that the answer is correct for $n - 1$. In this case, the formula changes to

$$S_{n-1} = \frac{n(n - 1)}{2}$$

So clearly, if $S_{n-1}$ is correct, then

$$S_n = S_{n-1} + n$$

Follow out the arithmetic:

$$= \frac{n(n-1)}{2} + n$$

$$= \frac{n(n-1) + 2n}{2}$$

$$= \frac{(n^2 + n)}{2}$$

$$= \frac{n(n+1)}{2}$$

But this is precisely what we would obtain by using the formula, and substituting $n$ for $n-1$. Hence the formula gives the correct result for $n$, assuming that it gives the correct result for $n-1$.

In turn, it gives the correct result for $n-1$, assuming it gives the correct result for $n-2$. Eventually, we reach the base case, where we can verify that the result is correct.

Try it on the first value in the sequence ($n = 1$, the "base case").

$$\frac{1(0)}{2} = 0$$

That checks. If you're worried check $n = 2$:

$$\frac{2(1)}{2} = 1$$

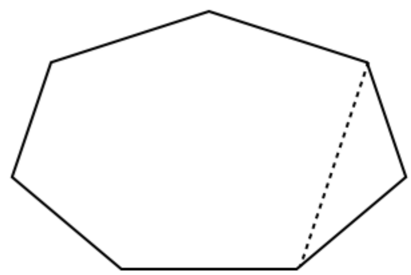
So the whole chain of reasoning is correct.

□

**geometric example**

In the figure below we have a polygon—an irregular heptagon. Actually, there are three polygons altogether, there is the heptagon with $n+1$ sides, the hexagon with only $n$ sides that would result from cutting along the dotted line, and the triangle that is cut off.

What we would like to do is to find a formula for the sum of the internal angles that depends only on the number of sides or vertices.



The first part of the answer is to guess. In the figure, you can see that by adding the extra vertex to go to the $n+1$-gon, we added a triangle, or perhaps you'd rather say than in going from $n+1$ to $n$ we lost a triangle.

In either case, the difference is 180°. The difference between having $n$ sides and $n+1$ sides is to add 180°.

The second part of the argument is to suppose that $n = 3$, in that case we must have simply 180° degrees for a triangle. So we guess that the formula may be

$$(n-2)180^\circ = S_n$$

where S is the sum of the angles in an $n$-gon.

We can use induction to prove that this formula is correct.

The proof has two parts. We must verify the formula for a base case like the triangle, which we've done. You may wish to check that it works for the square as well, but that's not strictly necessary.

The second part of the proof is to verify that in going from $n$ to $n+1$, we add another $180°$.

$$(n-2)180° + 180° \stackrel{?}{=} ((n+1)-2)180°$$

On the left-hand side, we have the sum of angles for $n$ sides, which we assume is correct, and then we just add $180°$ to it. On the right, we have substituted $n+1$ into the formula.

Now we need to show that these are equivalent. But of course

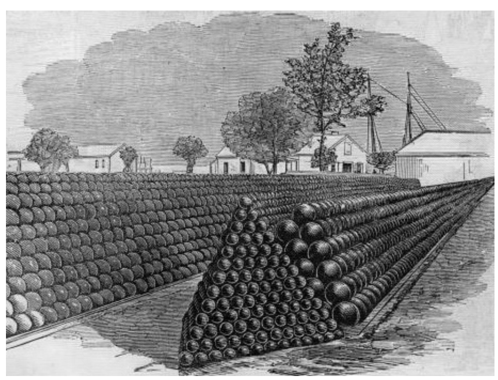
$$(n-2)x + x = ((n+1)-2)x$$

$$n - 2 + 1 = n + 1 - 2$$

$\square$

That is the inductive proof of the formula.

**sum of integer squares**



Have you ever seen a square stack of marbles, or cannonballs? The number of elements on each level is the square of a natural number. We ask, what is the total number:

$$S = 1^2 + 2^2 + \cdots + n^2 = ?$$

The answer is not obvious, but we will see how to prove it later using a method called induction. We will also see how to deduce it. For now it is enough to give the result:

$$S = \frac{1}{3} \cdot n \cdot (n + \frac{1}{2}) \cdot (n + 1)$$

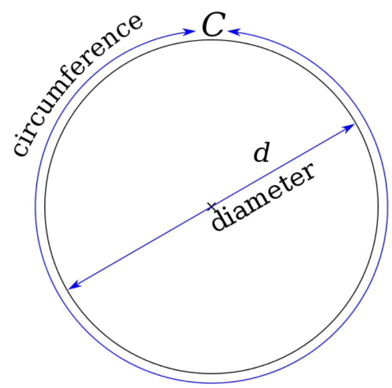
which is sometimes given as

$$S = \frac{n(n + 1)(2n + 1)}{6}$$

Given the formula for the volume of a cone or pyramid, the factor $1/3$ is not surprising.

# Chapter 7

# Pi is a constant




We began the book with a bold claim: the ratio of the circumference of a circle to its diameter is a constant, independent of the length of the diameter:

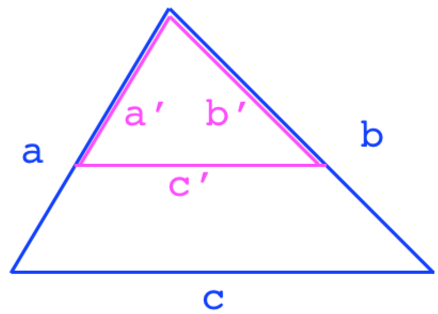
$$\pi = \frac{C}{d} = \frac{C}{2r}$$

We did not prove this theorem at the time but will do so now.

We need the idea of limits, which was introduced previously, and a property of similar triangles.

Let us first define similarity: two triangles are similar if all their three angles are equal. That is, one is a scaled-down version of the other.

The theorem is: if two triangles are similar, then their sides are proportional to each other.




Draw a horizontal line parallel to the base. Then the resulting small triangle has its sides in proportion to the original one as:

$$\frac{a'}{a} = \frac{b'}{b} = \frac{c'}{c}$$

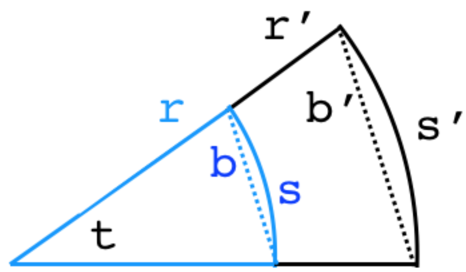
Now we can prove that $\pi$ is a constant.

Proof:

Consider two circles of different sizes on the same center, with inner radius $r$ and outer radius $r'$.

Divide the circles into $n$ equal sectors each with central angle $t$.

Then the arc length $s = C/n$ or $s' = C'/n$. For any particular $n$ we have that $C/s = C'/s'$. Thus, it suffices to show that the ratio $s/r = s'/r'$ (i.e. is constant) to prove the theorem.




In the limit as $n \to \infty$, i.e., as the example sector gets smaller and the total number of sectors gets very large, $b \approx s$ and $b' \approx s'$.

This is just Archimedes' argument, that as the number of sides of an inscribed regular polygon increases without limit, the perimeter of the polygon will be equal to the circumference of the circle. Therefore as $n \to \infty$

$$s = b; \quad s' = b'$$

But the two triangles are similar, because they share the angle $t$ and are both isosceles. Therefore, the ratio $b/r$ is equal to the ratio $b'/r'$ and then

$$\frac{s}{r} = \frac{s'}{r'}$$

so

$$\frac{C}{r} = \frac{C'}{r'}$$

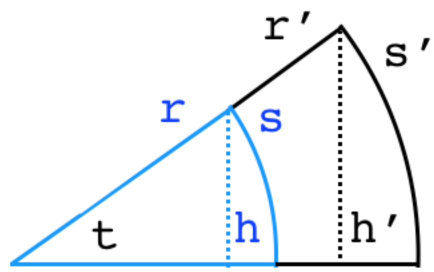
This completes the proof.

□

Second proof:

Here is a simple variant which assumes something we will prove in the section on sine and cosine. If this is confusing, it can easily be skipped.

Drop the altitude $h$ in each of the two similar triangles. The ratio $h/r$ is equal to $\sin t$, but the arc length $s = t$, measured in radians.




In the limit that $n \to \infty$, the ratio between $s$ and $h/r$ is equal to our "special limit":

$$\lim_{n\to\infty} \frac{t}{\sin t} = 1$$

If the ratio to the sine is equal to 1, so is the ratio to its inverse and thus the ratio $s/r$ is constant, which is what we wanted to prove.

$\square$

### Pi is irrational

This proof is too challenging for this book. You can read about it in wikipedia, or

https://mindyourdecisions.com/blog/2013/11/08/proving-pi-is-irrational-a-step-by-step-guide-to-a-simple-proof/

# Part III

# Lines and triangles

# Chapter 8

# Euclid

### Euclid and the postulates

Greek geometry starts hundreds of years before Euclid, whose life overlapped (on both ends) that of Alexander the Great. Euclid died about 270 BC, although his life and death are shrouded in mystery.

Euclid's book *Elements* is still an excellent place to begin surveying the foundations of geometry.

Euclid's geometry is mainly constructions (geometric figures) drawn with a pencil on a flat piece of paper, using a straight-edge or a compass or both.

Here are Euclid's first three postulates — statements that are assumed to be true:

∘ A straight line segment can be drawn joining any two points.

∘ Any straight line segment can be extended indefinitely in a straight line.

∘ Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.

Let us assume these as well.

We finesse the difficulty in defining what is meant by *straight* in the real world. If you've ever done any carpentry, for example, you probably know that unknown edges are determined to be straight by comparison with known straight edges. But the mental construct of "a straight line is the shortest distance between two points" gets around this limitation.

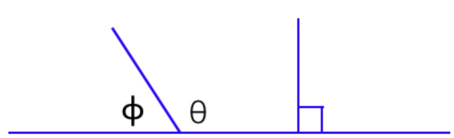
The fourth postulate is:

∘ All right angles are congruent, that is, equal to each other.

This one prompts a different question: what is a "right angle"?

If a line segment is drawn with one end on a line, let us refer to the two angles the line segment forms with the line as adjacent angles.

The definition of a right angle is this: if these two adjacent angles are equal, then they are both right angles.




On the left, one of the angles, $\phi$, is smaller than the other one, $\theta$.

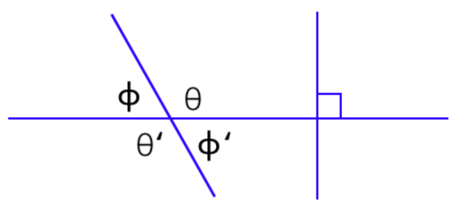
Alternatively, on the right, the two angles have equal measures. In this case we can conclude that both angles are right angles. A right angle is frequently designated by drawing a small square at the intersection. Since both angles are right angles, only one square is needed or usually drawn.

In all cases, the sum of the two angles $\phi + \theta$ is equal to two right angles or 180 degrees. There is nothing particularly special about 180 degrees for two right angles or 360 degrees for one whole turn.

Well, there is one thing: there are *approximately* 360 days in a year, which marks the sun's track across the sky. In his book, *Measurement*, Lockhart adopts the convention that one whole turn is equal to 1.

Later, we'll see that one whole turn is defined as $2\pi$ radians, and that convention turns out to be quite important.

Now, extend those lines below the horizontal




We said that the sum of the two angles $\phi + \theta$ is equal to two right angles, but so is the sum $\theta + \phi'$, for the same reason.

$$\phi + \theta = \theta + \phi'$$

We conclude that $\phi = \phi'$ and $\theta = \theta'$. On the right, if any one of the angles where two lines cross is a right angle, then all four are right angles.

This is called the vertical angle theorem.

There is a simple method to construct a line segment perpendicular to a line at a particular (given) point, or alternatively, through any point not on the line (see the video at the url):

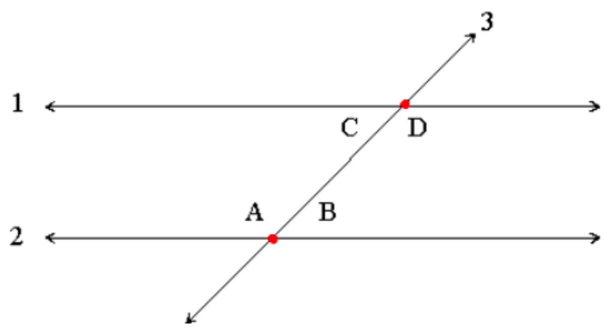
https://www.mathopenref.com/constperpextpoint.html

### parallel postulate

All this seems rather obvious.

The fifth and final postulate is more subtle:

◦ If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines inevitably must intersect each other on that side if extended far enough.




Line 1 and line 2 are parallel, if and only if, $A+C = B+D = 180 = 2$ right angles. This postulate is equivalent to what is known as the parallel postulate.

http://mathworld.wolfram.com/EuclidsPostulates.html

Consider a familiar situation where this is not true. Suppose we are doing geometry on the surface of a sphere, such as the earth. Two adjacent lines of longitude can be drawn so as to cross the equator at right angles, and the lines are parallel there, but they meet (intersect) at the poles.

The parallel postulate only holds for geometry on a *flat* surface.

### axioms

Euclid also lists five axioms. Here are two examples:

∘ Things that are equal to the same thing are also equal to one another.

∘ If equals are added to equals, then the wholes are equal.

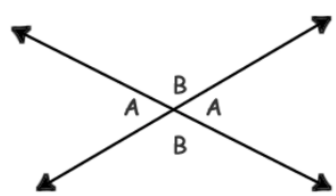
We will see how to proceed from the postulates and axioms to various proofs. *Given these assumptions*, we can prove theorems that must be true.

### Thales

I'm a big fan of William Dunham's books — three of them are listed in the References.

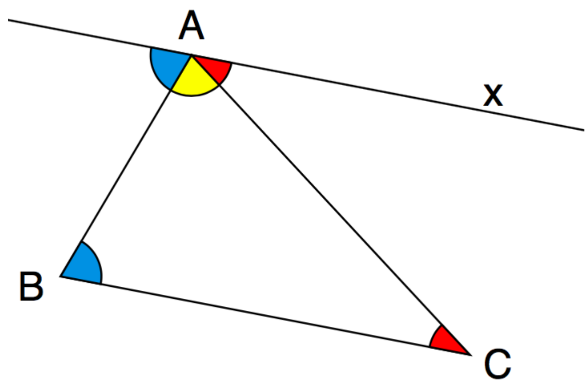
Dunham has written a lot about the history of mathematics in Greece, starting with Thales (624-546 BC), who was from a Greek town called Miletus on the coast of Asia Minor (modern Turkey). He lived long before Euclid. Although none of his writing survives, it is believed that Thales proved several early theorems including one we saw above

∘ The vertical angles formed by two straight lines crossing, are equal.



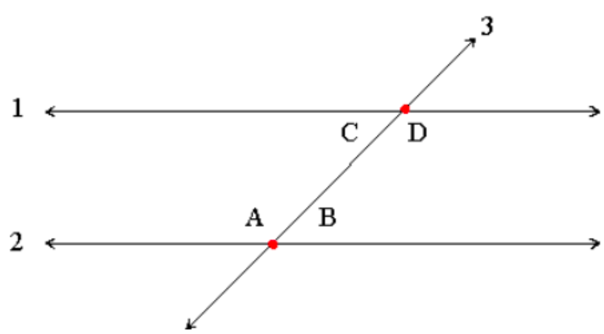

This theorem depends on a property of straight lines. In the proof, we used the axiom "equals added to equals are equal", alternatively "equals subtracted from equals are equal."

∘ The angle sum of a triangle is equal to two right angles.




This theorem depends in turn on a theorem which we laid the groundwork for above but did not state explicitly.

In the figure below, if 1 is parallel to 2, we said that $A+C=B+D=180$ degrees.




But we also know from the properties of two lines given above that $A+B=180$ degrees. So

$$A+C=180=A+B$$

and then

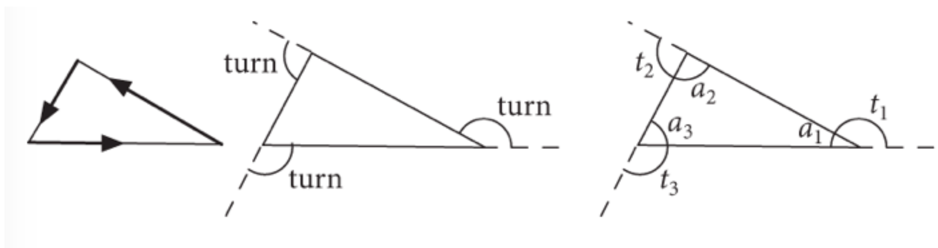
$$C=B$$

This is called the theorem on alternate interior angles. Given this, you can go back to the angles of a triangle problem and follow the colors to the proof.

**another proof**

Here is an alternative proof of the theorem on the sum of angles in a triangle adding to 180 degrees..

Imagine walking around the perimeter of a triangle in the counter-clockwise direction. At each vertex we turn left by a certain number of degrees, $t$, called the exterior angle. After passing through all three vertices, we must end up facing in the same direction as we started.




$$t_1 + t_2 + t_3 = 360$$

In addition, for each vertex, the interior angle $a_i$ plus the exterior angle $t_i$ add up to 180 degrees. If we add up all three pairs, we obtain

$$(t_1 + a_1) + (t_2 + a_2) + (t_3 + a_3) = 3 \cdot 180 = 540$$

By subtraction

$$a_1 + a_2 + a_3 = 180$$

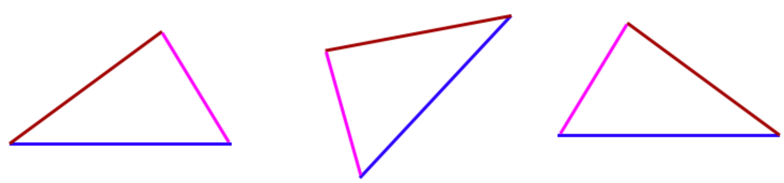
# Chapter 9

# Congruent triangles

### Congruence and similarity of triangles

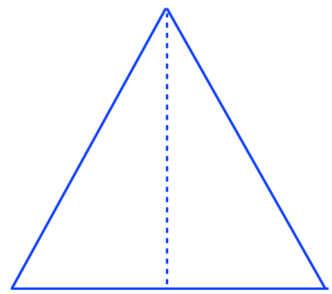
∘ Two triangles are *congruent* if and only if they have the same three side lengths. This is often abbreviated SSS (side-side-side).

By this definition, a triangle and its mirror image are congruent. The three triangles shown below are all congruent, even though the first is flipped (it is the mirror image of the other two).



Having the same three sides means that the shape is the same, and all three angles are the same — the shapes are superimposable, with the proviso that we allow the shape to be flipped over.

In this figure the two smaller triangles obtained by dividing an equilateral triangle in half, are congruent.

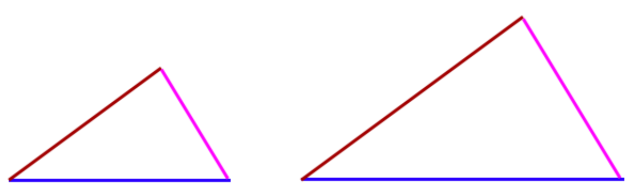


Some triangles are *similar* but not congruent, with all three angles the same but of different overall sizes. We could call this AAA (angle-angle-angle). For similar triangles, the three corresponding pairs of sides are in the same proportions, but re-scaled by the same constant of proportion.

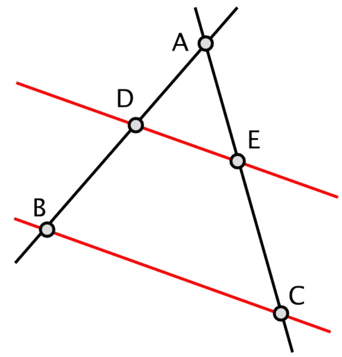
◦ Two triangles are similar if they have the same three angles.

Because of the angle sum theorem, if any two angles of a pair of triangles are known to be equal, then the third one must be equal as well.

Similar triangles have their sides in the same proportions.



Given any triangle, draw a line parallel to one side, which also joins the other two sides. The new triangle with that side as its base is similar to the given triangle. Similarity means that all the angles are equal. This is easily proved using the theorem on alternate interior angles.




In this example, these ratios are all equal

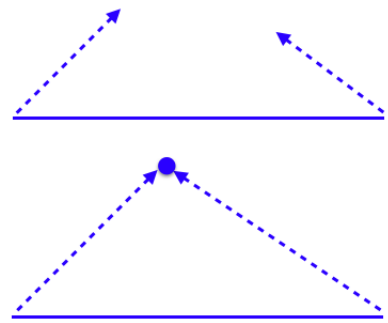
$$\frac{AD}{AB} = \frac{AE}{AC} = \frac{DE}{BC}$$

In addition to SSS (side-side-side), there are other conditions that lead to congruence of two triangles when they are satisfied, namely

- SAS (side-angle-side)
- ASA (angle-side-angle)
- AAS (angle-angle-side)

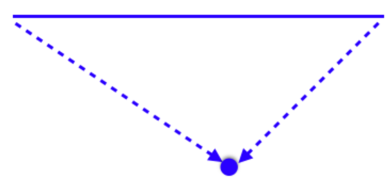
### constructions

Again, the way I think about these conditions is to imagine trying to construct a triangle from the given information, and ask whether it is uniquely determined. Suppose we know ASA. The situation is thus:



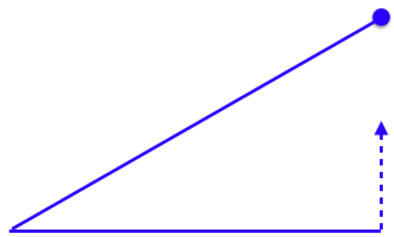
Plot the known side and start two other sides from the ends of that side containing the known angles. They must cross at a unique point.

But... actually, if we start the two lines pointing below the horizontal, there is another solution, the mirror image. This triangle is also congruent to the one above.



Alternatively, knowing two angles means we also know the third, because they must add to 180 degrees. For this reason, ASA and AAS imply that we have exactly the same information, because we know all three angles and (this part is important) we also know *which* two angles flank the known side.

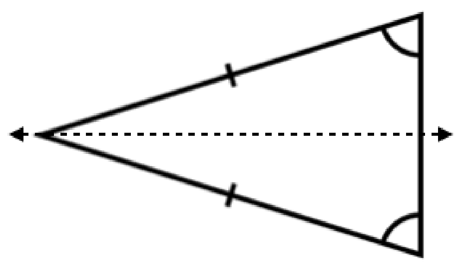
For a right-triangle, if the hypotenuse and one leg are equal, the two triangles are congruent.



