Why mathematics needed to be arithmetised.

Part 1

On the classical attitude towards numbers, arithmetic, geometry, constructibility and infinity

Chris Fenwick

Uploaded on 09/04/2026

fenwick_chris@hotmail.com, uczlcfe@ucl.ac.uk

Table of content

1 Introduction

1.1 A radical shift in the approach to mathematics from school maths to university maths

> "The shift from the natural mathematics in school to formal mathematics in university continues to present formidable obstacles to the learner whose experience is based on symbolic computation in arithmetic and algebra and visual perception in geometry, blended together in a natural approach to the calculus." [48]

Suppose you are an A-level student who has just sat his exams in June. You are going to start a mathematics degree in October, just four months away. You know everything about the mathematics you have been taught. For example, you know about functions such as e^x, $\ln x$ and $\sin x$. You know these to be curves and you know how to graph them. You know what happens to them as x approaches plus infinity or minus infinity simply by looking at the behaviour of the curves at the extreme ends of the x-axes.

In terms of the definition of trig functions you know these to be the ratio of relevant sides of a right-triangle. As for of e^x some of you may have seen the expression $\lim_{n\to\infty}(1 + x/n)^{n/x}$ but you consider this to simply be an alternative way of writing e^x. And as for $\ln x$, you see this as the inverse of e^x so there is no need to define $\ln x$.

But what if I told you that this is not the case. What if I told you that you don't know the real definition of the trig functions? What if I told you that there is indeed an actual definition for e^x, and $\ln x$? And this is where a transition is about to take place in your understanding of the nature of mathematics as you start your degree. The truth is that the trig functions are not defined as you have been taught at school. For example, one definition (of several) for $\sin x$ is

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

Also, e^x and $\ln x$ have their own definitions:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

$$\ln x = \int_1^t \frac{1}{t}\, dt$$

Even π is defined such that this gives the number we know of:

$$\pi = 2\int_{-1}^{1} \sqrt{1 - x^2}\; dx.$$

The first two equations you may have seen before. They are the Maclaurin expansions of $\sin x$ and e^x respectively.

Now, you may wonder what the difference is between saying $\sin x =$ opposite/hypotenuse and saying $\sin x$ is given/defined by its Maclaurin series, or between saying $\pi =$ circumference ÷ diameter and saying that it is the integral above. It is in these differing definitions that the transition between school mathematics and university mathematics lies.

For example, the right-hand side of the equation involving e^x is not simply a different way of "writing" e^x. The right-hand side actually *defines* e^x. This means that e^x *exists* or is *constructed* as a function on the basis of the arithmetic operations of the right-hand side, and not because e^x happens to be "e^x". The same applies to the last two equations. The fact is that $\ln x$ and π are defined by the right-hand side of these equations. And more than this, there are multiple versions of the definition for e^x, $\sin x$, and $\ln x$.

Returning to what you were taught at school, you know about the concept of differentiation and integration. But what if I told you that, although you know the definition of the derivative and the integral to be

$$\frac{df}{dx} = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

and

$$\int_a^b f(x)\ dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i)\ \delta x$$

you don't know the basis for these definitions. Your current understanding of the concept of differentiation is that it relates to sequences of secants through two points on a curve, such that these two points become infinitely close so as to ultimately produce a tangent at one point. As for integration, you know this to be related to the sum of areas of an infinite number of infinitely thin rectangles under a curve.

But in the world of rigorous mathematics (which is the world you are now entering) the way of thinking about the concept of differentiation and integration as being based on secants, curves, tangent lines, rectangles, and the idea of "infinitely close" is no longer used. It hasn't been used for at least 130 years. This type of thinking (the type you were taught at school) is what we call geometric thinking. However, modern mathematics (the type taught at university and used at a professional level) is founded on arithmetic. Numbers and arithmetic.

And this transition from a geometric approach to mathematics to an arithmetic approach is substantial and even radical. It is completely different to the way in which you thought about mathematics at school. And in moving from school to university you are moving from an environment where you were taught predominantly by mathematics teachers to an environment where you are taught by professional mathematicians.

For sure there will be topics in your undergraduate course that you will be familiar with. These will simply be extensions of maths topics you learnt at school. For example, you will learn about partial differentiation and multiple integration. You will learn about line integrals and surface integrals. You will learn about ordinary differential equations and partial differential equation. Although these topics develop your school calculus much further, the mathematical thinking required of these is the same, or similar enough to, that which you developed at school. As such they should pose little problem in terms of the required mindset.

But you will also meet totally new topics that you won't have studied before (unless you have done some private studies on these). Topics such as abstract algebra, group theory, number theory, and the killer of them all for a lot of 1st year students: real analysis. It is not just that you won't have seen these topics before. It is that they require a radical shift in the way you think mathematically. A shift from a geometric perspective to an arithmetic/axiomatic perspective. This shift is not trivial and is not easy because none of you have been taught this latter way of thinking, but you are suddenly thrown in at the deep end four months after doing thirteen years of doing school mathematics.

This radical shift from a geometric perspective to an arithmetic perspective took place over approximately 400 years from the 16th century, but particularly during the 1800s and into the early 1900s. Why did this shift happen? The reason lay in the inability of geometry to provide rigorous justification to ever developing and deepening understanding of mathematics. For example,

- Pythagoras (570BC – 490BC) and his followers thought that $\sqrt{2}$ existed only geometrically, namely as a diagonal of a unit square. It certainly did not exist as a number. This was because the Pythagoreans only believed in integers and ratios of integers. Since $\sqrt{2}$ could not (and cannot) be written as a ratio of a rational fraction, p/q where p and q are integers, they considered $\sqrt{2}$ as a number to be irrational. The diagonal line of a square clearly has a beginning and an end. In other words, this line is finite (just as integers are finite). But $\sqrt{2}$ has an infinite number of decimal places and can therefore not be expressed as a number in a finite way. Yet today modern maths classifies $\sqrt{2}$ (and all irrationals) as numbers, and we ourselves have no problem accepting $\sqrt{2}$ as a number (although there are great subtleties in thinking about $\sqrt{2}$ as a number);

- Also, in the Europe of the 16th and 17th centuries mathematicians thought negative numbers to be false or absurd, and not possible as numbers even though such numbers occurred as solution to equations. Similarly, mathematicians didn't believe that $\sqrt{-1}$ was a number, and therefore called it imaginary. Yet today modern maths classifies things like -1 and $\sqrt{-1}$ as numbers, and we ourselves have no problem accepting negative numbers and imaginary number (even if it seems weird to call $\sqrt{-1}$ a number);
- In terms of arithmetic, and up until the 17th