In the figure, imagine the hypotenuse swinging on the hinge of its vertex with the horizontal base. There is only one angle where it will terminate on the vertical side with the correct length. This determines the angle between the known sides, or alternatively, the length of the third side.

**another theorem from Thales**

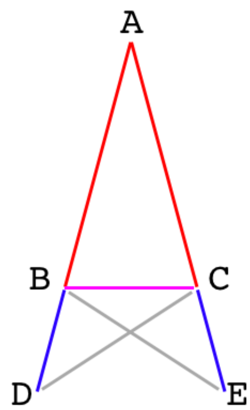
∘ The base angles of an isosceles triangle are equal.



My favorite proof of this theorem is from symmetry. Draw a line from the vertex between the two equal sides to the midpoint of the base opposite. If you turn the triangle over along this axis, we obtain the same triangle back again.

Alternatively, just say AAS or use the previous theorem on right triangles.

Euclid's proof is here:



AB = AC (given)
AD = AE (given)
ΔABE = ΔACD (SAS)

BD = CE (subtraction)
BE = CD (congruent Δ)
ΔBCE = ΔBCD (SSS)

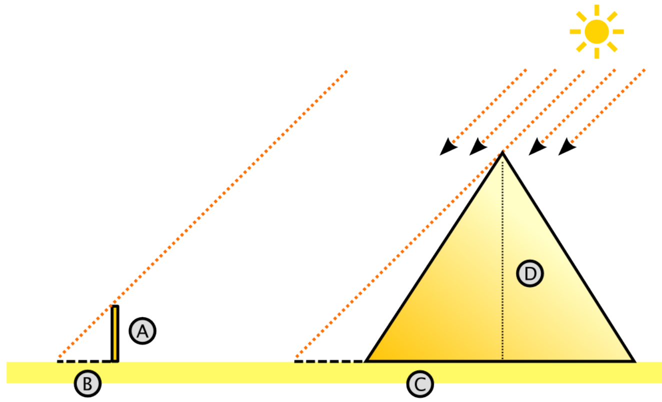
∠BCE = ∠CBD (congruent Δ)
∠ABC = ∠ACB (subtraction)

**pyramid height**

As we said, Thales was from Miletus and he lived around 600 BC. Thales is believed to have traveled extensively around the Mediterranean and was probably of Phoenician heritage, famous sailors.

During his travels, he went to Egypt, home to the great pyramids at Giza, which were already ancient then. They were built just around around 2560 BC (dated by reference to Egyptian kings) and were already 2000 years old at that time!

The story is that Thales asked the Egyptian priests about the height of the Great Pyramid of Cheops, and they would not tell him. So he set about measuring it himself. The current height is 480 feet. He used similar triangles.




**platonic solids**

https://en.wikipedia.org/wiki/Platonic_solid

> In three-dimensional space, a Platonic solid is a regular, convex polyhedron. It is constructed by congruent (identical in shape and size) regular (all angles equal and all sides equal) polygonal faces with the same number of faces meeting at each vertex. Five solids meet these criteria.

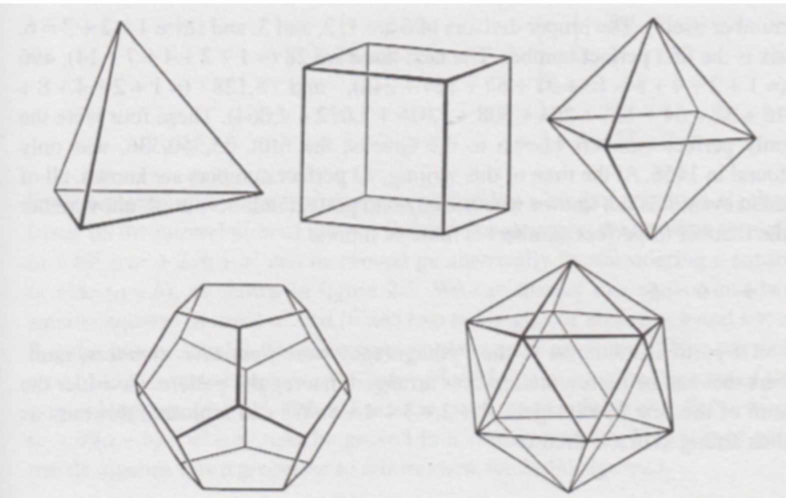


Figure 2.3. The five regular or Platonic solids

These are: (i) tetrahedron, (ii) cube, (iii) octagon, (iv) dodecagon, and (v) icosahedron.

There is a wonderful, simple proof that there are only five of them. Any solid requires at least three sides meeting at each vertex, otherwise the joint between two sides can just flap, like a hinge. Furthermore, the total of all the vertex angles added up must be less than 360 degrees, since otherwise the figure would be planar, not 3-dimensional.

So, three equilateral triangles total $60 \times 3 = 180$, four total $60 \times 4 = 240$ and five total $60 \times 5 = 300$. Six would be a hexagon lying in the plane. Three squares total $90 \times 3 = 270$, while four give a square array in the plane. Finally, three pentagons give $108 \times 3 = 324$. And that's it. Three hexagons would give $120 \times 3 = 360$, which gives an array in the plane.

Proving that all the angles and side lengths come out correctly, so that the possible solids actually can be constructed is another matter, however. Euclid devotes book XIII of *The Elements* to this:

https://mathcs.clarku.edu/~djoyce/elements/bookXIII/bookXIII.html#props

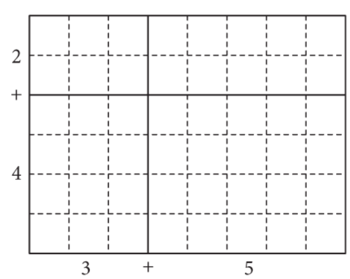
# Chapter 10

# Area

One aspect of calculus will be to determine the area of figures in the plane, particularly figures bounded by curves, as well as volumes in space. This is the magic of calculus, that we can make curves conform to rectilinear concepts of area and volume.

Since this introductory section is about Euclidean geometry, let's just say a few words about the area of a triangle. But we'll start with the rectangle.

To find the area of a rectangle, we must first fix a unit length. Then multiply the width by the height.




This particular figure (from Lockhart) shows the distributive law in action:

$$
\begin{aligned}
&(3+5)\cdot(4+2)\\
&=3\cdot 4+3\cdot 2+5\cdot 4+5\cdot 2\\
&=48
\end{aligned}
$$

Any combination of numbers that add up to 8, times any combination of numbers that add up to 6, gives the same result.

The next figure is a parallelogram, a four-sided figure whose two pairs opposite sides are parallel (left panel). As a consequence of the theorems we saw previously, the opposing angles are equal, and the adjacent angles add up to 180 degrees.


To find the area, we cut off a right triangle from the left and re-attach it on the right. The angles add up to form a straight line along the base and a right triangle at the upper right. The area is clearly $h \times b$.

What about triangles? Well, every triangle can be turned into a parallelogram, by attaching a rotated image of itself, like this:

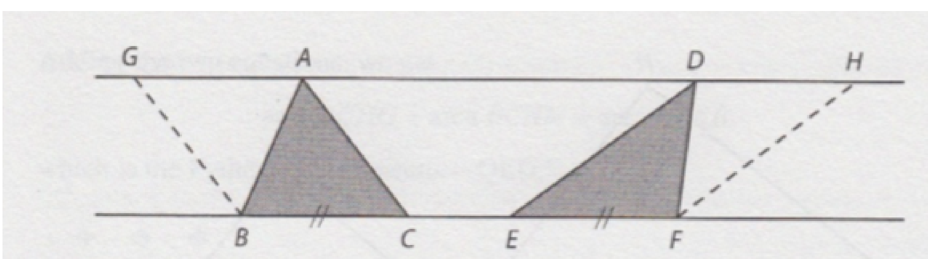
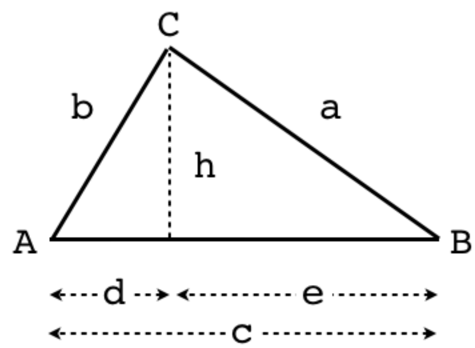



Figure 3.1. Proposition I 38: Triangles *ABC* and *DEF* have the same area

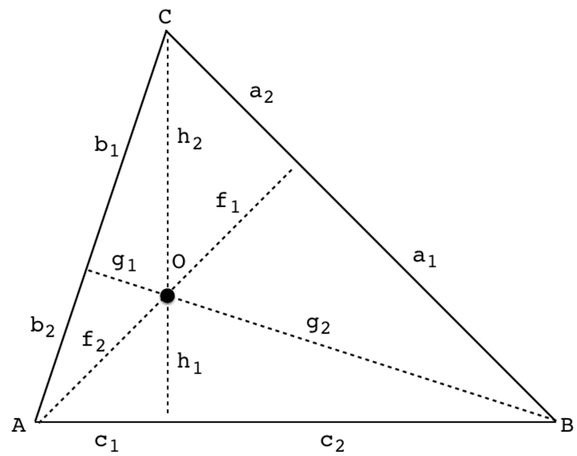
An acute triangle is on the left and an obtuse triangle on the right. Since the area of each triangle is one-half that of its corresponding parallelogram (because we added the same area to make the parallelogram), the area of a triangle is one-half the base times the height.




Here, the area is $hc/2$.

We could choose any side of the triangle to be the base and then multiply $1/2 \times$ base $\times$ height to get the area. We must always get the same answer!

If you accept the argument about the parallelogram above, it must be true, because the area of the triangle has to be the same no matter how you calculate it. Here's a proof:




In $\triangle ABC$ with sides $a, b, c$, drop the three altitudes from each of the three vertices to form right angles on the opposing sides. Ceva's theorem says that these altitudes cross at a single point (we will prove this later). Label the parts of the sides and the altitudes as shown in the diagram.

The area of the whole $\triangle ABC$ is equal to the sum

$$\triangle BOC + \triangle AOC + \triangle AOB$$

Using the rule, *twice* the area is

$$2A = af_1 + bg_1 + ch_1$$

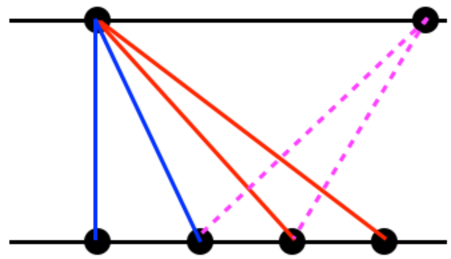
But each of these smaller areas can be computed in different ways. In particular $\triangle BOC$ can be viewed as having base $g_2$ and height $b_1$, while $\triangle AOB$ can be viewed as having base $b_2$ and height $g_2$, so (twice) the total area is also

$$\begin{aligned} 2A &= b_1g_2 + b_2g_2 + bg_1 \\ &= bg_2 + bg_1 = bg \end{aligned}$$

Similar calculations can be carried out for the other two sides. Hence the area is the same regardless of which side is chosen as the base.

□

A corollary is that all triangles with the same base and height have the same area.



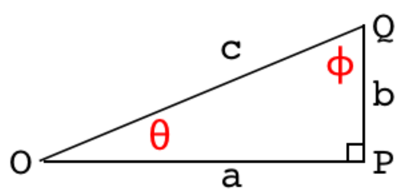
# Chapter 11

# Angle bisector

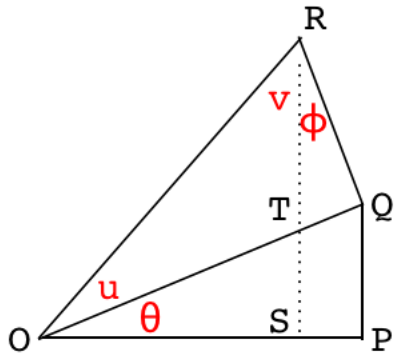
The main result we are headed for is the Pythagorean Theorem. Before we get there, however, it is worthwhile to continue our development of basic geometry with a discussion about right angles and right triangles.

A right triangle is a triangle containing one right angle. Right angles (and right triangles) are special. We saw previously that the definition of a right angle is that two of them add up to one straight line or 180 degrees. Since we proved that the sum of the three angles in any triangle is equal to one straight line, by extension, the sum of angles in any triangle is also equal to two right angles.

In the figure below, the angle at vertex $P$ is a right angle. It is common to mark a right angle with a little square, as shown, but these are a pain to draw, so I will not usually do that. The side opposite $P$, namely $c$, is the hypotenuse.



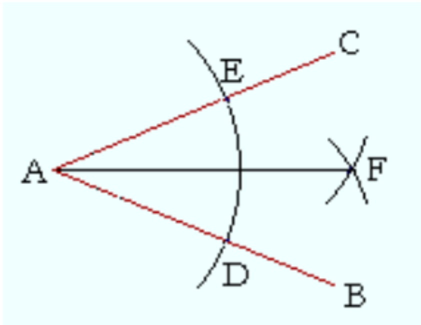
Since the sum of angles in a triangle is equal to two right angles, the sum of the angles $\theta$ and $\phi$ above is also equal to a right angle, or 90 degrees. Angles $\theta$ and $\phi$ are said to be complementary. This fact is often exploited in proofs. Here is an example we will see later on:




Suppose we are given that $\angle OPQ$ and $\angle OQR$ are right angles. We draw the altitude $RS$ and observe that the angle at vertex $S$ is a right angle. Therefore, in triangle $ORS$, the sum $\theta + u + v$ is equal to one right angle. At the same time, in triangle $OQR$, the sum $u + v + \phi$ is also equal to one right angle. Therefore, $\theta = \phi$. Further, $\triangle QRT$ and $\triangle OPQ$ are similar triangles.

**angle bisector**

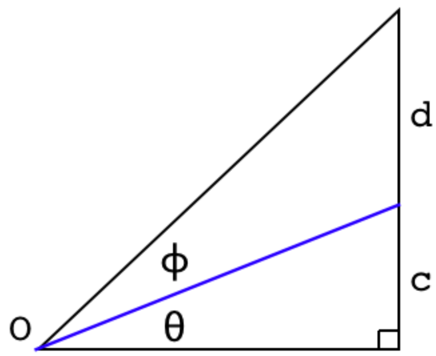
With that background, we now consider a classic problem: involving angle bisectors. Actually, before we do that, let's just show a method for constructing an angle bisector




To bisect angle $\angle BAC$ use the compass to mark off equal segments $AD$ and $AE$ and then mark off equal segments $DF$ and $EF$. The line segment $AF$ bisects the angle.

Proof: $\triangle ADF$ is congruent to $\triangle AEF$ by SSS. Therefore, $\angle CAF = \angle BAF$.

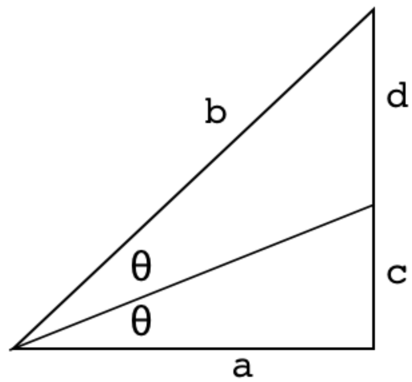
Now, back to our problem, and the diagram below.




Suppose we are given that the large triangle, and the bottom of the two smaller triangles are both right triangles.

We draw a line joining the vertex $O$ on the left with the side opposite. This line could in general be drawn anywhere, however two interesting cases are when the angle at $O$ is bisected, or when the side opposite is bisected.

These cases are different.



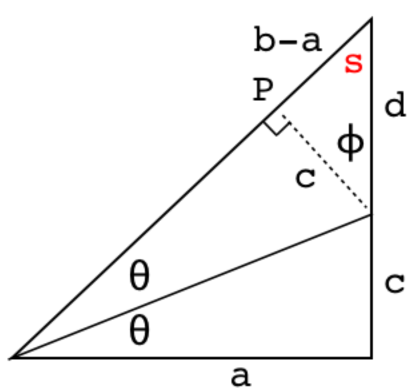

Here we have chosen the first possibility. We are in a position to prove an important theorem.

**angle bisector theorem**

With reference to the figure above, we are to prove that

$$\frac{d}{b} = \frac{c}{a}$$

Draw an altitude for the upper of the two small triangles, meeting the side of length $b$ at point $P$.



By congruent triangles (the two triangles each with vertex angle $\theta$), the altitude has length $c$.

By the rules for complementary angles discussed above:

$$2\theta + s = 90 = s + \phi$$

Hence, $2\theta = \phi$. We conclude that the smallest triangle at the top right of the figure is similar to the original. By similar triangles, we form the equal ratios of the hypotenuse to the adjacent angle (either $\phi$ or $2\theta$):

$$\frac{d}{c} = \frac{b}{a}$$

This is rearranged simply to give

$$\frac{d}{b} = \frac{c}{a}$$

which is what we were asked to prove.

$\square$

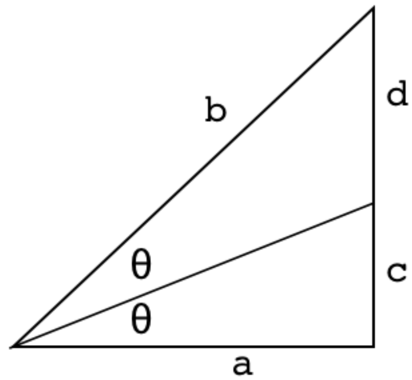
The result can be pushed a little further:

$$\frac{a}{b} = \frac{c}{d}$$

Here's the key point

$$\frac{a+b}{b} = \frac{c+d}{d}$$

$$\frac{a+b}{c+d} = \frac{b}{d} = \frac{a}{c}$$

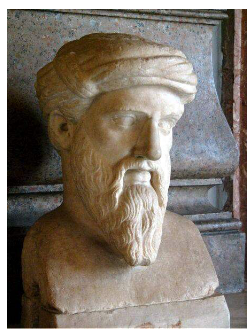



which is a surprising result and becomes important in looking at Archimedes method for approximating the value of $\pi$.

# Chapter 12

# Pythagoras

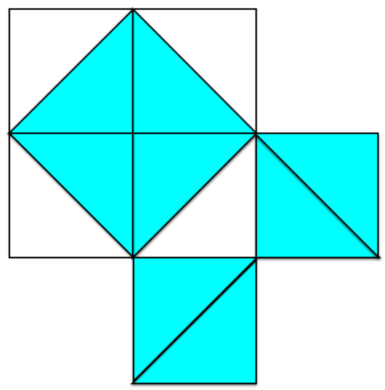
The most famous theorem of Greek geometry is also the most useful in Calculus.



Pythagoras (c.570-c.495 BC) was much younger than Thales but may have encountered him as a young man. Like many other Greek mathematicians, Pythagoras was not from the mainland, but from one of the islands, in his case, Samos, which is not far from Miletus, where Thales lived.

Pythagoras was famous as a philosopher as well as a mathematician. In fact, he founded a famous "school" and it is not sure now which of the theorems developed by this school are due to Pythagoras, and which to his disciples. It is not even clear whether the Pythagorean theorem, as we know it, was known to Pythagoras.

However, it's pretty certain that they knew something. The $3, 4, 5$ right triangle and many other Pythagorean triples (see below) had been known for a thousand years (since 1500 BC). Here is a special case, easily proved, for an isosceles right triangle.

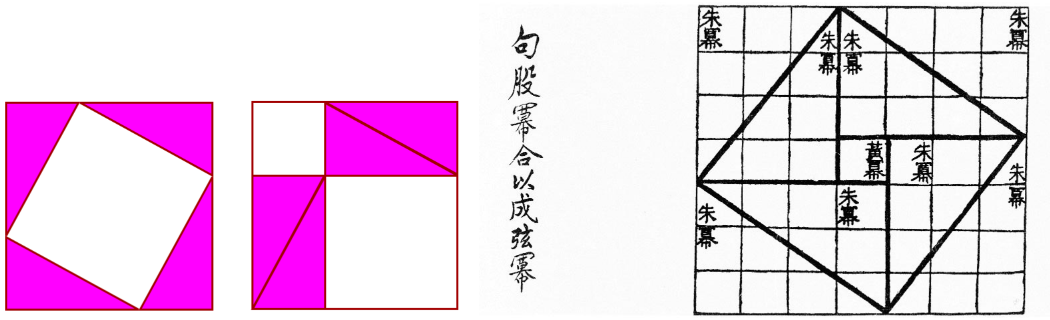


The area of the square on the hypotenuse is equal to twice the area on each side.

There are literally hundreds of proofs of the general theorem, that if $a$ and $b$ are the shorter sides of a right triangle and $c$ is the hypotenuse, then

$$a^2 + b^2 = c^2$$

This one is sometimes called the "Chinese proof." I can easily imagine proceeding from the figure above to this one by simply rotating the inner square.




It really needs no explanation, but ..

In the left panel we have a large square box that contains within it a white square, whose side is also the hypotenuse of the four identical right triangles contained inside. Altogether the four triangles plus the white area add up to the total.

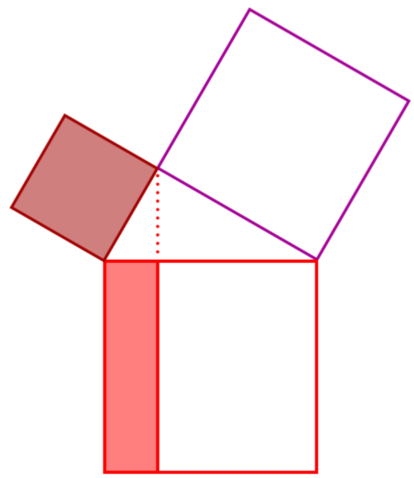
We simply rearrange the triangles. Now we evidently have the same area left over from the four triangles, because they still have the same area and the surrounding box has not changed.

But clearly, now the white area is the sum of the squares on the second and third sides of the triangles. Hence the two white squares on the right are equal in area to the large white square on the left. □

This proof is contained in the Chinese text Zhoubi Suanjing (right panel, above).

https://en.wikipedia.org/wiki/Zhoubi_Suanjing

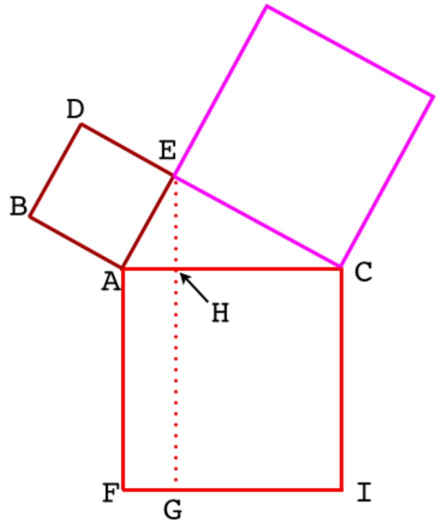
**Euclid's proof**



My favorite proof relies on the construction above (Euclid $I.47$, sometimes called the "bridal chair" or the "windmill"), where the central triangle is a right triangle, and the other constructions are squares. It is a bit more detailed, but it is one of only a few places in the book that we actually show a proof from Euclid, which is a justification for including it.

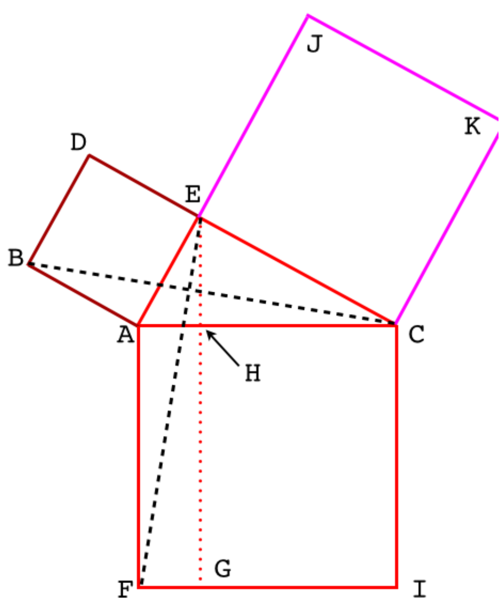
What we will show is that the part of the large square in red is equal in area to the entire small square, in maroon.

We label some points as shown:




First, drop a vertical line $EHG$, constructing the rectangle $AFGH$.

Finally, sketch dotted lines for the long sides of two triangles:




The crucial point is this: we will show that triangle $\Delta ABC$ is congruent to triangle $\Delta AEF$.

Use "side-angle-side". The two sets of sides are evidently equal

$$AB = AE, \quad AC = AF$$

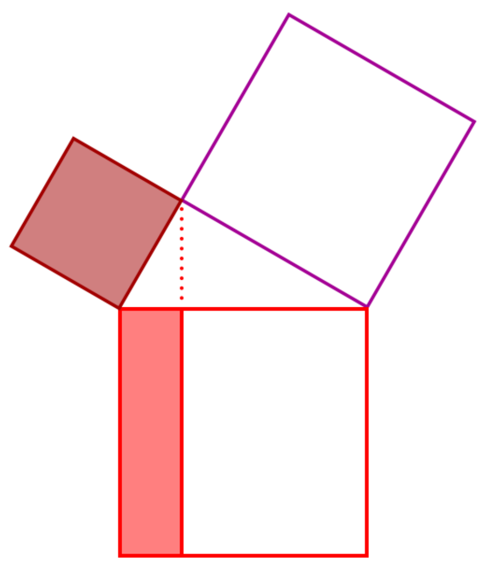
because these are given as sides of two squares.

What about the included angle? Both angle $\angle BAC$ and $\angle EAF$ contain a right angle plus the shared angle $\angle EAC$. So they are themselves equal, and thus we have proved the congruence relationship:

$$\Delta ABC = \Delta AEF$$

The next part of the proof is to tilt triangle $\Delta ABC$ to the left and see that it has base $AB$ and altitude $AE$ so its area is one-half that of the small square $ABDE$. On the other hand triangle $\Delta AEF$ has base $AF$ and altitude $AH$ (as well as $FG$) so its area is one-half that of the rectangle $AFGH$.

Hence we have proved that the two colored areas in this figure are equal:

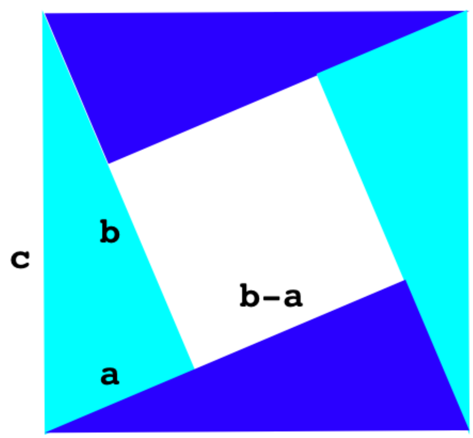


Finally, we could proceed to do the same thing on the right side of the figure, but we just appeal to symmetry. All the equivalent relationships will hold. □

**algebraic proofs**

The following proofs are algebraic ones. Not so pretty, but fast.

Arrange 4 identical right triangles as shown in the figure below. The four triangles plus a small central square form a larger quadrilateral which is also a square.




The angles at the corners of the quadrilateral, at the points flanking the hypotenuse $c$, are right angles, because they are formed by addition of two complementary angles of congruent right triangles. Since the quadrilateral has four internal right angles and equal length sides, it is a square.

Now just calculate the area of the parts. We have four identical right triangles with sides $a$ and $b$, plus the central square with sides $b - a$. The area is

$$A = 4 \cdot \frac{1}{2}ab + (b - a)^2$$

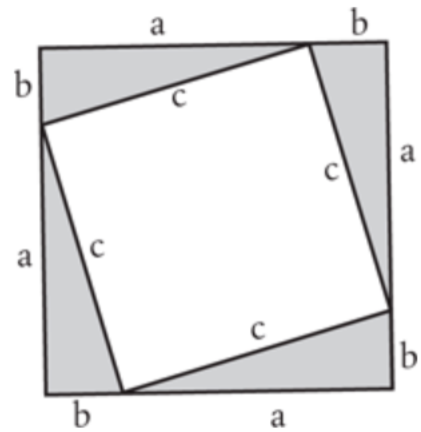
$$= b^2 + a^2$$

But the area is also the square of side $c$.

$\square$

We have used various properties proved earlier, e.g. that the sum of the angles of any triangle is 180 degrees.

Here is a very similar proof:

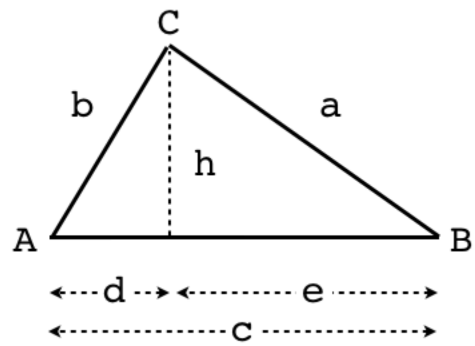



In this figure, the right triangles are aligned so that the big square has sides which combine the lengths $a + b$ and have area $(a + b)^2$. But we can also calculate the area as the sum of its components, namely, central tilted square plus the four triangles:

$$(a + b)^2 = c^2 + 4 \cdot \frac{ab}{2}$$
$$a^2 + b^2 + 2ab = c^2 + 2ab$$
$$a^2 + b^2 = c^2$$

$\square$

For the third algebraic proof, divide a right triangle into two smaller ones by dropping an altitude, which meets the base at a right angle.



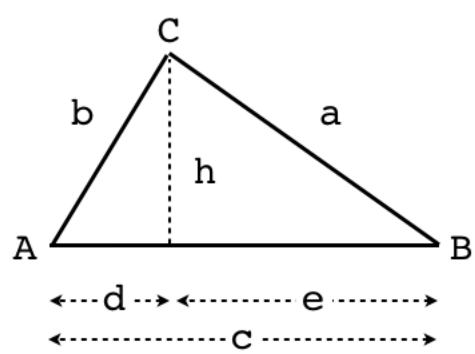

By complementary angles, these three triangles are all similar (e.g., the angle between sides $b$ and $h$ is equal to that between sides $a$ and $c$). So we can construct ratios of sides that are equal.

We need a relationship involving $a^2$.

$$\frac{h}{b} = \frac{e}{a} = \frac{a}{c}$$

We choose the ones involving $c$ and $e$:

$$a^2 = ce$$




and also a relationship involving $b^2$:

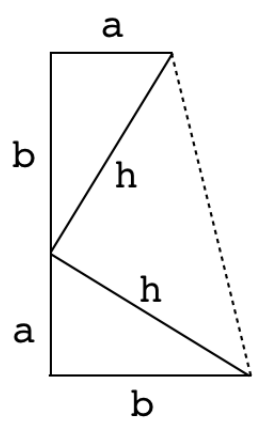
$$\frac{b}{d} = \frac{c}{b}$$
$$b^2 = cd$$

Putting the two together:

$$a^2 + b^2 = ce + cd$$
$$= c(d + e) = c^2$$

Which is what we wanted to prove. $\square$

There are more than 300 proofs of this theorem, including one by a President of the United States, James A. Garfield.




Here is his proof:

Draw a right triangle and a rotated copy as shown. The area of the quadrilateral is the product of the side $(a + b)$ and the *average* of $a$ and $b$:

$$A = (a + b) \cdot \frac{1}{2}(a + b)$$

If you don't see this right away, calculate the area as a rectangle plus a triangle (whose side is not shown but drops vertically from the top right corner).

$$a(a+b)+\frac{1}{2}(b-a)(a+b)$$

with the same result so

$$A=\frac{1}{2}a^2+ab+\frac{1}{2}b^2$$

But it is also the sum of

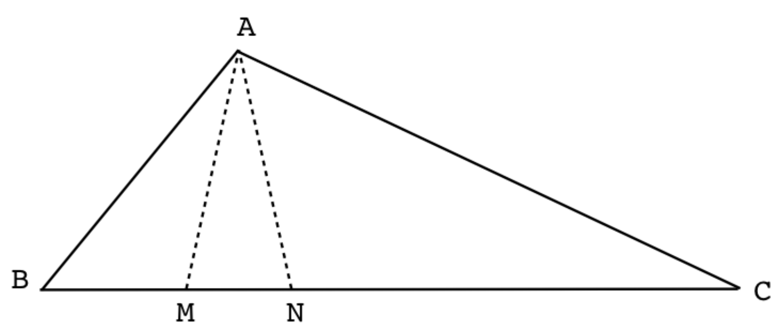
$$\frac{1}{2}ab+\frac{1}{2}ab+\frac{1}{2}h^2$$

Equate the two and the result follows almost immediately.

□

**Corollary**

There are several important corollaries of the Pythagorean theorem. We'll derive one later called the law of cosines. Here is another from the Islamic geometer Ibn Quorra, who brought algebraic techniques, shunned by the Greeks, to geometry.




Let $\triangle ABC$ be *any* triangle (here it is obtuse). Draw $AM$ and $AN$ so that the new angles $\angle AMB$ and $\angle ANC$ are equal to $\angle A$. The corresponding triangles are similar to the original, because they share the angle of measure $A$ plus one other from the original triangle.

Then

$$BM : AB = AB : BC$$

Thus, $AB^2 = BM \times BC$. Similarly

$$NC : AC = AC : BC$$

So $AC^2 = NC \times BC$ Therefore

$$\begin{aligned} AB^2 + AC^2 &= BM \times BC + NC \times BC \\ &= (BM + NC) \times BC \end{aligned}$$

In the case where the angle at vertex $A$ is a right angle, then $M$ coincides with $N$, and $BM + NC = AC$, and this reduces to the Pythagorean theorem.

### Pythagorean triples

The simplest right triangle with integer sides is $3, 4, 5$:

$$3^2 + 4^2 = 5^2$$

any multiple $n$ will work

$$(3n)^2 + (4n)^2 = (5n)^2$$

but that's not so interesting. The triples which are not multiples of another triple are called *primitive*.

To go further, we can use Euclid's formula. For every integer $m, n$, with $m > n$, a Pythagorean triple is given by

$$a = m^2 - n^2 \quad b = 2mn \quad c = m^2 + n^2$$

It is better to choose $m$ and $n$ of opposite parity (one even and one odd). Otherwise, $a$, $b$ and $c$ will all be even and the triple won't be primitive.

It is easy to see why this works:

$$a^2 + b^2 = (m^2 - n^2)^2 + (2mn)^2$$
$$= m^4 - 2m^2n^2 + n^4 + 4m^2n^2$$
$$= m^4 + 2m^2n^2 + n^4$$
$$= (m^2 + n^2)^2 = c^2$$

Suppose $a = 5$. The two squares with a difference of 5 are

$$3^2 - 2^2 = 5$$

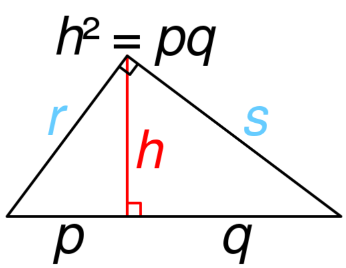
So $b = 2mn = 12$ and $c = 3^2 + 2^2 = 13$. And indeed $5^2 + 12^2 = 13^2$.

https://en.wikipedia.org/wiki/Pythagorean_triple#Enumeration_of_primitive_Pythagorean_triples

A thousand years before Pythagoras, the Babylonians knew the triple $4601, 4800, 6649$. It seems unlikely that they found this by random search.

**geometric mean**

As a slight detour from calculus, but on the topic of this chapter




According to the figure, the altitude of a right triangle is related to the two segments along the base by

$$h^2 = pq$$

$$h = \sqrt{pq}$$

That is, $h$ is the geometric mean of these two values $p$ and $q$.

The proof is simple. Using the Pythagorean theorem with the two small triangles (also right triangles), we obtain:

$$h^2 + p^2 = r^2$$

$$h^2 + q^2 = s^2$$

Summing

$$2h^2 + p^2 + q^2 = r^2 + s^2$$

Using the theorem with the big triangle:

$$r^2 + s^2 = (p + q)^2$$

$$= p^2 + 2pq + q^2$$

Equating the two expressions for $r^2 + s^2$ we get:

$$2h^2 + p^2 + q^2 = p^2 + 2pq + q^2$$

$$h^2 = pq$$

According to wikipedia:

https://en.wikipedia.org/wiki/Geometric_mean

The fundamental property of the geometric mean is that (letting $m$ be the *geometric mean* here):

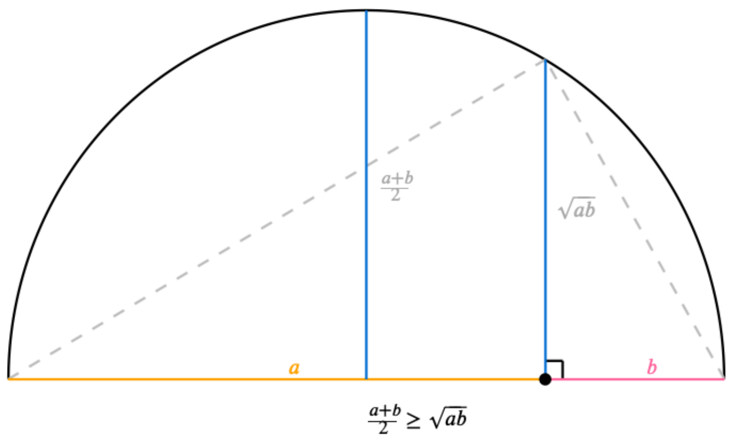
$$m \; [ \; \frac{x_i}{y_i} \; ] \; = \frac{m(x_i)}{m(y_i)}$$

and one consequence is that

> This makes the geometric mean the only correct mean when averaging normalized results; that is, results that are presented as ratios to reference values.

A number of examples are given in the article.

Last: a proof-without-words that the geometric mean is always less than or equal to the arithmetic mean.



# Part IV

# Circles

# Chapter 13

# Circles

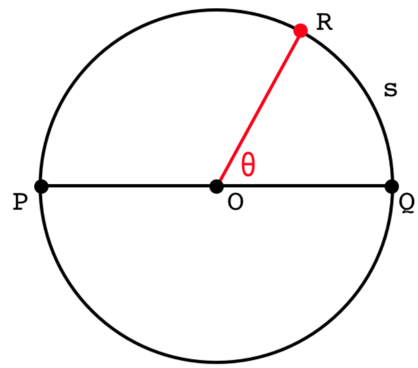
From a previous chapter, Euclid's third postulate was:

◦ Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center. The tool to do this is called a compass:

https://en.wikipedia.org/wiki/Compass_(drawing_tool)

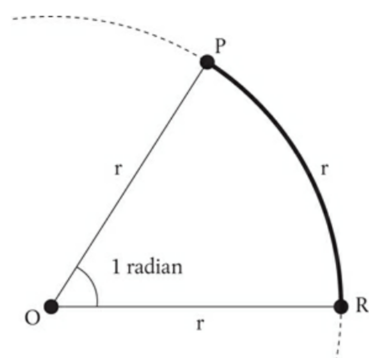
If the radius is extended so that it cuts the circle at two points, it is called a diameter. We saw previously that one can construct a line perpendicular to any given line. If that line is constructed perpendicular to the diameter at the point where it meets the circle, the new line is called a tangent line. By definition, the tangent line touches the circle at a single point.

**arcs of a circle**




In calculus and analytical geometry angles are defined in terms of radians of arc. For a unit circle with radius = 1, the total circumference is $2\pi$, so the arc swept out by the angle $\theta$ is in the same ratio to $2\pi$ as the ratio of the angle's measure in degrees to 360°.

It seems natural then to adopt the arc length as a measure of the angle, where 360° is equal to $2\pi$ *radians*, and an angle of 90°, for example, a right angle, is equal to $\pi/2$ radians.




**72.** Definition of a radian.

We say that the angle $\theta$ is equal to the arc it sweeps out on the circumference, in radians.

$$s = \theta$$

To convert some more measures of angles in degrees to radians:

$$180^\circ = \pi,\ 90^\circ = \frac{\pi}{2}$$

$$60^\circ = \frac{\pi}{3},\ 45^\circ = \frac{\pi}{4},\ 30^\circ = \frac{\pi}{6}$$

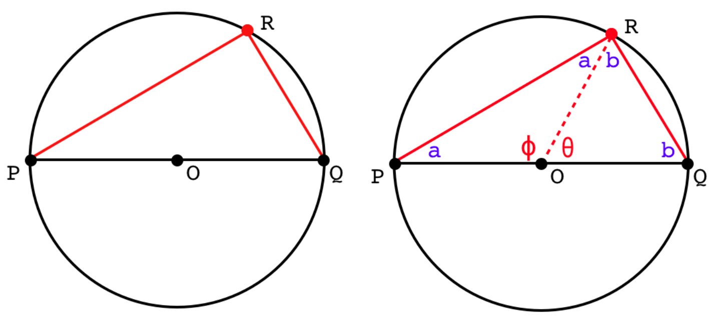
## Thales' theorem

In this chapter, we introduce a few more theorems concerning circles, starting with the last of Thales' theorems:

◦ Any angle inscribed in a semicircle is a right angle.

Now, think of three points on the circumference of the circle as forming a triangle. If two points are on a diameter of the circle, the angle formed at any arbitrary but distinct third point is always a right angle.

To prove: $\angle PRQ$ is a right angle.




Solution: Draw the radius OR. Notice that $\triangle OPR$ and $\triangle OQR$ are both isosceles.

Label the respective base angles $a$ and $b$. By considering that together the sum of the angles of $\triangle PQR$ can be written:

$$2a + 2b = \pi$$

$$a + b = \frac{\pi}{2}$$

But this is the measure of $\angle PRQ$.

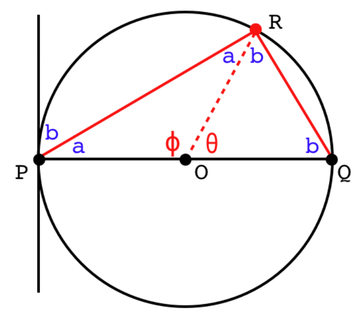
In addition, the arcs swept out by angles $a$ and $b$ (OPR and OQR on the diameter) clearly add up to $\pi$. This suggests that:

$$a = \frac{\theta}{2}$$
$$b = \frac{\phi}{2}$$

Proof:

$$2a + 2b = \pi = 2a + \phi$$
$$\phi = 2b$$

Consider the chord PR and draw the tangent at P.

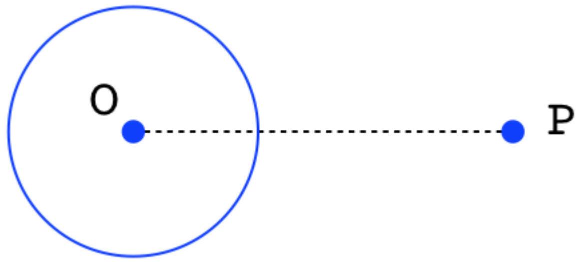



The arc between the tangent and the chord equals 2b because it is the same arc as cut off by $\angle\ PQR$ (which is $\angle\ b$).

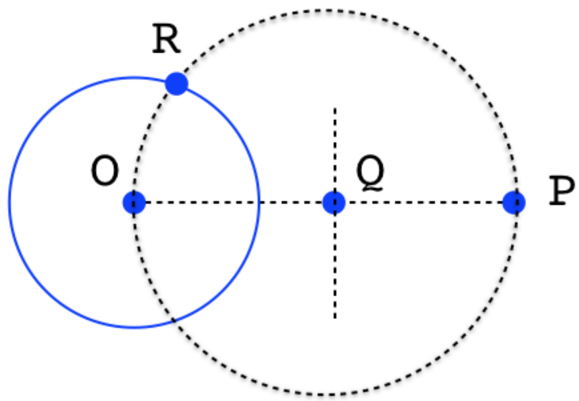
Take a chord of the circle, draw the diameter and the tangent. The same rule applies to both angles: one between the chord and the diameter, and the second between the chord and the tangent. The arc is twice the measure of the angle.

**tangents**

Thales theorem provides a way to construct the tangent to a circle passing through any exterior point $P$.

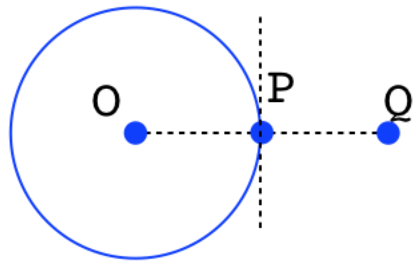



Use OP as the diameter of a circle. Draw the line segment $OP$ and divide it in half by erecting the perpendicular bisector. Use that distance as the radius of a new circle. The point $R$ is the intersection of the two circles.




By Thales theorem, $\angle ORQ$ is a right angle, and since $OR$ is a radius of the original circle, $RQ$ is the tangent at the point $R$.

To construct a tangent on a circle at a given point $P$



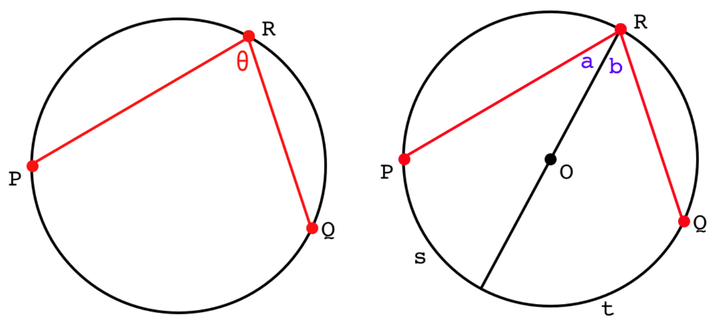

Extend $OP$ to $Q$ such that $OP$ is equal to $PQ$. Construct the perpendicular bisector at $P$. That is the tangent of the circle.

# Chapter 14

# Arcs of a circle

Having established some basic facts about circles we can do a bit more. We will use some of these results later on.

One is to generalize the result for all arcs. The examples so far contain the diameter in some way. Consider the arc swept out by the angle $\theta$ in this figure.



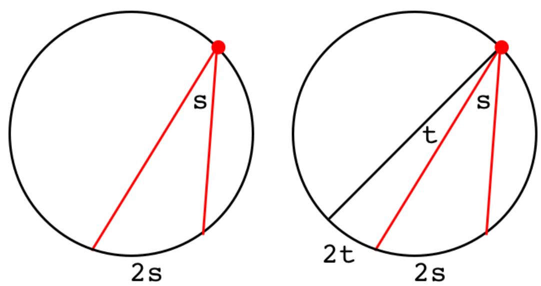

We can prove that the measure of the angle $\theta$ is equal to the 1/2 the arc swept out between P and Q. For a simple proof, draw the diameter: By our previous work:

$$b = \frac{t}{2}$$

$$a = \frac{s}{2}$$

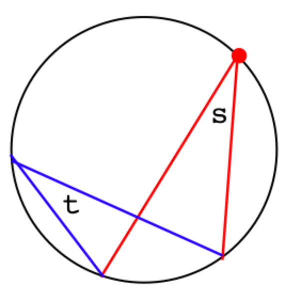
$$\theta = a + b = \frac{s+t}{2}$$

We have proved the theorem for two cases: where the diameter is one line segment flanking the angle, and where the angle includes the diameter. However, the theorem is true even if the angle does not include the diameter.




On the right, draw the diameter. Notice that we have two arcs which include the diameter: one with angle $t$ and one with angle $s+t$. We obtain the generalized arc with angle $s$ by subtraction.

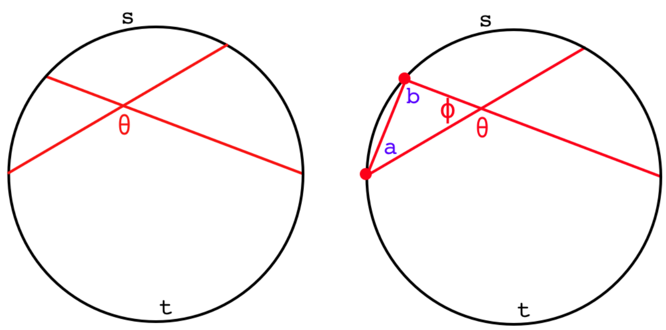
As a corollary, any two angles with vertexes on the circle that cut off the same arc are equal. In the figure, $s = t$. Also the triangles are similar triangles.



## Intersecting chords

Given two chords, to prove:

$$\theta = 1/2(s+t)$$

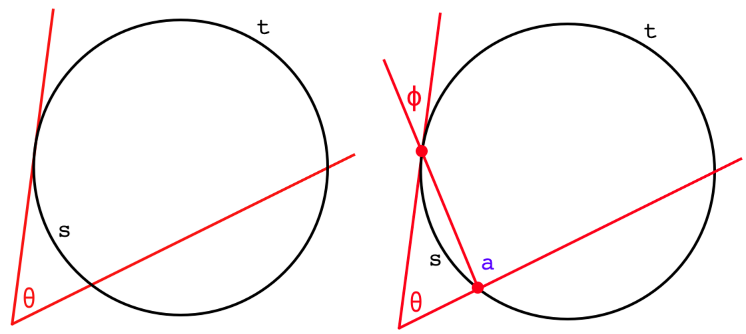



$\theta$ is the average of the two arc lengths. Solution: Draw a triangle.

$$a = \frac{s}{2}$$
$$b = \frac{t}{2}$$
$$a + b = \theta = \frac{s+t}{2}$$

## Tangent and secant

Rather than having all three points on the circle, one is now outside. We have the same arc swept out by the endpoints ($t$), but the included angle is now smaller, and there is a new small piece of arc length $s$.




To prove:

$$\theta = \frac{t-s}{2}$$

Solution: Draw the triangle. By our previous work (and supplementary angles):

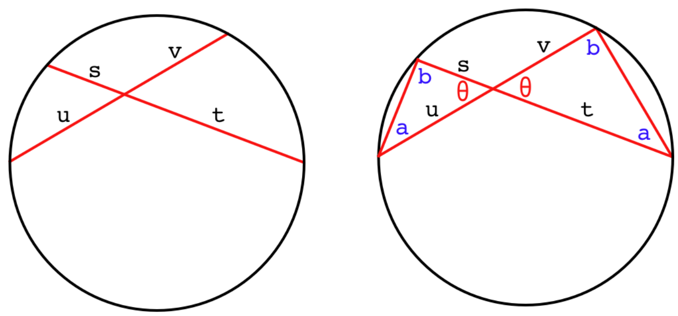
$$\phi = \frac{s}{2}$$
$$a = \frac{t}{2}$$

by supplementary angles:

$$\theta + \phi = a$$
$$\theta = \frac{t}{2} - \frac{s}{2}$$
$$= \frac{t-s}{2}$$

## Chord segments

Finally, there is a simple algebraic relationship between chord segments. Draw two chords of the circle and label the lengths of the segments as shown (note: $s$ and $t$ do not refer to arcs any more).




Draw the two triangles. Notice that the two angles labeled $a$ are equal because they sweep out the same arc of the circle, and similarly for the two angles labeled $b$. By similar triangles:

$$s/u = v/t$$

$$st = uv$$

# Chapter 15

# Eratosthenes

This part of the book is focused on geometry, and we take a look at Eratosthenes in this chapter as an important Greek scholar.

The widely held theory, that the ancient world believed the earth to be flat, is just wrong. People with any level of sophistication not only knew the earth is roughly spherical but also knew its size within a few percent of the true value.

One likely basis is the false story that Columbus had trouble getting financing for his proposed trip to China because everyone thought he would fall off the edge of the earth. This was a tall tale invented by Washington Irving, who also made up several remarkable fables about George Washington.

The real reason the Italians and the Portuguese thought Columbus would fail is that they had a pretty good idea of the size of the spherical earth and thus of the distance to China, while the over-optimistic Columbus believed it was about half the true value. The prospective financiers knew that he was not able to carry the supplies necessary for a trip of this length.

Morris Kline (*Mathematics and the Physical World*) says that the error is due to geographers after Eratosthenes, who reduced the estimated circumference from 24,000 to 17,000 miles.

**Eratosthenes**

Views of the Greek philosophers on the earth and its sphericity are detailed here

https://www.iep.utm.edu/thales/#SH8d

Here is a partial quotation:

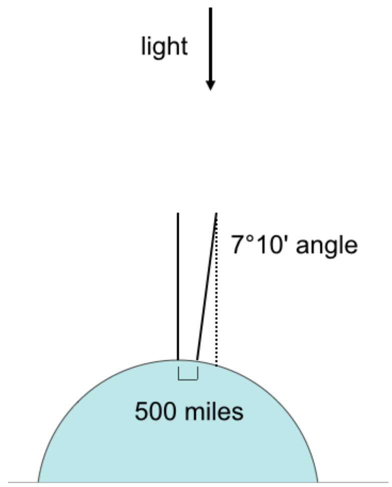
> There are several good reasons to accept that Thales envisaged the earth as spherical. Aristotle used these arguments to support his own view [...] . First is the fact that during a solar eclipse, the shadow caused by the interposition of the earth between the sun and the moon is always convex; therefore the earth must be spherical. In other words, if the earth were a flat disk, the shadow cast during an eclipse would be elliptical. Second, Thales, who is acknowledged as an observer of the heavens, would have observed that stars which are visible in a certain locality may not be visible further to the north or south, a phenomen[on] which could be explained within the understanding of a spherical earth.

https://en.wikipedia.org/wiki/Eratosthenes

Eratosthenes (ca. 276 - 195 BCE) measured the circumference of the earth from this observation: at high noon on June 21st there was no shadow seen at Syene, e.g., allegedly from a stick in the ground. Some people say it was a deep well, where the bottom was illuminated at midday.

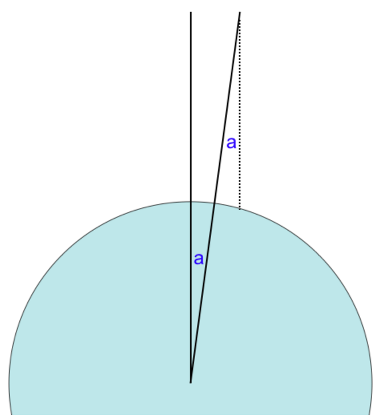
Syene is presently known as Aswan. It is on the Nile about 150 miles upstream of Luxor, which includes the famous site called the Valley of the Kings. At 24.1 degrees north latitude, Aswan or Syene is close enough to having the sun directly overhead on June 21. (The "Tropic of Cancer" is at 23 degrees, 26 minutes north).


This news about the lack of a shadow at Syene reached Alexandria, a famous center of learning of the ancient world. Alexandria lies on the Mediterranean some 500 miles north of Syene, and anyone there who was looking could observe that at high noon on June 21st there *was a shadow*. This shadow Eratosthenes measured to be some 7 degrees and a bit (7 degrees and 10 minutes).




A full 360 degrees divided by 7 degrees and a bit is approximately 50. So we can calculate on this basis that the circumference of the earth is about $50 \times 500 = 25000$ miles. That's pretty close to the correct value.

For this calculation, we assume that the sun's rays are effectively parallel (not a bad assumption given a distance of 93 million miles). Then we just use this:




an application of the alternate-interior-angles theorem.

It is curious how the distance from Alexandria to Syene was calculated [Kline]. "Camel trains, which usually traveled 100 stadia a day, took

50 days to reach Syene. Hence the distance was 5000 stadia...It is believed that a stadium was 157 meters." We obtain

$$157 \times 5000 \times 50 = 39,250 \text{ km}$$

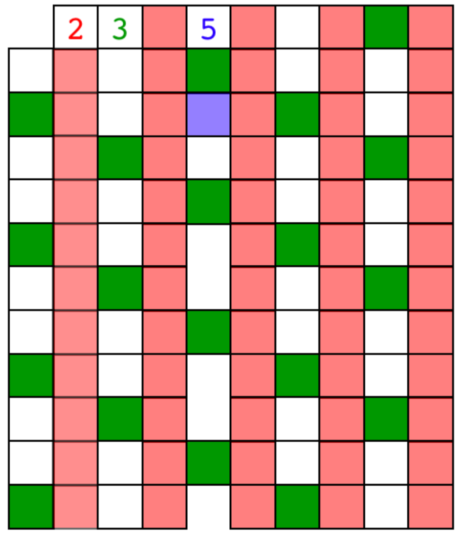
That's a much better estimate than a method that relies on camels really deserves.

### The sieve of Eratosthenes

Eratosthenes is famous in mathematics for his "sieve" which allows one to compute the prime numbers in an economical fashion. This next section really has nothing to do with calculus, but it is inspired mathematics.

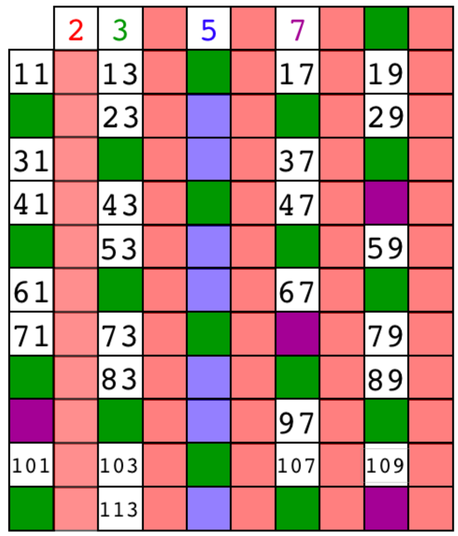
The sieve is operated by first enumerating all the integers to some upper limit (here 120). To do things manually it is convenient to use rows with 10 values, so there are 12 of them in all. Most of the boxes have not yet been numbered.

Starting with the first prime number, 2 (red), eliminate all the numbers divisible by 2 (all the even numbers). Here this has been done by coloring red all of the squares in the even numbered columns (all numbers ending in $2, 4, 6, 8, 0$).




Next, do the same thing with 3 (green). 6 was already eliminated previously, but all odd multiples of 3 like 9 and 15 go away at this step.

The next larger number that still has a white square is 5. The only squares eliminated are the white ones in the fifth row. The first value specifically eliminated at the 5 step is 25. Continue with 7, eliminating $49, 77, 91$ and 119.




The sieve ends when the number for the beginning of the next round, the smallest number not yet eliminated, is greater than the square root of the upper limit (here $\sqrt{120}$). So 7 is used for the last round, because after that round the smallest remaining integer is 11, but we terminate since $11^2 > 120$.

The graphic shows all the numbers which have yet to be eliminated after the round of 7. All of these numbers, 11, 13, 17, and so on, as well as those used as divisors for each round of the sieve $(2, 3, 5, 7)$, are prime numbers.

By testing for division by $2, 3, 5$ and 7, we have found the first 30 prime numbers.

From a performance standpoint, it is important that we do not need to carry out division. All that is really needed is repeated addition. Coding this algorithm in, say, Python is a good challenge. A bigger challenge is to come up with a method to *grow* the list of primes on demand. This can be done by keeping track of the first value to be tested above the limit, for each prime in the current list.

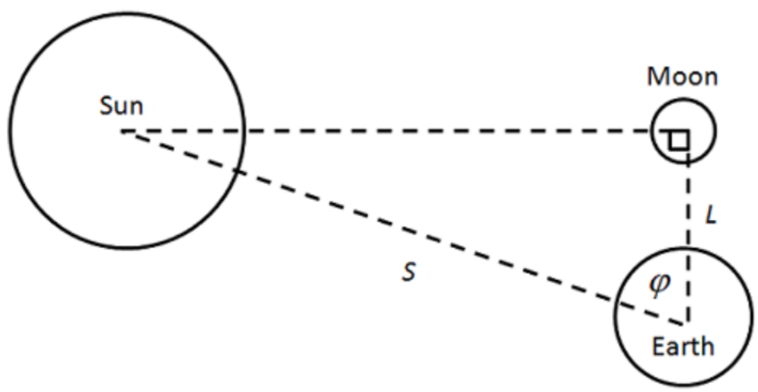
## Aristarchus

Aristarchus of Samos (310-230 BCE) wrote a famous book in which he calculated the relative sizes of the sun and the moon and their distances from earth.

One straightforward observation is that the apparent size of the sun and moon in the sky is about the same. This can be seen during a solar eclipse, or observed at any other time by holding a disk up at a fixed distance from the eye, (while taking care to block most of the sun's rays). The value is approximately one-half degree.

Since the distance to the sun is much greater than that to the moon (see below), we can infer that the sun is much larger than the moon.

The central idea of Aristarchus is that, at half moon, the geometry of the three orbs is like this:




In other words, when the phase is half moon and that moon is exactly overhead, the sun has not yet set, but is a bit above the horizon.

If $S$ is the distance to the sun and $L$ is that to the moon, he estimated that

$$18 < \frac{S}{L} < 20$$

with the same ratio for their sizes. Unfortunately, this is not a particularly good estimate. The true value is about 390. Aristarchus obtained a value of 20 for the Earth-Moon distance in Earth radii. The correct value is about 60. Much better estimates were obtained later, by Hipparchus and Ptolemy.

However, Aristarchus made up for this by being the first person to propose a heliocentric theory of the solar system: that the earth and planets rotate around the sun.

https://en.wikipedia.org/wiki/On_the_Sizes_and_Distances_(Aristarchus)

**quick estimate**

Here is an estimate for the earth-moon distance based on a lunar eclipse.

One measures the time it takes for a complete, total eclipse. From the first shadow of the earth on the moon to the last, that time is about 3 hr. The moon has moved approximately 1 earth diameter in its orbit in that time.

However, we must correct for the fact that the first and last shadows occur on opposite edges of the moon. It was noted that the shape of the eclipse suggests the earth's diameter (at that distance) is about 2.5 moon diameters. So the moon has actually moved (2.5 + 1.0)/2.5 = 1.4 earth diameters in the given time. The relevant time becomes 2.14 hr.

Any correction for the true size of the earth's diameter is minimal because the earth-moon system is so far from the source of illumination.

The other piece of information we need is the time for a full revolution, one lunar cycle. This part is tricky. Naively, you'd look for the moon to be in the same place against the fixed stars (27 days, c. 8 hr). This is off because the earth has moved in the meantime — there is a parallax error. As a rough correction, mutliply by 360/330 degrees. The result in hours is 715.

The circumference of the orbit is then

$$715/2.143 = 333$$

earth diameters.

This gives a radius of 53 earth diameters, which is not too far from 60.

# Chapter 16

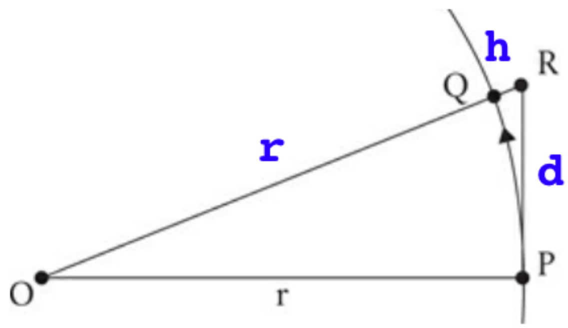
# Circular orbits

### Pythagoras and Newton

A previous chapter looked in detail at Pythagoras' Theorem, which is used incessantly from here on out. Here, we explore one use of the Pythagorean theorem and provide a taste of orbital mechanics, which is a particular focus of calculus. Newton made early calculations similar to these, which increased confidence about his famous inverse-square law and inspired the mathematics that led to the explanation of elliptical orbits.

Although the orbits of the planets around the sun are ellipses, they are very nearly circular and we will make that approximation for what follows here.

We use the Pythagorean Theorem to make another approximation. Using $r$ for the (fixed) radius of the orbit for the moment, because the construction has capital letters for the points, including the symbol $R$:




$$r^2 + d^2 = (r+h)^2 = r^2 + 2rh + h^2$$
$$d^2 = 2rh + h^2$$

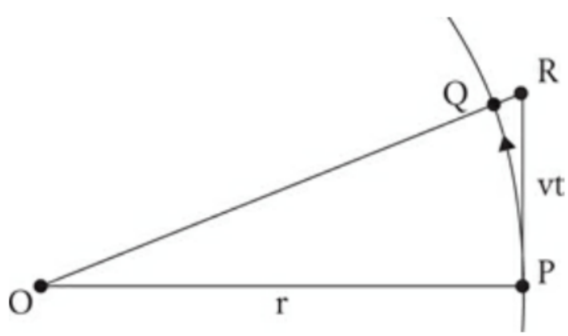
If $h << r$ then we can ignore the very small quantity $h^2$ and obtain

$$d^2 = 2rh$$
$$r = \frac{d^2}{2h}, \qquad h = \frac{d^2}{2r}$$

If the planet were not accelerated, then it would move from $P$ to $R$, a distance $d$, and this is equal to the velocity $\times$ time:

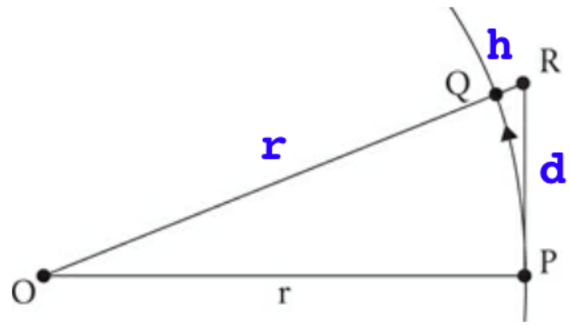
$$d = vt$$

At this point, we use an idea from calculus. *For a small enough segment of the orbit*, this distance $PR$ is the same as the arc length $PQ$.



So we substitute for $d^2 = (vt)^2$ into the equation from above

$$h = \frac{d^2}{2r} \approx \frac{(vt)^2}{2r}$$




Also, for a small enough part of the orbit (again), $h$ and $d$ are perpendicular to each other as well.

At this point we use the additional assumption that the force is directed toward the sun. We might say that the distance *fallen* by the planet in this short time is $h$.

By the standard equation of motion, under gravitational acceleration $g$ is related to $h$ and the time $t$ by this equation:

$$h = \frac{1}{2}gt^2$$

We combine the two different expressions for $h$

$$h = \frac{1}{2}gt^2 \approx \frac{(vt)^2}{2r}$$

$$g \approx \frac{v^2}{r}$$

Note: we have not covered this yet. If this idea (dependence on $t^2$) is completely new to you, you may want to come back to this part after going through the first **chapter** on calculus.

The equation $a = v^2/r$ comes even more easily with a little bit of calculus and the use of vectors. See **here**.

### Kepler's Third Law

The famous mathematician Johannes Kepler (of whom much more later also), working with observational data from Tycho Brahe, had the following values for the radius $R$ of the (assumed circular) orbit and the period $T$ (time for completion of one orbit), for five planets.

Orbital data for the six planets known in Kepler's time

| | $\overline{r}$ (units of $\overline{r}_{\text{Earth}}$) | $T$ (years) |
|---|---|---|
| Mercury | 0.387 | 0.241 |
| Venus | 0.723 | 0.615 |
| Earth | 1.000 | 1.000 |
| Mars | 1.524 | 1.881 |
| Jupiter | 5.203 | 11.862 |

On the basis of this data, Kepler published his **third law** (in 1619, about 10 years after the first two). K3 states that

$$T^2 = kR^3$$

The square of the period is proportional to the cube of the radius of the orbit. The data in the table has been scaled so that $k = 1$.

For a circular orbit, the orbital speed, the magnitude of the velocity $v = |\mathbf{v}|$, is constant.

The period times the speed is equal to the circumference.

$$vT = C = 2\pi R$$
$$T = \frac{2\pi R}{v}$$

K3 above says that

$$R^3 = T^2$$
$$= \frac{(2\pi)^2 R^2}{v^2}$$

Hence

$$v^2 \approx \frac{1}{R}$$

We showed above that the acceleration for a circular orbit is

$$a = \frac{v^2}{R} = v^2 \cdot \frac{1}{R}$$

so we conclude that that

$$g = a \approx \frac{1}{R} \cdot \frac{1}{R} = \frac{1}{R^2}$$

if the acceleration of gravity $g$ is directed toward the sun, with a magnitude that is inversely proportional to the square of the distance, then we can explain Kepler's third law by running this chain of reasoning in reverse.

**comparing the moon to an apple**

Earlier we worked out that the acceleration is

$$a = \frac{v^2}{R}$$

Let's figure out the acceleration of the moon. We make a decision to work in English units for this one.

The moon averages about 237 thousand miles from earth (221.5 - 252.7 thousand miles). The earth's circumference is about 24.9 thousand miles so its radius is about 3.96 miles. Thus, the ratio of the moon's distance to the center of the earth, compared to my distance to the center of the earth, is about 60 : 1 (ranging between 56-64).

What is the moon's velocity? The distance it travels in one complete orbit (in feet) is:

$$2\pi \cdot 2.4 \times 10^5 \cdot 5280$$

The time that takes in seconds is

$$28 \cdot 24 \cdot 3600$$

$$v = \frac{2\pi \cdot 2.4 \times 10^5 \cdot 5280}{28 \cdot 24 \cdot 3600}$$

The acceleration is $v^2/R$ so we square everything except the radius.

$$a = \frac{(2\pi)^2 \cdot 2.4 \times 10^5 \cdot 5280}{(28 \cdot 24 \cdot 3600)^2} = 0.0085$$

That's in feet per second.

We compare this value to the acceleration measured at the surface of the earth, which is 32.2 in the same units. The ratio is 3788, which is just over $(61.5)^2$.

Newton:

> I began to think of gravity extending to the orb of the Moon . . . and computed the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth . . . & found them answer pretty nearly. All this was in the

two plague years of 1665-1666. For in those days I was in the prime of my age for invention & minded mathematicks and Philosophy more than at any time since.

# Part V

# Continuum of numbers

# Chapter 17

# Fundamental theorem of arithmetic

This chapter doesn't have much to do with calculus, it is here because it is a key result of Euclid's, though he never stated the theorem in its modern form. It also shows some more sophisticated but (I hope) not too challenging proofs. It can be skipped without interfering with the development of the rest of the text.

We're going to show that every integer has a **unique prime factorization**.

$$n = p_1 \cdot p_2 \dots p_n$$

First, we need a preliminary result, which is called **Euclid's lemma**.

Every natural number $n > 1$, i.e. every positive integer greater than 1, is either prime, or it is the product of two smaller natural numbers $a$ and $b$.

But the same is true of $a$ and $b$ in turn.

Therefore, every number that can be factored into $a$ and $b$ is the product of the prime factors of $a$ times the prime factors of $b$.

Suppose a given prime $p$ divides $n = ab$, i.e. $p|n$. Then either $p|a$ or $p|b$ (or both).

## Proof of existence

The proof is by induction. (I know we haven't covered induction yet, so you can consider this a demonstration of how the method works rather than a real proof. For more information, see **here**.

Assume the lemma is true for all numbers between 1 and $n$.

It is certainly true for say, $n < 101$, because we can check each case. Start with $n = 101$.

∘ If $n$ is prime (as it is here) there is nothing to prove and we move to $n + 1$.

∘ $n$ is not prime, then there exist integers $a$ and $b$ (with $1 < a \leq b < n$) such that $n = a \times b$.

∘ By the induction hypothesis, since $a < n$ and $b < n$, $a$ has prime factors $p_1 p_2 \dots$ and $b$ has prime factors $q_1 q_2 \dots$ so

$$n = ab = p_1 p_2 \dots q_1 q_2 \dots$$

This shows that there exists a prime factorization of $n$.

## Proof of uniqueness

To show that the prime factorization is unique suppose that $n$ is the smallest integer for which there exist two different factorizations:

$$n = p_1 p_2 \cdots = q_1 q_2 \dots$$

Pick the first factor $p_1$. Since $p_1$ divides $n = q_1 q_2 \ldots$, by Euclid's lemma, it must divide some particular $q_j$. Rearrange the $q$ so that $q_j$ is first.

But since $p_1$ divides $q_1$ and both are prime, it follows that $p_1 = q_1$.

Now continue the same process with all the factors $p_i$.

As wikipedia says now:

This can be done for each of the $m$ $p_i$'s, showing that $m \leq n$ and every $p_i$ is some $q_j$. Applying the same argument with the p and q reversed shows $n \leq m$ (hence $m = n$) and every $q_j$ is a $p_i$.

$\square$

**Hardy proof**

Hardy and Wright (*Theory of Numbers*, sect. 2:11) have a second proof, which is given here (almost) verbatim:

Let us call numbers which can be factored into primes in more than one way, *abnormal*, and let $n$ be the smallest abnormal number.

Different factorization:

The same prime $P$ cannot appear in two different factorizations of $n$, for, if it did, $n/P$ would be abnormal and yet $n/P < n$, the smallest abnormal number.

Thus, we have that

$$= p_1 p_2 \cdots = q_1 q_2 \ldots$$

where the $p$ and $q$ are primes, and no $p$ is a $q$ and no $q$ is a $p$.

If there exist abnormal numbers with two such factorizations, they must be completely different.

**the contradiction**

We may take $p_1$ to be the least $p$ or $q$ (if it's $q$, switch labels). Since $n$ is composite, $p_1^2 \leq n$.

The same is true for $q_1$ and (since $p_1 \neq q_1$), we have that $p_1 q_1 < n$

Hence, if $N = n - p_1 q_1$, we have $0 < N < n$ and also that $N$ is not abnormal.

Now $p_1 | n$ and since $N = n - p_1 q_1$, so $p_1 | N$.

Similarly $q_1 | N$. Hence $p_1$ and $q_1$ both appear in the unique factorizations of both $N$ and $p_1 q_1$.

From this it follows that $p_1 q_1 | n$ and hence $q_1 = n/p_1$. But $n/p_1$ is less than $n$ and has the unique prime factorization $p_2 p_3 \ldots$.

Since $q_1$ is not a $p$, this is impossible. Hence there cannot be any abnormal numbers, and this is the fundamental theorem.

□

# Chapter 18

# Rationals

The integers are great, they give us an infinite supply of numbers.

However, there is a problem with division. For

$$p \in \mathbb{N}, \quad q \in \mathbb{Z}$$

very often the result of $p \div q$ is not contained in $\mathbb{N}$ or even in $\mathbb{Z}$. We say these sets are not *closed* under division.

For example $3 \div 2 = ?$

So, we just leave the result as

$$\frac{p}{q} = \frac{3}{2}$$

where $p/q$ is in "lowest terms", i.e. they have no common factor other than 1. Of course if $p$ is a factor of $q$ or $q$ is a factor of $p$, then we can divide both top and bottom by whichever is smaller (to yield an integer in the case where $q < p$).

$q$ must not be zero because division by zero is not defined. We *could* choose to allow division by zero, but would quickly run into logical contradictions.

**interpolation**

Now, consider two rational numbers, not equal. Let

$$s = \frac{p_1}{q_1} \quad t = \frac{p_2}{q_2}$$

Suppose $s < t$.

The *average* of these two rational numbers is:

$$r = \frac{1}{2} \; [ \; s + t \; ]$$

Then

$$2r = s + t$$

$$2r - 2s = t - s$$

We have that $s < t$, so $t - s > 0$ and then

$$r - s > 0$$

$$r > s$$

A similar argument will show that

$$r < t$$

so

$$s < r < t$$

□

Thus, one can always find a new rational number that lies between two known rational numbers.

**decimal representation**

Every rational number can be represented as a decimal, using the method called long division.

Consider 1/2

$$2\overline{)1.000}$$

We say that 2 does not *go into* 1, since $2 > 1$, so we have the first part of our result as 0, followed by a decimal point. But 2 does go into 10 exactly 5 times, giving 0.5. The remainder is zero and so the division process terminates.

Consider 1/8.

$$8\overline{)1.000}$$

∘ 8 goes into 10 once, leaving 2 as remainder

∘ 8 goes into 20 twice, leaving 4.

∘ 8 goes into 40 exactly 5 times with no remainder.

The result is 0.125.

The other possibility is that in going through the process a remainder comes up that has been seen previously. If this happens then the sequence will repeat forever.

If we don't terminate with zero, then this must eventually happen, because there are only as many as $q$ possible remainders.

Thus, for example

$$1/7 = 0.142857142857\ldots$$

which contains 142857, repeating.

**decimals to fractions**

Conversely, every repeating decimal can be represented as a rational number. For example

```
      1 x r =      0.142857142857...
1000000 x r = 142857.142857...
 999999 x r = 142857
```

$$r = \frac{142857}{999999} = \frac{1}{7}$$

since $7 \times 142857$ equals 999999 exactly.

You can do this trick with

```
     r = 0.333
10 x r = 3.33
 9 x r = 3
```

$$r = \frac{3}{9} = \frac{1}{3}$$

or even

```
     r = 0.4999
10 x r = 4.999
 9 x r = 4.5
```

$$r = \frac{4.5}{9} = \frac{1}{2}$$

and

```
     r = 0.9999
10 x r = 9.999
 9 x r = 9
```

$$r = \frac{9}{9} = 1$$

This is one of the subtleties of numbers. In what sense can we say that

$$0.5 = 0.4999\ldots$$

$$1 = 0.9999\ldots$$

Most everyone is OK with the example $1/3 = 0.3333\ldots$ but some may be uneasy with the other two.

Ultimately, we justify the result as defined by evaluation of a limit.

Consider 0.9999. If $n$ is the number of places in the result, then as $n \to \infty$ the number being shown approaches 1 as its limit. We'll come back to this after considering the real numbers.

**ordering**

For two rational numbers $a$ and $b$ there are only three cases: either $a = b$, $a < b$ or $b < a$.

$$\frac{p}{q} < \frac{s}{t} \iff pt < qs$$

$p/q$ is less than $s/t$ if and only if $pt < qs$. Ordering of the integers guarantees ordering of the rational numbers.

Note: we used the property of the integers that if

$$a < b$$

then for $c > 0$

$$ca < cb$$

**intervals**

We denote the numbers greater than $u$ and less than $v$ as lying in the interval $(u, v)$. With parentheses, the interval described is *open*, it does not include the boundary values.

To describe a *closed* interval, write $[u, v]$. This interval includes all the values in the first one, plus it also includes $u$ and $v$.

Because of the density property described below, any interval such as

$$I = [0, 1]$$

contains an *infinite* quantity of rational numbers.

**density**

Consider the set of all points

$$x = \frac{p}{10^n}$$

for all natural numbers $n$ and integers $p$.

It is clear that simply by increasing the value of $n$, we can construct a set of equally spaced rational numbers as tightly clustered as we wish.

The rational numbers are said to be *dense* on the number line.

The method for computing the average of two rational numbers could be used to achieve the same thing. The result is:

**theorem**

• Between *any* two rational numbers it is always possible to find another rational number.

We might describe this situation by writing that

$$\forall\ u, v \in \mathbb{Q}\ \exists\ w \in \mathbb{Q} \mid w \in (u, v)$$

That's a mouthful!

- the symbol $\forall$ means "for every" or "for all".
- the symbol $\exists$ means "there is" or "there exists".
- the symbol $\mid$ means "such that".

For every open interval whose bounds are rational numbers, there exists another rational number between $u$ and $v$.

# Chapter 19

# Irrationals

There is a big problem with rational numbers which you probably know: some numbers cannot be expressed as the ratio of two integers, as a first example, the number which when multiplied by itself is equal to 2, written $\sqrt{2}$.

The discovery that one cannot find integer $p$ and $q$ such that

$$(\frac{p}{q})^2 = 2$$

is due to the Pythagorean school and was most unwelcome since it screwed up their cherished theory of the universe.

Some say that they drowned the guy who discovered it by throwing him overboard, and that his name was Hippasus. Like most stories about Greek mathematicians, the truth is unknown.

We will see that there is a similar problem (called irrationality) with $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, etc.

Proof.

For $\sqrt{2}$:

We assume that there does exist a rational number $p/q$ such that

$$\frac{p}{q} = \sqrt{2}$$

We will show that this assumption leads to a contradiction.

A crucial part of the proof is that we suppose $p/q$ to be in lowest terms and in particular, that $p$ and $q$ are not both even. It would be easy to recognize the case if they were both even, for then each would have their terminal digit in the set $\{0, 2, 4, 6, 8\}$.

Another fact we will need is that every odd number, when squared, gives an odd result. Proof: every odd number can be written as $2k+1$ (for non-negative integer $k$) and then

$$(2k+1)^2 = 4k^2 + 4k + 1$$

which is an odd number. Therefore, if $n^2$ is even, $n$ is also even.

So go back to

$$\frac{p}{q} = \sqrt{2}$$

Move the $q$ term to the right-hand side and square both sides:

$$p^2 = 2q^2$$

This implies that $p^2$ and $p$ are even, using the result from above. So we can write that $p = 2m$. But now

$$(2m)^2 = 2q^2$$
$$2m^2 = q^2$$

which implies that $q$ is *also* even.

Recall that we started with the assumption that $p$ and $q$ are not both even. We have reached a contradiction. We conclude that there do not exist two integers $p$ and $q$ such that $p/q = \sqrt{2}$.

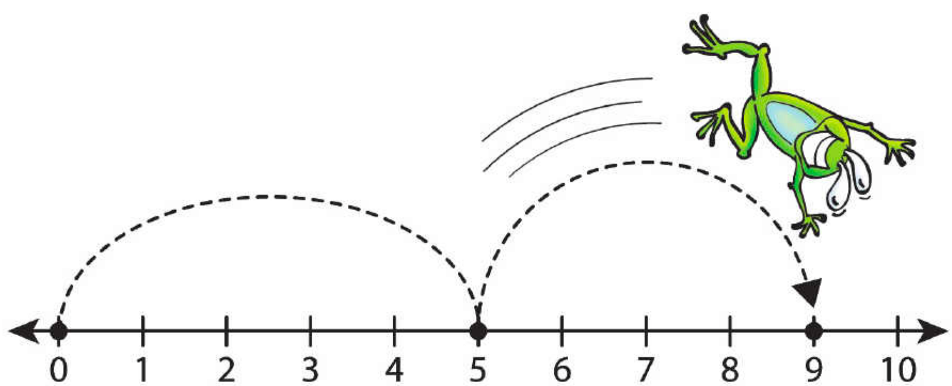
**discussion**

To quote Hardy (*A Mathematician's Apology*):

> The proof is by reductio ad absurdum, and reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers *the game*.

The numbers like $\sqrt{2}$ are said to be *irrational* numbers and the set of these, plus all the other numbers is called the set of real numbers $\mathbb{R}$.

This led Dedekind to formulate the famous Dedekind cut. Visualize the standard number line as an infinite line on (an infinite) piece of paper.




Each real number corresponds to a cut, a knife-edge coming down somewhere on this number line. Every other number that is not equal to this one, is either $>$ or $<$ the number specified by the cut.

One position is $\sqrt{2}$, one is $7/4$ and so on.

**proof using prime factors**

The **fundamental theorem of arithmetic** says that any positive integer greater than 1 can be expressed as a product of its prime factors

$$n = p_1 \cdot p_2 \dots p_i$$

where this factorization is unique (if the factors are sorted first), and multiple copies allowed. For example

$$60 = 2 \cdot 2 \cdot 3 \cdot 5$$

A corollary says that the square of any integer (a perfect square) has an even number of prime factors since

$$n^2 = p_1^2 \cdot p_2^2 \dots p_i^2$$

In the expression from above

$$p^2 = 2q^2$$

the number of prime factors on the left is therefore even, but the number on the right is odd. This is a contradiction. Therefore $p$ and $q$ cannot both be integers.

**continued fractions**

Square roots can be represented as continued fractions. Some smart person figured out that we can write this:

$$(\sqrt{2} - 1)(\sqrt{2} + 1) = 2 - 1 = 1$$

Now, rearrange to get a substitution we will use repeatedly

$$\sqrt{2} - 1 = \frac{1}{\sqrt{2} + 1}$$

Add one and subtract one on the bottom right:

$$\sqrt{2} - 1 = \frac{1}{2 + \sqrt{2} - 1}$$

And substitute for $\sqrt{2} - 1$:

$$= \frac{1}{2 + \frac{1}{\sqrt{2}+1}}$$

Lather, rinse, and repeat:

$$= \frac{1}{2 + \frac{1}{2+\sqrt{2}-1}} = \frac{1}{2 + \frac{1}{2+\frac{1}{\sqrt{2}+1}}}$$

Clearly, this goes on forever.

$$\sqrt{2} - 1 = \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \dots}}}$$

Add 1 to the value of the *continued fraction* to get an expression for the square root of 2.

The numerators are all 1, so this is called a simple continued fraction. The continued fraction representation of $\sqrt{2}$ is usually written as $[1 : 2]$, meaning that there is an initial 1 followed by repeated 2's.

This fraction goes on forever (since $\sqrt{2}$ is irrational). One can view the existence of the infinite continued fraction as a proof of irrationality.

We can turn the above into an approximate decimal representation of $\sqrt{2}$, by truncating the infinite expansion at the .... Then the last fraction is $5/2$. Invert and add, repeatedly:

```
2 + 1/2 = 5/2
2 + 2/5 = 12/5
2 + 5/12 = 29/12
2 + 12/29 = 71/29
2 + 29/71 = 171/71
2 + 71/171 = 413/171
```

To terminate we need to use that initial 1:

```
1 + 171/413 = 584/413 = 1.414043
```

To six places, $\sqrt{2} = 1.414213$. We have only three places, but can get more (convergence is relatively slow, however).

### geometric proof

There are many other proofs of the irrationality of the square root of 2.

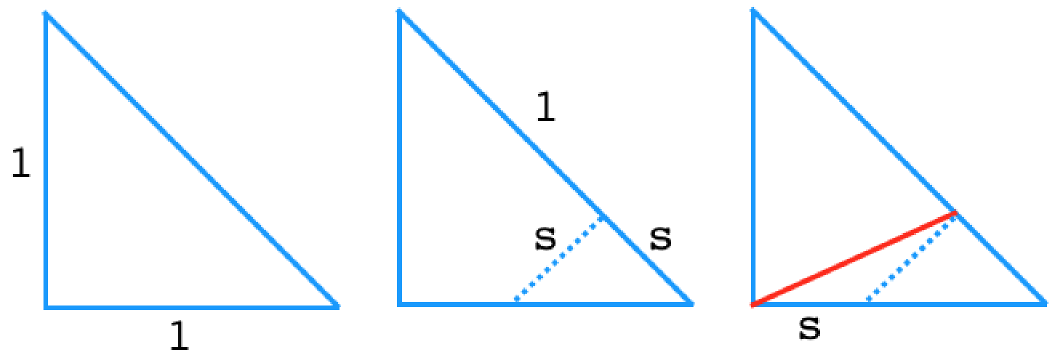
https://www.cut-the-knot.org/proofs/sq_root.shtml

Here we will look at one more, before considering a more general proof for all non-perfect squares. This one is from Tom Apostol (see the link). A more elaborate exposition is:

https://jeremykun.com/2011/08/14/the-square-root-of-2-is-irrational-geometric-proof/

Draw an isosceles triangle with side length 1, then Pythagoras tells us that the hypotenuse is equal in length to $\sqrt{2}$ (left panel).

Our hypothesis is that the length of the hypotenuse is a rational number, and that its ratio to the side is in "lowest terms".




Now mark off the length of the side on the hypotenuse and erect a perpendicular (middle panel). The new small triangle that is formed is also isosceles (it is a right triangle and it also contains one of the complementary angles of the original right triangle). By hypothesis, its side length $s$ is the difference of two rational numbers, so it is a rational number.

Furthermore, the triangle with the red base and blue sides of length 1 is isosceles (right panel), so by complementary angles the triangle with the red base and one side a dotted line has equal angles at its base and so is isosceles. All the lengths marked $s$ are equal.

Therefore, the hypotenuse of the new, small right triangle is a rational number, since it is equal to $1 - s$.

We are back where we started, with an isosceles triangle that has all rational sides.

It is clear that this process can continue forever. The sides will never be in "lowest terms" because we can always form a new similar but smaller right triangle, which amounts to evenly dividing both the sides and the hypotenuse by a rational number.

**general proof**

Every integer which is not a perfect square has a square root that is irrational.

Proof:

Let $D$ be a positive integer but not the square of an integer.

Then, there exists a positive integer $\lambda$ such that

$$\lambda^2 < D < (\lambda + 1)^2$$

Now suppose there does exist a rational number $t/u$ whose square is $D$. We will show that this leads to a contradiction.

$$\frac{t}{u} = \sqrt{D}$$

$$t^2 - Du^2 = 0$$

As previously, we can assume that $u$ is the smallest positive integer with this property since otherwise we could find $\gcd(t, u)$ and divide.

Substituting into the first expression above:

$$\lambda^2 < (\frac{t}{u})^2 < (\lambda + 1)^2$$

$$\lambda < \frac{t}{u} < \lambda + 1$$

(This last step is justified because $\lambda > 0$ and so are $D$ and $\lambda + 1$).

$$\lambda u < t < (\lambda + 1)u$$

We next consider two integers. First, define $v = t - \lambda u$. By manipulating the left-hand inequality above, we see that $v > 0$. With the right-hand inequality.

$$t < (\lambda + 1)u$$

$$t - \lambda u < u$$

Thus, $v = t - \lambda u$ is also certainly less than $u$ and we have established that

$$0 < v < u$$

We then define the integer $s = Du - \lambda t$. We will show that $Du > \lambda t$ which means that $s > 0$.

From the left-hand inequality

$$\lambda u < t$$

$$\lambda < \frac{t}{u}$$

but

$$\frac{t}{u} = D\frac{u}{t}$$

hence

$$\lambda < D\frac{u}{t}$$

which means that $Du > \lambda t$ as required.

Therefore, we have established that both $s$ and $v$ are positive integers with $v < u$.

In the next section we will prove that $s$ and $v$ have the same property as $t$ and $u$, namely

$$s^2 - Dv^2 = 0$$

This is a contradiction, since we supposed that $u$ was the smallest integer such that $s^2 - Dt^2 = 0$. Thus, there is no such $u$ and $t$ and $\sqrt{D}$ is irrational.

□

Proof of our lemma:

This is just some messy algebra. I expand a bit from the answer given in the reference, and work the algebra in reverse.

We defined $s = Du - \lambda t$ so

$$s^2 = D^2u^2 - 2Du\lambda t + \lambda^2 t^2$$

and $v = t - \lambda u$ so

$$Dv^2 = D(t^2 - 2\lambda ut + \lambda^2 u^2)$$

Subtracting, we obtain:

$$\begin{aligned} s^2 - Dv^2 &= D^2u^2 - 2Du\lambda t + \lambda^2 t^2 - Dt^2 + 2D\lambda ut - D\lambda^2 u^2 \\ &= D^2u^2 + \lambda^2 t^2 - Dt^2 - D\lambda^2 u^2 \end{aligned}$$

which factors magically into

$$= (\lambda^2 - D)(t^2 - Du^2)$$

but the second term is zero by the definition of $t$ and $u$, so the whole thing is zero, which means that

$$s^2 - Dv^2 = 0$$

as required.

□

**other irrational numbers**

There are many other irrational numbers besides these square roots. The proof that $e$ is irrational is easy, but since we haven't introduced the exponential yet we need to wait.

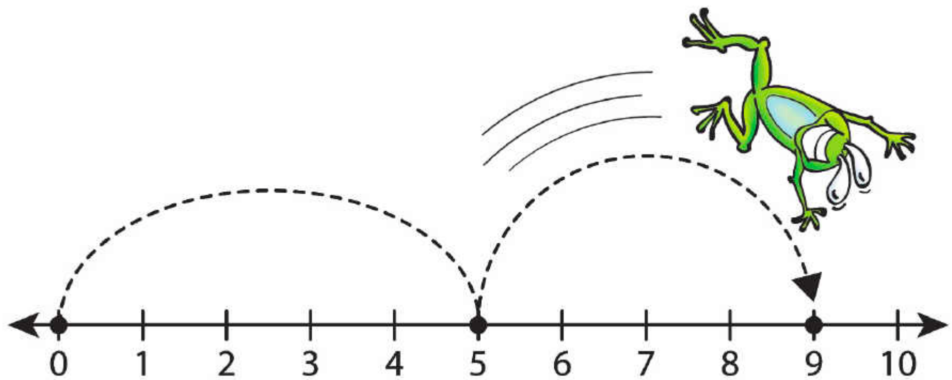
The proof that $\pi$ is irrational is a bit harder, so we defer that as well.

## density

### number line

A simple tool to visualize all of the real numbers is the familiar number line. Here is the number line with numbers marked from $\mathbb{N}$, but obviously we could also draw one for $\mathbb{Z}$ or $\mathbb{Q}$.

We explore the application of the number line to $\mathbb{R}$ as we proceed.




We might simply assume that to every point on the number line there corresponds a rational or irrational number, and that this total collection obeys the same laws of arithmetic as the rational numbers do.

As mentioned above, the need for the real numbers is indicated by empty "holes" in the number line corresponding to the irrational numbers like $\sqrt{2}$.

A problem that arises is how to specify an irrational number non-geometrically and other than as the solution to an equation such as $r^2 = 2$. In all cases we write particular real numbers as *approximations*. For example, the square root of 2 lies between 1 and 2 because

$$1^2 = 1 < 2$$

$$2^2 = 4 > 2$$

Implying that $\sqrt{2} < 2$. At the second place:

$$1.4^2 = 1.96 < 2$$

$$1.5^2 = 2.25 > 2$$

Implying that $\sqrt{2} < 1.5$. At the third:

$$1.41^2 = 1.9881 < 2$$

$$1.42^2 = 2.0164 > 2$$

Implying that $\sqrt{2} < 1.42$. and at the seventh place

$$1.414213^2 = 1.9999984093689998.. < 2$$

$$1.414214^2 = 2.0000012377960004 > 2$$

and so on.

We can never write down the decimal value of $\sqrt{2}$ exactly, but only approximate it to greater and greater precision. The decimal value goes on forever.

Because any repeating decimal can be written as a fraction, we know that the sequence cannot repeat.

The real number $\sqrt{2}$ is defined to be the limit of this sequence

$1.4, 1.41, 1.414, \ldots 1.414214 \ldots$

as the number of terms $n \to \infty$.

In a similar way, the number $e$ can be viewed as

$$\lim_{n\to\infty} (1 + \frac{1}{n})^n$$

And the number $\pi$ can be viewed as the limit of the method of exhaustion applied to the area of a unit circle.

# Chapter 20

# Continuum of numbers

In a previous chapter we showed that given any two rational numbers one can find a rational number which lies between them.

Three related statements are also true. We will show that

∘ for any two rational numbers one can find a real number which lies between them

∘ for any two real numbers one can find a rational number which lies between them

∘ for any two real numbers one can find a real number which lies between them

### continuum

∘ Between any two *real* numbers it is always possible to find a rational number.

Proof: pick

$$N \in \mathbb{N} \text{ such that } N > \frac{1}{b-a}$$

Then

$$\frac{1}{N} < b - a$$

Define the set $\mathbf{A}$ as follows:

$$\mathbf{A} = \{ \frac{m}{N} : \ m \in \mathbb{N}\}, \quad \text{a subset of } \mathbb{Q}$$

The claim is that

$$\mathbf{A} \cap (a, b) \neq \emptyset$$

There do exist numbers within the open interval $(a, b)$ that are in the set $\mathbb{Q}$.

The proof is by contradiction. Assume on the contrary that the set $\mathbf{A}$ does not contain a rational number lying inside this interval. In other words:

$$\mathbf{A} \cap (a, b) = \emptyset$$

Now, find the largest integer $m_1$ such that $m_1/N < a$ (it is OK if $m_1$ is equal to 0). Then the next rational number in $\mathbf{A}$ must be larger than $b$ since the set intersection is empty:

$$\frac{m_1 + 1}{N} > b$$

But this implies that

$$\frac{m_1 + 1}{N} - \frac{m_1}{N} > b - a$$

$$\frac{1}{N} > b - a$$

which contradicts our condition on $N$ above. Hence the assumption is false and so

$$\mathbf{A} \cap (a, b) \neq \emptyset$$

Thus there must exist a rational number $r$ in $\mathbf{A}$ such that $a < r < b$.

**example**

Consider the open interval: $(\sqrt{2}, \sqrt{3})$.

$$a = \sqrt{2} \approx 1.414$$

$$b = \sqrt{3} \approx 1.732$$

$$b - a \approx 0.3178$$

$$\frac{1}{b-a} \approx 3.1462$$

Pick $N \geq 4$, for example

$$N = 4: \quad 1.414 < \frac{6}{4} = 1.5 < 1.732$$

$$N = 5: \quad 1.414 < \frac{8}{5} = 1.6 < 1.732$$

$$N = 6: \quad 1.414 < \frac{9}{6} = 1.5 < 1.732$$

(In this case $N = 2$ and $N = 3$ happen to work as well).

◦ Between any two rational numbers it is always possible to find a real number.

One proof consists of finding a *particular* irrational in the interval $(a, b)$, where $a$ and $b$ are rational. For $a < b$, we simply add to the number $a$ the following

$$c = \frac{\sqrt{2}}{2}(b - a)$$

$c$ is smaller than $b - a$ (because $\sqrt{2}/2 < 1$) so the result $a + c$ lies between $a$ and $b$. We also know that $c$ is irrational, because $\sqrt{2}$ times any rational number is irrational. Finally, $a + c$ is irrational because adding $\sqrt{2}$ times a rational number to any rational number produces an irrational number.

Proof of the first preliminary requirement: $\sqrt{2}$ times a rational is irrational. Suppose for integer $p, q, r, s$ we have

$$\sqrt{2}\,\frac{p}{q} = \frac{r}{s}$$

then

$$\sqrt{2} = \frac{rq}{ps}$$

But the right-hand side is rational, so this is a contradiction.

For the second requirement, again by contradiction suppose

$$\sqrt{2}\,\frac{p}{q} + \frac{s}{t} = \frac{u}{v}$$

for integer $p, q, r, s, u, v$. But the right-hand side of

$$\sqrt{2} = \frac{q}{p}\left(\frac{u}{v} - \frac{s}{t}\right)$$

is rational, so this is a contradiction.

Note in passing that powers are different. What do you think about

$$r = \sqrt{2}^{\sqrt{2}}$$

You may think $r$ is "likely" to be irrational. Just a mess. But how about

$$r^{\sqrt{2}}$$

Whether $r$ is rational or irrational

$$r^{\sqrt{2}} = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{2} = 2$$

!!

◦ Between any two real numbers it is always possible to find another real number. This one is subtle. Suppose the two real numbers are "really, really close."

We suppose that they are not equal, so they must be different, say $a < b$.

Since they are different, at some stage in the decimal expansions of $a$ and $b$, there must be a first position at which $a$ and $b$ differ. If $b$ does not have a 0 at the next position, terminate there and that will be $c$.

For example:

$$a = 1.23456789129..$$

$$b = 1.23456789133..$$

$$c = 1.23456789130..$$

$b$ must have some digit following this first position where it does not match $a$, and which is also not equal to zero (otherwise it would be a terminating decimal and thus a rational number). So we can always find a place to terminate to form $c$.

**Eternity is a very long time, especially towards the end.**

(credited to Woody Allen)

**variations of infinity**

In other words there is *no least number* $x$ such that $x > 0$, for example, and no greatest number $x$ such that $x < 1$.

Proof: Assume that $m$ is the smallest number $> 0$. The rational number $m/2 < m$ is also greater than zero, but smaller than $m$. Thus, $m$ is not the smallest positive number.

In general, there is no number that is the closest number to another number.

That is actually OK. Here's what's really weird. Cantor proved that the set $\mathbb{Q}$ is *countably finite*. Each element in $\mathbb{Q}$ can be paired in order with a member of $\mathbb{N}$.

The idea of the proof is to show that one can set up a correspondence between $\mathbb{N}$ and $\mathbb{Q}$, assigning each number $r \in \mathbb{Q}$ in a particular order to $1, 2, 3, \ldots$. Here is the figure from Courant and John:

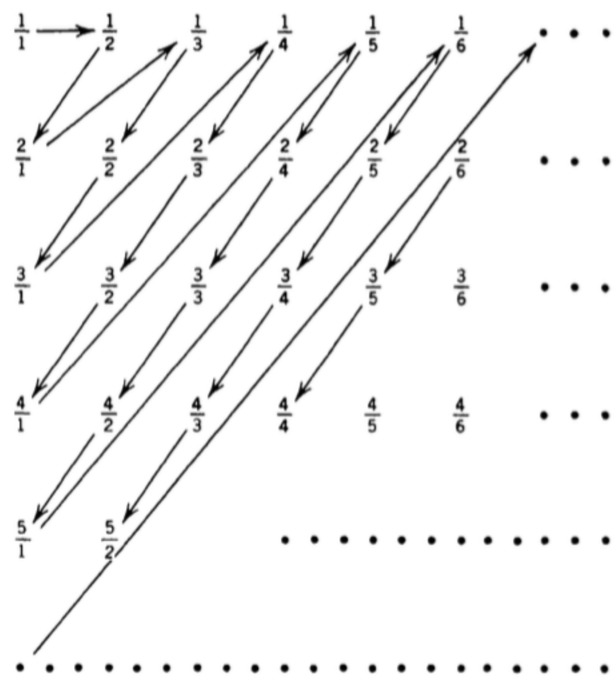



**Figure 1.S.1** Denumerability of the positive rationals.

Basically, each row contains all the rational numbers with a particular numerator, and each column contains all the numbers with a particular denominator, arranged in strict increasing order.

Next, set up the sequence indicated by the arrows:

$$\frac{1}{1} \quad \frac{1}{2}, \frac{2}{1} \quad \frac{1}{3}, \frac{2}{2}, \frac{3}{1} \quad \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1} \quad \frac{1}{5}, \frac{2}{4}, \frac{3}{3}, \frac{4}{2}, \frac{5}{1} \quad \frac{1}{6}, \frac{2}{5}, \frac{3}{4}, \frac{4}{3}, \frac{5}{2}, \frac{6}{1} \cdots$$

Then remove all fractions that are duplicates because they are not in lowest terms.

$$\frac{1}{1} \quad \frac{1}{2}, \frac{2}{1} \quad \frac{1}{3}, \frac{3}{1} \quad \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1} \quad \frac{1}{5}, \frac{5}{1} \quad \frac{1}{6}, \frac{2}{5}, \frac{3}{4}, \frac{4}{3}, \frac{5}{2}, \frac{6}{1} \cdots$$

Finally, each $r$ in this sequence is assigned to a natural number (in the sequence $\mathbb{N}$), establishing the denumerability property. $1/3$ is paired with 4 and $3/1$ is paired with 5, and so on.

Cantor showed that such a correspondence (which we just established for $\mathbb{Q}$), is impossible for $\mathbb{R}$. The proof of this is not hard, but we will skip it here. You can check out the chapters on Georg Cantor in Dunham's *Journey Through Genius*.

Thus, the rational numbers are said to be "countably infinite", while the real numbers are not countable. (There is also a proof that the transcendental numbers are much more numerous than the non-transcendental ones).

We say that the set of numbers greater than 0 has *no least element*. We can test this by picking the smallest rational member imaginable, but subsequently, we can always find a smaller rational element (say, by halving that number).

And once we get really close with the small rational element, there are infinitely more irrational than rational ones waiting beyond. And yet, given any such very close irrational number, we can always find a smaller rational number, still larger than the bound.

I told you it was weird.

This property of the real numbers, that there is no closest number to any given number, accounts for virtually all of the theoretical difficulties in calculus which are solved by the use of limits and the apparatus of $\delta$ and $\epsilon$ or alternatively, neighborhoods. We will get to that in a bit.

# Part VI

# Analytic geometry and trigonometry

# Chapter 21

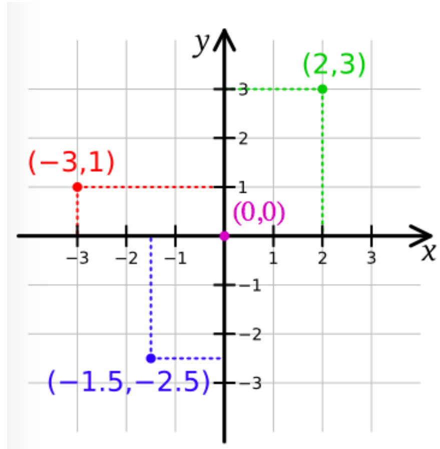
# Analytic geometry

It is difficult today to put ourselves in the place of those who tried to reason about mathematics through the ages.

The Greeks lacked algebra, and although the Romans worked with numbers they did not have decimal notation. The concept of 0 came much later (from India), and even in the Middle Ages there was as yet no such thing as the equals sign =, which dates from 1557.

https://en.wikipedia.org/wiki/Table_of_mathematical_symbols_by_introduction_date

The invention of analytic geometry is often ascribed solely to Descartes, but Fermat also had his own version. There are two fundamental ideas.




The first is to orient two number lines on a piece of paper, at right angles, and then consider pairs of numbers $(x, y)$ in the 2D plane. Such pairs or tuples are called points.

Descartes published this idea in 1637. The presentation would be difficult to recognize as our current system, but the germ is there: axes where the position of a variable could be marked. Only the positive numbers would be shown, and the axes not necessarily perpendicular. As to the proofs, here is wikipedia on the subject:

> His exposition style was far from clear, the material was not arranged in a systematic manner and he generally only gave indications of proofs, leaving many of the details to the reader. His attitude toward writing is indicated by statements such as "I did not undertake to say everything," or "It already wearies me to write so much about it," that occur frequently. In conclusion, Descartes justifies his omissions and obscurities with the remark that much was deliberately omitted "in order to give others the pleasure of discovering [it] for themselves."

The second idea of analytic geometry is to plot all the points that satisfy some mathematic relationship between $x$ and $y$, for example the parabola $y = x^2$.

To do this, pick a few values of $x$ and calculate the corresponding values of $y$. For example: $(0, 0), (\pm 1, 1), (\pm 2, 4), \ldots$. Plot these points, and then finally, sketch the graph of the curve, without actually trying to plot *all* of the individual points (of which there is an infinite number). We make the assumption here that the function being plotted is continuous, so that the sketch of a curve between two points that are close enough together will be fairly smooth and if the $x$-values are close to the plotted $x$, the corresponding $y$-values will not be not too different from the plotted $y$.

**point**

A point is simply an ordered pair $(x, y)$ such as $(1, 3)$. Often points have integer components, but they don't have to be.

**distance formula**

The $x$- and $y$-axes are perpendicular to one another (a fancy word for that is *orthogonal*).

Suppose we pick two particular points $(s, t)$ and $(u, v)$, plot them on a graph, and then draw the line that connects them. Recall Euclid's first two postulates:

◦ A straight line segment can be drawn joining any two points.

◦ Any straight line segment can be extended indefinitely in a straight line.

The distance between the two points is given by the Pythagorean formula, where $\Delta x$ is the change in $x$ and $\Delta y$ is the change in $y$:

$$d = \sqrt{\Delta x^2 + \Delta y^2}$$

It is often easier to use the squared distance and avoid the square root:

$$\begin{aligned} d^2 &= \Delta x^2 + \Delta y^2 \\ &= (s-u)^2 + (t-v)^2 \end{aligned}$$

Switching the order of $(s, t)$ and $(u, v)$ doesn't change the result.

### formulas for a line

Now we want to derive an equation that describes (is valid for) all the points or pairs of values $(x, y)$ on this line. A general approach is to say that the line has some slope $m$, which is defined as $\Delta y$, divided $\Delta x$:

$$m = \frac{\Delta y}{\Delta x} = \frac{y - y'}{x - x'}$$

This is called the *point-slope equation*. For any two particular points $(s, t)$ and $(u, v)$ one can plot a line between them. The slope is

$$m = \frac{s-u}{t-v}$$

One can write the two points in either order, with the same result since:

$$\frac{s-u}{t-v} = \frac{u-s}{v-t}$$

Depending on the details, the value of $m$ might be zero, for a horizontal line, where all the values of $y$ are the same (which happens when $s = u$). Or it might be undefined, for a vertical line, where all the values of $x$ are identical ($t = v$).

In most cases, however, $m \neq 0$ and $m \in (-\infty, \infty)$. That is, $m$ is usually non-zero and not infinite.

Except in the case of the vertical line, we can write

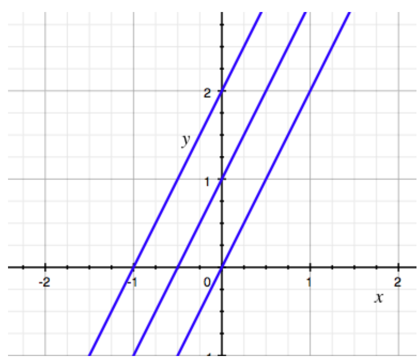
$$y = mx + y_0$$

for any point $(x, y)$ on a given line, where $y_0$ is the $y$-intercept, the value of $y$ obtained when $x = 0$.

[ The choice of $b$ for the $y$-intercept is the usual notation, but it conflicts with another $b$ that we will see in a minute. ]

$y = mx + y_0$ is the *slope-intercept equation* of the line.

The equation of a line is determined by both the slope and one point on the line, for example the $y$-intercept. One can draw a whole family of parallel lines with the same slope and different $y$-intercepts. Here are three lines $y = 2x + y_0$ for $y_0 = \{0, 1, 2\}$.




The value of $x$ corresponding to $y = 0$ is the $x$ intercept

$$x_0 = -\frac{y_0}{m}$$

The point-slope equation is easily derived from the second one. Suppose we have $y = mx + y_0$:

Plugging in for specific points $(s, t)$ and $(u, v)$ we have

$$t = ms + y_0$$

$$v = mu + y_0$$

Subtracting:

$$v - t = m(u - s)$$

which rearranges to give the desired result.

**intersections**

Often one has two lines (or curves) and we want to find the point(s) that lie on both. We might have

$$y = 2x - 1$$

$$y = -x + 8$$

Substitute from the second into the first:

$$2x - 1 = -x + 8$$

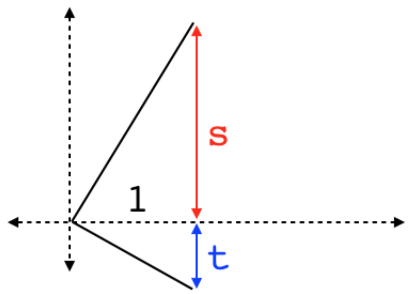
$$3x = 9$$

$$x = 3$$

From the first equation, $y = 5$, and we check that $x = 3, y = 5$ solves the second equation as well.

**orthogonality**

If two lines cross each other at right angles we say they are *orthogonal*. In that case the slopes have a special relationship. Their product is equal to $-1$.




Here is a simple proof. Draw the two lines going through the origin, forming a right angle there. The first has slope $s$, so it goes through the point $(1, s)$, the second goes through $(1, t)$.

Recall from the chapter on the Pythagorean theorem that the altitude squared is equal to the product of the two pieces of the base. Here:

$$1^2 = 1 = st$$

These are the lengths, i.e. the absolute values of the slopes. Thus $|s| = 1/|t|$. But

Clearly the sign of $t$ is negative. So we arrive at

$$s \cdot (-t) = 1$$

$$m_1 = -\frac{1}{m_2}$$

We'll see a natural easy proof of this once we look at trigonometry.

### formula for a circle

A circle can be defined as all the points at the same distance from a central point, let us label that point $(h, k)$. The distance from the points to the center is the radius, denoted $r$.

Using the Pythagorean theorem, we can calculate the square of the distance from the origin as

$$r^2 = (x - h)^2 + (y - k)^2$$

The simplest circles are those whose central point is the origin of the coordinate system. In that case the equation simplifies to

$$r^2 = x^2 + y^2$$

Usually, we know the value of $r$ and we want to write an equation for $y$ in terms of $x$. Then

$$y^2 = r^2 - x^2$$
$$y = \sqrt{r^2 - x^2}$$

**formula for a parabola**

A general formula for a parabola with its vertex at the point $(h, k)$ is

$$y - k = a(x - h)^2$$

where $a$ is called the shape factor. It governs how steeply the curve rises (and by its sign, in which direction it opens).

Multiplying out:

$$y - k = a(x^2 - 2xh + h^2)$$
$$y = ax^2 - 2ahx + ah^2 + k$$

In this form the cofactors are usually simplified as

$$y = ax^2 + bx + c$$

where

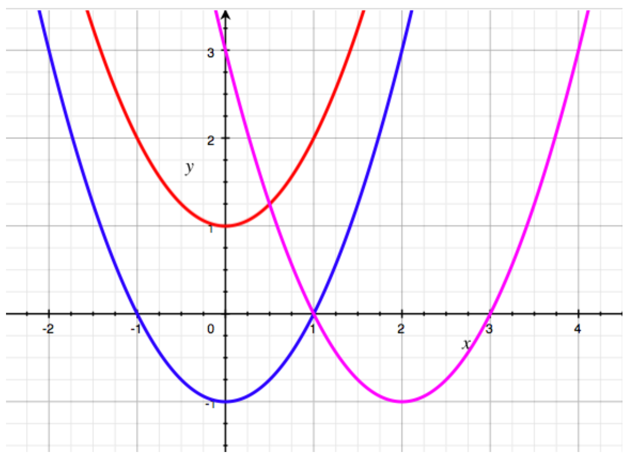
$$b = -2ah; \quad c = ah^2 + k$$

If the equation is given in the second form then we can find:

$$h = -\frac{b}{2a}$$

$$k = c - ah^2 = c - \frac{b^2}{4a}$$

Probably the most common thing we're asked to do with a quadratic equation like this is to find the roots, the values of $x$ for which $y = 0$ is a solution. These are the points where the graph of the curve crosses the $x$-axis. (It may not do so, it is possible to have 0, 1 or 2 roots).




In the figure, the red curve does not cross the $y$-axis. Its equation is $y = x^2 + 1$, and there are no solutions, no values of $x$ that solve the equation when $y = 0$.

$$0 = x^2 + 1$$

$$x^2 = -1$$

To find the roots of

$$ax^2 + bx + c = 0$$

We can guess solutions by trying to factor into a form like:

$$(x - s)(x - t) = 0$$

but roots do not have to be integers (or even rational). An arguably more productive and certainly more general approach is the process of *completing the square.*

First, multiply through by $1/a$ and rearrange:

$$x^2 + \frac{b}{a}x = -\frac{c}{a}$$

The key insight is to recognize that if we add $(b/2a)^2$ to both sides, the left-hand side will become a perfect square:

$$x^2 + \frac{b}{a}x + (\frac{b}{2a})^2 = -\frac{c}{a} + (\frac{b}{2a})^2$$

$$(x + \frac{b}{2a})^2 = -\frac{c}{a} + (\frac{b}{2a})^2$$

$$x + \frac{b}{2a} = \pm\sqrt{-\frac{c}{a} + (\frac{b}{2a})^2}$$

Multiplying top and bottom of the first term under the square root gives a common factor:

$$x + \frac{b}{2a} = \pm\sqrt{-\frac{4ac}{4a^2} + (\frac{b}{2a})^2}$$

which can come out of the square root and then matches what's in the second term on the left-hand side:

$$x + \frac{b}{2a} = \pm\frac{\sqrt{-4ac + b^2}}{2a}$$

which we rearrange slightly to give the standard *quadratic formula*:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

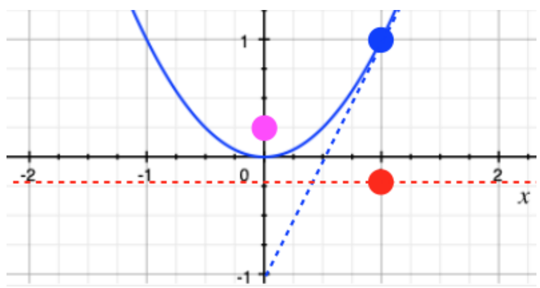
**focus and directrix**

There is also a classic geometric definition of the parabola.

Based on what we said above, we can transform any parabola of the form $y = ax^2 + bx + c$ into a $(y - k) = a(x - h)^2$. If we're interested in the shape of the parabola and don't care about its absolute location, then without loss of generality, we can translate any parabola to the origin of coordinates, with equation $y = ax^2$, so let us just work with that.

Now, pick a point on the $y$-axis a distance $p$ up from the origin, colored magenta in the figure. This point is called the focus.

Then draw a line parallel to the $x$-axis which intersects the $y$-axis the same distance $p$ below the origin. This line is called the directrix. It is colored red and is dashed.




The parabola consists of all those points whose distance to the focus is equal to the vertical distance to the directrix.

Pick an arbitrary point on the parabola (in blue), with coordinates $(x, ax^2)$. The squared distance to the focus (magenta) is

$$x^2 + (ax^2 - p)^2$$

where $\Delta x$ is just equal to $x$ and $\Delta y$ is equal to $y - p$, with $y = ax^2$.

The squared distance to the directrix (red) is $(ax^2 + p)^2$.

For the correct choice of $p$ these distances will be equal:

$$x^2 + (ax^2 - p)^2 = (ax^2 + p)^2$$

We have $(m-n)^2$ on the left-hand side and $(m+n)^2$ on the right-hand side, so the result will have $4mn$ on the right hand side:

$$x^2 = 4apx^2$$

Divide by $x^2$

$$1 = 4ap$$

$$p = \frac{1}{4a}$$

The shape factor $a$ determines the distance of the focus from the origin, we label that distance as $p$. The equation of the directrix is $y = -p$.

**slope of the tangent**

It will turn out that the slope of the tangent to $y = ax^2$ at any fixed point $x$ is equal to $2ax$.

This is literally the first result from differential calculus, but we will also see a way to find it using analytical geometry in the next chapter, as well as a vector approach later on.

Thus, the equation of a line passing through the point $(x, ax^2)$ with the given slope is

$$y' - ax^2 = 2ax(x' - x)$$

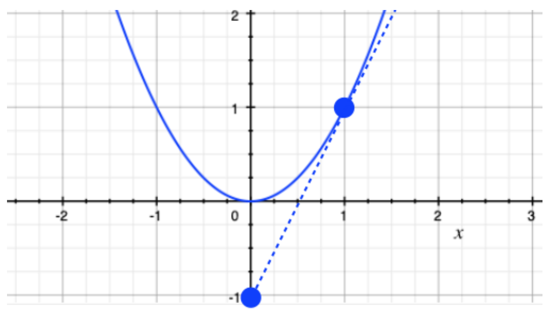
where $(x', y')$ is any other point on the line.

What *that* means is that the $x$-intercept $x_0$ of the tangent line is:

$$-ax^2 = 2axx_0 - 2ax^2$$

$$
\begin{aligned}
ax^2 &= 2axx_0 \\
x &= 2x_0 \\
x_0 &= \frac{1}{2}x
\end{aligned}
$$

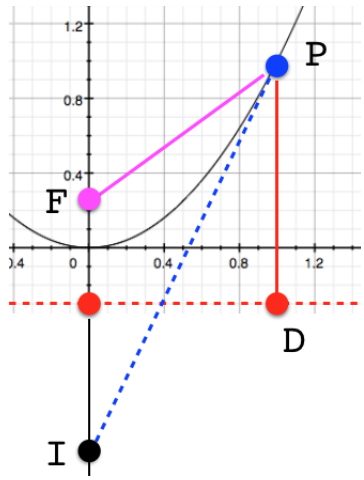
The tangent line passes through the $x$ axis halfway back toward the origin.




And what *that* means is that the $y$-intercept is symmetrical with the original point (as far below the $x$-axis as the point is above it). Here's the algebra:

$$
\begin{aligned}
y_0 - ax^2 &= 2ax(0 - x) \\
y_0 &= -ax^2
\end{aligned}
$$

And then finally, if the point on the parabola is $P$, the focus $F$, the intersection with the directrix $D$, and the $y$-intercept $I$




the quadrilateral $FPDI$ is a regular parallelogram with all four equal sides, and its long diagonal (the tangent line) makes equal angles with $FP$ and $PD$.

And what *that* means is that if $PD$ is extended vertically, the angle it makes with the tangent line is equal to the angle between $FP$ and the tangent line, so that for example, all vertical light rays entering a parabola will reflect and then come together at the focus.

# Chapter 22

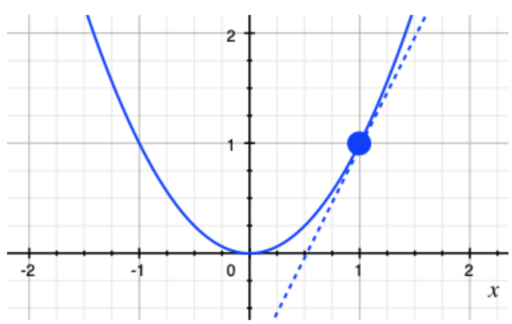
# Slope of a parabola

In this chapter, we show how to find the slope of parabola at any point using classical methods.

### part 1

Consider the simplest parabola: $y = x^2$.

The point $(1, 1)$ is on the curve, because $(x = 1, y = 1)$ satisfies the equation $y = x^2$.




Suppose we know that the slope of the tangent to the curve at the point $(1, 1)$ is equal to 2.

(Using calculus to find this result is trivial, we'll also show a noncalculus method in part three, below).

The equation of the tangent line is

$$y_2 - y_1 = m(x_2 - x_1)$$

Plugging in for $(x_2, x_1) = (1, 1)$ (and just writing $(x, y)$ for $(x_1, y_1)$):

$$y - 1 = 2(x - 1)$$

$$y = 2x - 1$$

Now suppose that we knew only the parabola and this slope, but we did not know the point where the tangent meets the curve, and so do not know the $y$-intercept.

We have the equation of a line:

$$y = 2x + y_0$$

We seek points which are simultaneously on the line and the curve. They must satisfy both equations.

Since this is a tangent line, we seek the value for which this expression has only a single solution. The tangent "kisses" the curve at a single point.

So, substitute for $y$ from the equation for the curve:

$$x^2 = 2x + y_0$$

$$x^2 - 2x - y_0 = 0$$

Now look at the quadratic formula we would use to solve this equation for $x$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

There is a single solution when the part under the square root (called the discriminant) is equal to zero.

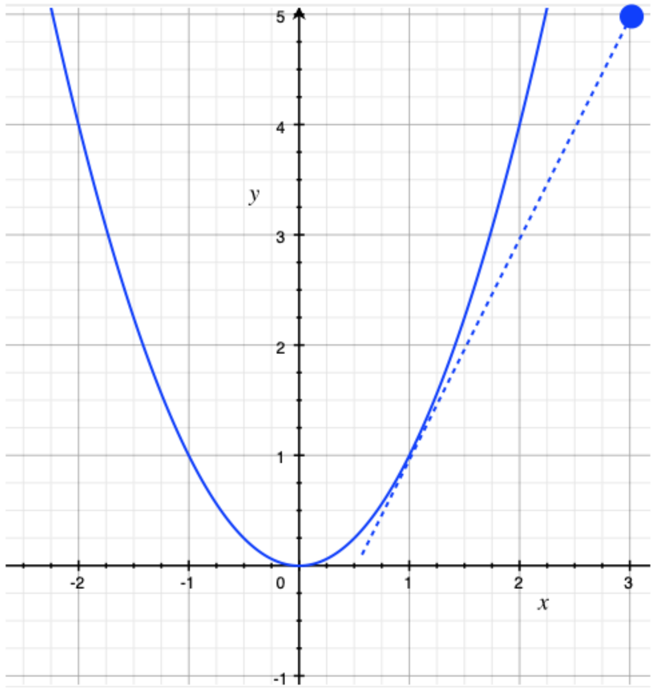
$$b^2 - 4ac = 0$$
$$b^2 = 4ac$$
$$(-2)^2 = 4(-y_0)$$
$$y_0 = -1$$

Therefore, the equation of the tangent line is $y = 2x - 1$, which matches what we had before.

In general, $y = 2x + y_0$ is a *family* of lines. For $y_0 = -1$, there is a single solution for $x$ to be both on the line and the parabola. For $y_0 < -1$, there are no solutions, while for $y_0 > -1$ there are two solutions, because the line actually traces out a secant of the parabola, passing through the curve at two points.

### part 2

Now suppose we have the same parabola and a point not on the parabola, but in the plane and outside of the ”cup” of the parabola, such as $(3, 5)$. We seek the equations of tangent lines to the parabola that go through this point.




There will be two of them. We show just one in the figure.

The equations of lines passing through this point, with different slopes $m$ are given by:

$$(y_2 - y_1) = m(x_2 - x_1)$$

Here, let $(x_2, y_2)$ be $(3, 5)$ and then multiply by $-1$, and drop the subscript, to obtain:

$$y - 5 = m(x - 3)$$

Since values of $(x, y)$ are both on the line and the parabola $y = x^2$, we can plug in for $y$:

$$x^2 - 5 = mx - 3m$$

$$x^2 - mx + (3m - 5) = 0$$

As before, solutions are given by the quadratic equation. The value of the slope $m$ giving a single solution (zero discriminant) is:

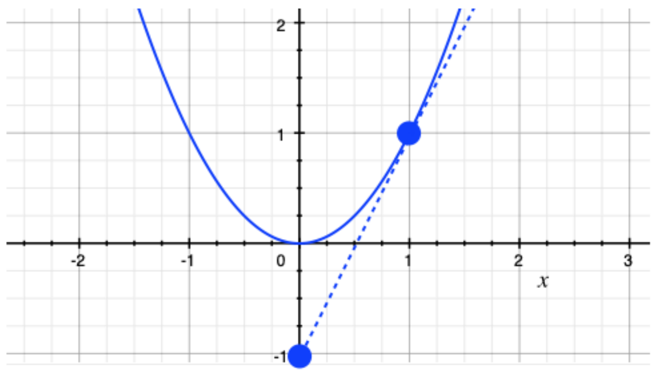
$$(-m)^2 - 4(3m - 5) = 0$$

$$m^2 - 12m + 20 = 0$$

$$(m - 2)(m - 10) = 0$$

$$m = 2, \quad m = 10$$

We knew the first one already, because the point $(3, 5)$ is on the line $y = 2x - 1$. This is the tangent to the curve at $(1, 1)$, which has slope $m = 2$.




Actually, there is always another solution. Any vertical line (with infinite slope) passes through only a single point on the parabola.

Basically what this amounts to is that in the equation

$$x = \frac{m \pm \sqrt{(-m)^2 - 4(3m - 5)}}{2}$$

as $m$ gets very large, only the term $(-m)^2$ matters under the square root, so we have

$$x = \frac{m \pm \sqrt{(-m)^2}}{2}$$

and as $m \to \infty$, $m - \sqrt{m^2} \to 0$.

**part 3**

Now suppose we are given the same parabola again and also a point on it such as $(x_1, y_1)$.

Any line through that point has the equation:

$$y - y_1 = m(x - x_1)$$

To find the equation of a tangent line through that point we need the slope $m$.

If there is a point $(x, y)$ that is on the line and *also* on the parabola, it must satisfy $y = ax^2$ as well, so:

$$ax^2 - ax_1^2 = m(x - x_1)$$

$$ax^2 - mx - ax_1^2 + mx_1 = 0$$

Certainly $x = x_1$ is a solution.

The value of $m$ must be such that there are *no other solutions*.

Write the quadratic equation to solve for $x$:

$$x = \frac{m \pm \sqrt{m^2 - 4a(mx_1 - ax_1^2)}}{2a}$$

There is a single solution when the discriminant is zero, that is, when

$$x = \frac{m}{2a}$$

$$m = 2ax$$

Since $x = x_1$ for the tangent line

$$m = 2ax_1$$

as expected.

The slope of the tangent line is $2ax_1$ and in particular, at the point $(1,1)$, the slope is equal to 2.

That's the answer, but there are two points to follow up on. We should plug the answer into these two equations and check what happens. We need

$$ax^2 - mx - ax_1^2 + mx_1 = 0$$

and

$$m^2 - 4(mx_1 - ax_1^2) = 0$$

For the first one:

$$ax^2 - 2ax^2 - ax_1^2 + 2ax_1$$

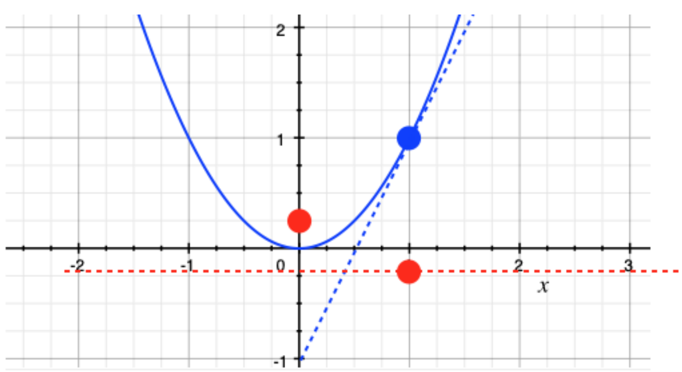
This is certainly equal to zero when $x = x_0$.

Then

$$m^2 - 4a(mx_1 - ax_1^2)$$
$$4a^2x^2 - 4a(2ax_1 - ax_1^2)$$
$$4a^2x^2 - 8a^2x_1^2 + 4a^2x_1^2$$

is also equal to zero when $x = x_0$.

**alternate solution**

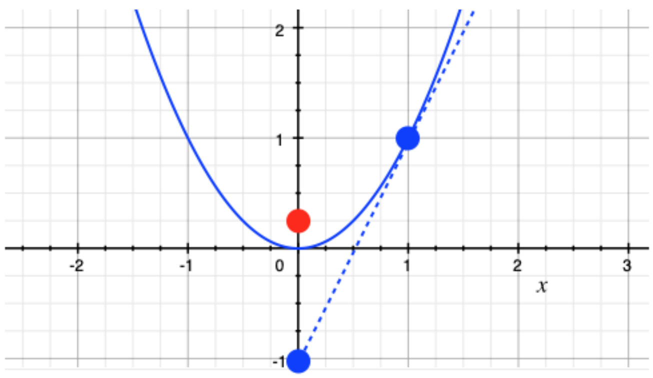


A parabola is defined geometrically by its focus, which is the point $(p, 0)$ for a centered parabola.

The focus is paired with a directrix, which is the line $y = -p$ for a vertex at the origin.

All points on the parabola lie at the same distance $d$ from the focus and the directrix.

A relatively advanced fact about the parabola is that any tangent line intersects the $y$-axis at the same distance $d$ from the focus.




Which is to say that if we draw a triangle in the above diagram using the two blue points and one red one, the two blue points are the vertices of equal angles and the triangle formed is isosceles.

For $y = x^2$, consider the point $(x, x^2)$, and find the distance to the focus squared as

$$d^2 = (x)^2 + (x^2 - p)^2$$

$$d^2 = x^2 + x^4 - 2x^2p + p^2$$

Call the $y$-intercept $k$ so then

$$k + d = p$$

$$d^2 = p^2 - 2pk + k^2$$

Equating the two expressions:

$$p^2 - 2pk + k^2 = x^2 + x^4 - 2x^2p + p^2$$

$$k^2 - 2pk = x^2(1 + x^2 - 2p)$$

In this case, we know $x = 1$ and $p = 1/4$ so

$$k^2 - \frac{k}{2} - (2 - \frac{1}{2}) = 0$$

We factor to obtain:

$$(k + 1)(k - \frac{3}{2}) = 0$$

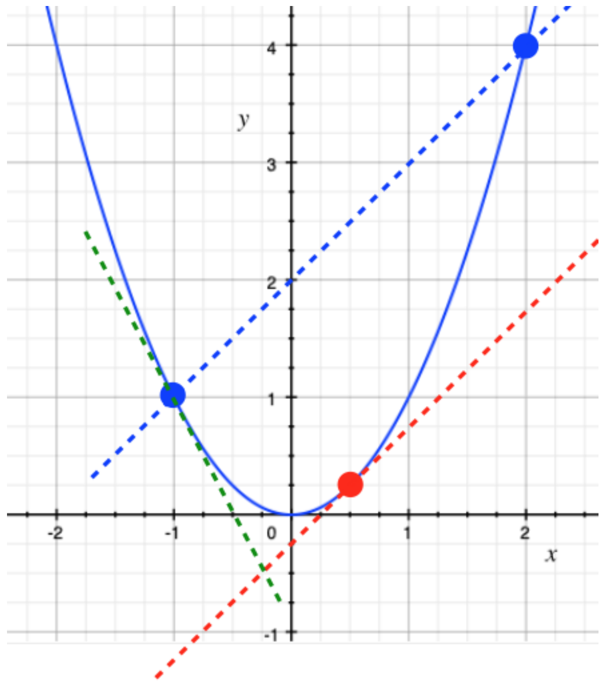
$k = -1$ was our solution above.

[ ?? What is the significance of the other solution?? ]

**further comment**

The slope of the parabola has some simple interesting properties. For example, pick any two points $(x, y)$ and $(x', y')$ on our standard parabola.

The slope of the line that connects those two points is equal to the slope of the parabola at the point whose $x$-value is halfway in between.




For the first part:

$$m = \frac{y' - y}{x' - x}$$

$$= \frac{ax'^2 - ax^2}{x' - x}$$

$$= a\ [\ \frac{x'^2 - x^2}{x' - x}\ ]$$

$$= a(x' + x)$$

For the midpoint

$$x_m = \frac{1}{2}(x' + x)$$

and the slope is

$$2a \cdot \frac{1}{2}(x' + x)$$

$$= a(x' + x)$$

A similar result is that if we pick any two points $(x, y)$ and $(x', y')$, and draw their slopes, the point where the two slope lines meet has its $x$-value exactly halfway in between $x$ and $x'$.

**circle**

Suppose we have a unit circle and an external point $(x, y)$. We wish to find the equation of the tangent line to a point on the circle. Call that point $(a, b)$.

Circles are special. Any tangent is perpendicular to the radius at the point of tangency.

The line through $(a, b)$ and the origin has slope $b/a$ since

$$m = \frac{b - 0}{a - 0}$$

If we have two lines with slopes $m_1$ and $m_2$ and they are perpendicular, the product is $-1$. So the tangent to the circle at $(a, b)$ has slope $-a/b$ and the line through $(x, y)$ and $(a, b)$ is

$$-\frac{a}{b} = \frac{y - b}{x - a}$$
$$-ax + a^2 = by - b^2$$

We also have that $a^2 + b^2 = 1$ so

$$ax + by = 1$$

Substitute into the equation of the circle:

$$a^2 + (\frac{1 - ax}{y})^2 = 1$$

$$
\begin{aligned}
a^2y + 1 - 2ax + a^2x^2 &= y \\
(y + x^2)a^2 - 2xa + (1 - y) &= 0
\end{aligned}
$$

We have a quadratic in $a$. For a particular $x$ and $y$, we can solve for $a$.

### ellipse

I found a problem on the web that extends this to the ellipse:

https://math.stackexchange.com/questions/834392/equations-of-lines-tangent-to-an-ellipse

In working that problem, I ended up with a quartic equation (fourth power). This is, quite literally, a mess.

Here's a great idea for ellipse problems: Stretch and rescale the problem to one involving a circle, by using a *change of variable*. Suppose the ellipse is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Let $x = au$ and $y = bv$. Then

$$
\begin{aligned}
\frac{a^2u^2}{a^2} + \frac{b^2v^2}{b^2} &= 1 \\
u^2 + v^2 &= 1
\end{aligned}
$$

The ellipse has become a unit circle!

Apply the same transformation to any points in the problem, solve the problem, and then reverse the transformation.

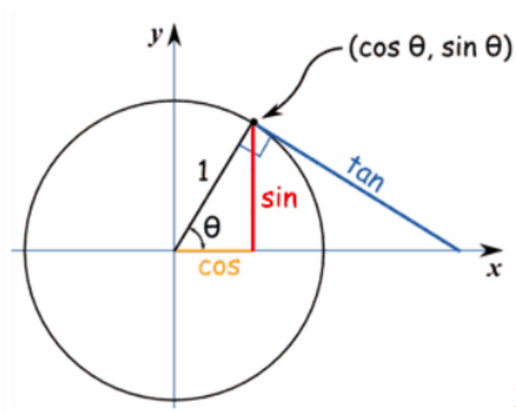
# Chapter 23

# Six functions

The most elementary trigonometric functions are sine and cosine.

### basic definitions

The "unit circle" is a circle of radius 1 with its center positioned at the origin of coordinates, the place where the $x$ and $y$ axes cross. From the diagram you can see that any point $(x, y)$ on the unit circle can be described in radial coordinates as $(\cos\theta, \sin\theta)$.




That is:

$$x = \cos\theta \quad\quad y = \sin\theta$$

If the circle has radius $r$ then

$$x = r\cos\theta \qquad y = r\sin\theta$$

The tangent is

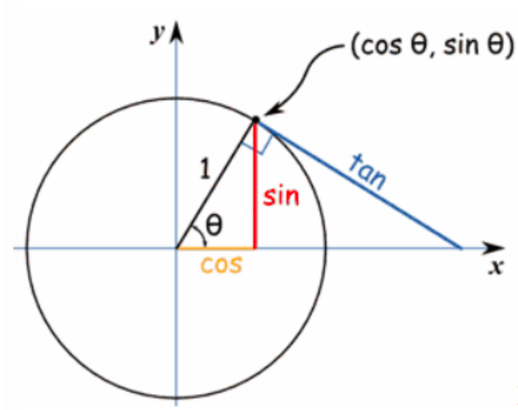
$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

In the diagram, all three right triangles are similar.

Thus, by similar triangles, the blue side has this relationship

$$\frac{\text{blue side}}{1} = \frac{\sin\theta}{\cos\theta}$$

which explains why it is labeled as it is.




Stewart:

> The mathematicians of ancient India built on the Greek work to make major advances in trigonometry. They [used] the sine (sin) and cosine (cos) functions, which we still do today. Sines first appeared in the Surya Siddhanta, a series of Hindu astronomy texts from about the year 400, and were developed by Aryabhata in Aryabhatiya around 500. Similar ideas evolved independently in China.

The other functions are the inverses of sine, cosine and tangent, namely: cosecant, secant and cotangent. The secant (inverse cosine) comes up, but the other two are not especially important in calculus. However, they do come up in one context that we will look at, Archimedes determination of the value of $\pi$. The crucial step in that approach will turn out to be the calculation of the cotangent of the half-angle $\theta/2$ given the values of cotangent and cosecant for angle $\theta$.

The main relationship or identity is derived from the Pythagorean theorem. We had above that for a unit circle

$$x = r\cos\theta \qquad y = r\sin\theta$$

Since $x$ and $y$ are the sides of a right triangle whose hypotenuse is $r$

$$x^2 + y^2 = r^2$$

and for a unit circle

$$\cos^2\theta + \sin^2\theta = 1$$

which is usually written

$$\sin^2\theta + \cos^2\theta = 1$$

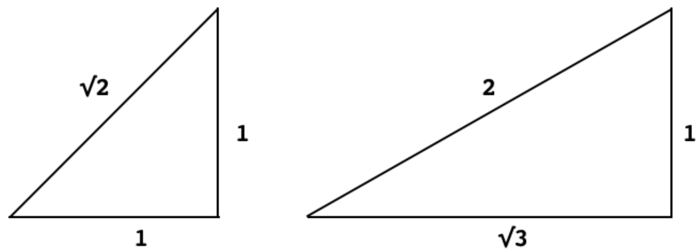
and transformed to

$$1 + \tan^2\theta = \sec^2\theta$$

It is assumed you've studied trigonometry before.

We can easily determine the values for these functions for three special cases.

The first is the angle 45 degrees or $\pi/4$. Draw an isosceles right triangle with sides of length 1 (left panel).




Then the hypotenuse has length $\sqrt{2}$ (from Pythagoras) and the values are

$$\sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} = \cos\frac{\pi}{4}$$

$$\tan\frac{\pi}{4} = 1$$

For the other two, bisect an equilateral triangle and erase one half (right panel). The smaller angle is 30 degrees or $\pi/6$ and its complement is 60 degrees or $\pi/3$.

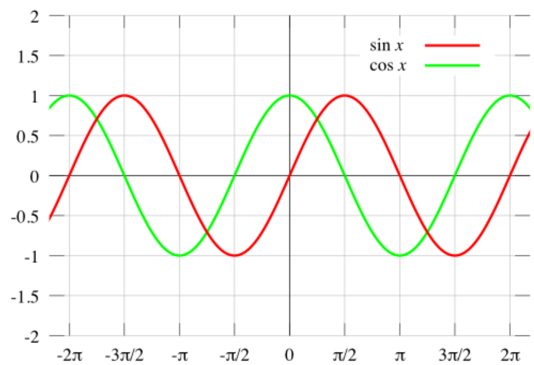
The values are

$$\sin\frac{\pi}{6} = \frac{1}{2} = \cos\frac{\pi}{3}$$

$$\cos\frac{\pi}{6} = \frac{\sqrt{3}}{2} = \sin\frac{\pi}{3}$$

$$\tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}$$

**graph**

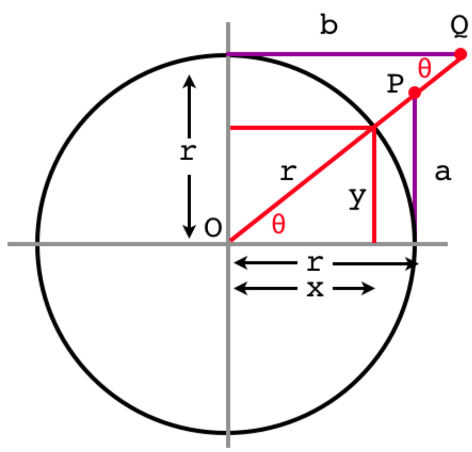



Savov:

> The sine function represents a fundamental unit of vibration. The graph of $\sin(x)$ oscillates up and down and crosses the $x$-axis multiple times. The shape of the graph of $\sin(x)$ corresponds to the shape of a vibrating string.

**visualization of all six functions**

Consider a unit circle. Extend the radius with the angle $\theta$ and then draw the vertical and horizontal tangents to the circle $a$ and $b$.



The original triangle with sides $x, y, r$ is similar to the triangle with sides $r, a, OP$, and both are similar to the triangle with sides $b, r, OQ$.

$$x, y, r \sim r, a, OP, \sim b, r, OQ$$

By similar $\triangle$

$$\frac{a}{r} = \frac{y}{x} = \tan\theta$$

But $r = 1$ so

$$a = \tan\theta$$

If you imagine a point moving around the circle $a$ will get very large as $\theta \to \pi/2$, and in fact, approaches $\infty$ there (becomes undefined).

The segment $OP$ is (by similar $\triangle$) to $r$ as

$$\frac{OP}{r} = \frac{r}{x}$$

$$OP = \frac{1}{\cos\theta} = \sec\theta$$

The horizontal from the y-axis to Q is $b$. Consider $\theta$ near the top of the figure. By similar $\triangle$, the relations we had were

$$r/b = y/x = \tan\theta$$

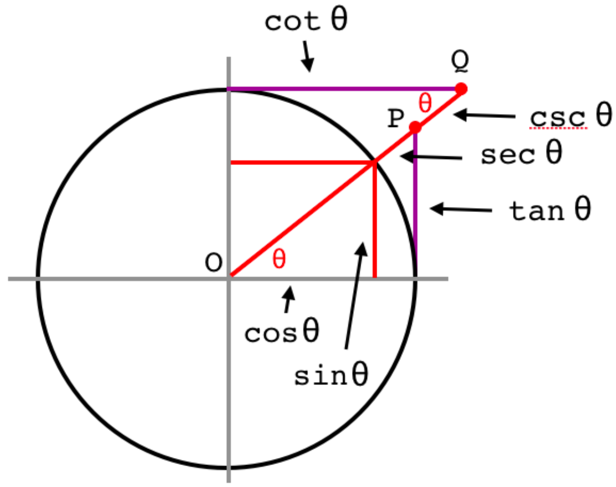
since $r = 1$

$$b = \frac{r}{\tan\theta} = \frac{1}{\tan\theta} = \cot\theta$$

Finally

$$r/OQ = 1/OQ = \sin\theta$$

$$OQ = \frac{1}{\sin\theta} = \csc\theta$$


cot θ
Q
θ
P
csc θ
sec θ
tan θ
O
θ
cos θ
sin θ

# Chapter 24

# Sum of angles

### cosine of a sum

The sum of angle formulas (i.e. formulas for the sine and cosine of the sum or difference of two angles) are used often in calculus, not only for working problems, but even in finding an expression for the "derivative" of sine and cosine.

You really must know them. I think it's so important that we will show three ways of finding these formulas — not all in this chapter. The easiest way to remember them uses Euler's equation, and we won't be ready for that until later. See **here**.

There are four equations: $\sin s \pm t$ and $\cos s \pm t$.

I've memorized only this one:

$$\cos s - t = \cos s \cos t + \sin s \sin t$$

By $\cos s - t$ we mean $\cos(s - t)$, but have left off the parentheses.

Say "cos cos" and then recall the difference in sign.

**check**

I like this version because it can be checked easily. Set $s = t$:

$$\cos s - t = \cos 0 = 1 = \cos^2 s + \sin^2 s$$

which is our favorite trigonometric identity and obviously correct.

**change signs**

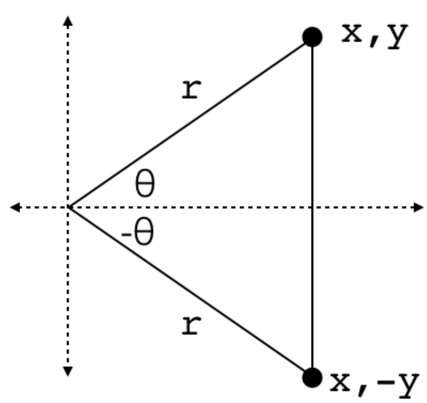
For $\cos s + t$ flip the sign on the second term.

$$\cos s + t = \cos s \cos t - \sin s \sin t$$

This is simply a result of the fact that

$$\cos -\theta = \cos \theta$$

$$\sin -\theta = -\sin \theta$$




The diagram shows the reason: $\cos \theta = \cos -\theta = x/r$ while $\sin \theta = y/r = -(\sin -\theta) = -(-y/r)$.

Proof:

$$\cos(s - (-u)) = \cos s \cos(-u) + \sin s \sin(-u)$$

Since $\cos -x = \cos x$ and $\sin -x = -\sin x$:

$$\cos(s + u) = \cos s \cos u - \sin s \sin u$$

But $u$ is just a dummy variable (it could be any symbol), so

$$\cos(s + t) = \cos s \cos t - \sin s \sin t$$

**sine of a sum**

We will look at the proof for the sine formula later, for now just write it:

$$\sin s + t = \sin s \cos t + \sin t \cos s$$

Say "sin cos" and then, that here + goes with +. Like most things having to do with sine and cosine, there is a change of sign when moving from one to the other.

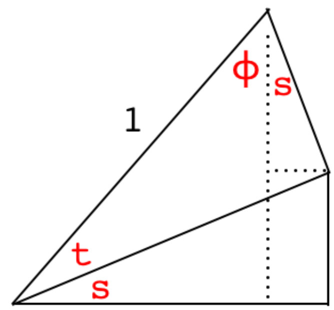
For $\sin s - t$, flip the sign on the second term, as before.

**proof**

Here is a geometric proof of both of the sum of angles formulas, using similar triangles. The key is to draw an inspired diagram.

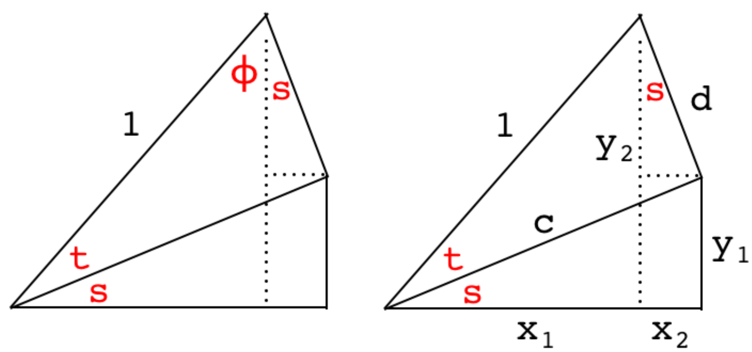
Consider a right triangle, with one of the angles labeled $s$. Construct another right triangle containing angle $t$, and scale it so that the base adjacent to angle $t$ is just as long as the hypotenuse of the triangle containing angle $s$, and draw them one on top of the other as shown:

Scale the joined triangles so that the hypotenuse of the second triangle has unit length. Our crucial insight is to draw vertical and horizontal dotted lines as shown below.




The angle $s$ is part of a right triangle with angle $t$ adjacent, where the third acute angle is $\phi$. But $\phi$ is also part of a second right triangle containing $t$ plus the angle adjacent to $\phi$. Therefore, that adjacent angle is also equal to angle $s$.

We add some labels to the sides of the triangles and calculate the sine and cosine of $s$, $t$ and $s + t$:




Since I already know the result I am looking for, I write what we had before

$$\cos s \cos t - \sin s \sin t$$

From the figure

$$\cos s = \frac{x_1 + x_2}{c}; \quad \cos t = \frac{c}{1}; \quad \cos s \cos t = x_1 + x_2$$

The sine of $s$ is a little trickier, look at the small right triangle at the top of the figure

$$\sin s = \frac{x_2}{d}; \quad \sin t = \frac{d}{1}; \quad \sin s \sin t = x_2$$

The difference is

$$\cos s \cos t - \sin s \sin t = x_1$$

but from the diagram it's clear that

$$\cos s + t = x_1$$

□

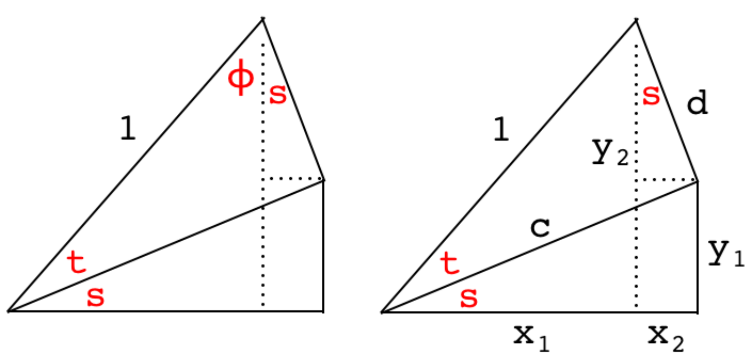
As a quick check we can ask what happens to the formula

$$\cos s + t = \cos s \cos t - \sin s \sin t$$

when $t = 0$. Then the first term is the cosine of $s$, and the second term is equal to 0. The formula is symmetrical with respect to $s$ and $t$.

**extension to sine**

Referring back to the diagram (and again, with our goal clearly in mind)




$$\sin s = \frac{y_1}{c}; \quad \cos t = \frac{c}{1}; \quad \sin s \cos t = y_1$$

$$\sin t = \frac{d}{1}; \quad \cos s = \frac{y_2}{d}; \quad \sin t \cos s = y_2$$

But

$$\sin s + t = y_1 + y_2 = \sin s \cos t + \sin t \cos s$$

Using the even/odd function rules, we get

$$\sin s - t = c + d = \sin s \cos t - \sin t \cos s$$

And that's all four of them.

**another calculation**

We found previously that

$$\sin \frac{\pi}{4} = \cos \frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

$$\sin \frac{\pi}{6} = \cos \frac{\pi}{3} = \frac{1}{2}; \quad \sin \frac{\pi}{3} = \cos \frac{\pi}{6} = \frac{\sqrt{3}}{2}$$

These angles correspond to 30, 45 and 60 degrees. It might be nice to have sine and cosine of 15 and 75 degrees as well. That would make even divisions of the first 90 degrees. We can get them as the sum and difference of $\pi/4$ and $\pi/6$.

Let $s = \pi/4$ and $t = \pi/6$. Then

$$\sin \frac{\pi}{12} = \sin s - t = \sin s \cos t - \sin t \cos s$$

$$= \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{3}}{2} - \frac{1}{2} \cdot \frac{1}{\sqrt{2}} = \frac{\sqrt{3} - 1}{2\sqrt{2}}$$

$$\cos \frac{\pi}{12} = \cos s - t = \cos s \cos t + \sin s \sin t$$

$$= \frac{\sqrt{3}}{2} \cdot \frac{1}{\sqrt{2}} - \frac{1}{2} \cdot \frac{1}{\sqrt{2}} = \frac{\sqrt{3}+1}{2\sqrt{2}}$$

We just check that $\sin^2\theta + \cos^2\theta = 1$:

$$\frac{(\sqrt{3}-1)^2 + (\sqrt{3}+1)^2}{(2\sqrt{2})^2}$$

$$= \frac{3 - 2\sqrt{3} + 1 + 3 + 2\sqrt{3} + 1}{8} = 1$$

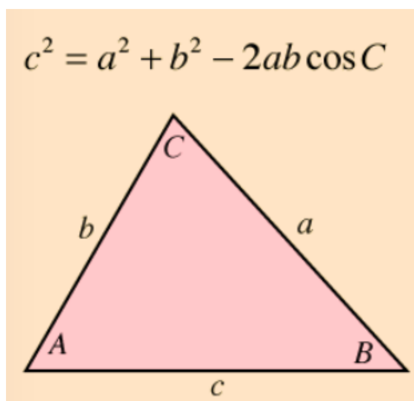
We can calculate similarly for $s + t = 5\pi/12$ or just switch sine and cosine from $\pi/12$.

# Chapter 25

# Law of cosines

## Law of cosines

Designate the lengths of a triangle's sides as $a, b, c$ and the angle between sides $a$ and $b$ as $C$ (because it is opposite side $c$). The law of cosines says that



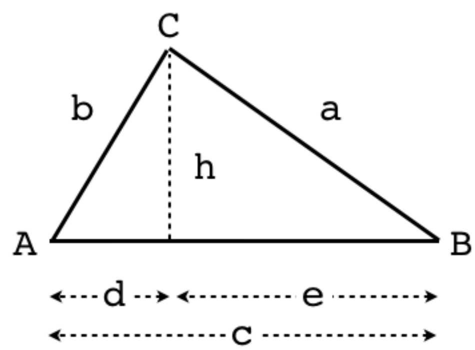

$$c^2 = a^2 + b^2 - 2ab\cos C$$

Lockhart calls this the "generalized" Pythagorean theorem. We can view the term $-2ab\cos C$ as a correction term which disappears in the case where $\angle C$ is 90 degrees.

## derivation

The result follows from the Pythagorean Theorem. (In fact, we can reuse the same diagram that was shown for the algebraic proof of the theorem).

For a triangle with sides $a$, $b$ and $c$ and angles opposite those sides $A$, $B$ and $C$, divide the third side into two lengths $c = d + e$ using the vertical altitude from vertex $C$.




$$a^2 - e^2 = h^2 = b^2 - d^2$$

So

$$a^2 = e^2 + b^2 - d^2$$

Since $d = c - e$ and thus $d^2 = c^2 - 2ce + e^2$:

$$a^2 = e^2 + b^2 - (c^2 - 2ce + e^2)$$

$$= b^2 - c^2 + 2ce$$

but $e = a \cos B$ so

$$a^2 = b^2 - c^2 + 2ac \cos B$$

rearrange to give a more familiar form (this is the law of cosines)

$$b^2 = a^2 + c^2 - 2ac\cos B$$

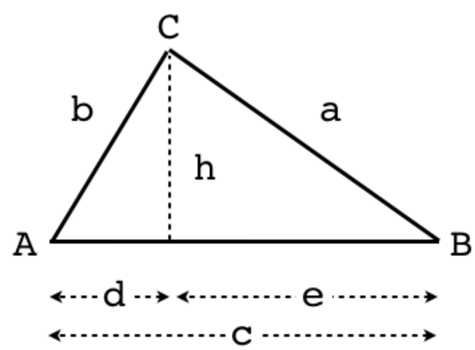
Any side of a triangle can be expressed in terms of the other two and the cosine of the angle between them. Thus, for example

$$c^2 = a^2 + b^2 - 2ab\cos C$$

$$a^2 = b^2 + c^2 - 2bc\cos A$$

## Law of sines

I'll just mention that there is another law called the law of sines. In contrast to the law of cosines, it is fairly trivial.




$$\frac{h}{b} = \sin A \qquad \frac{h}{a} = \sin B$$

Therefore

$$h = b\sin A = a\sin B$$
$$\frac{\sin A}{a} = \frac{\sin B}{b}$$

We could do the same construction and argument with $A$ and $C$ or $B$ and $C$. Therefore

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$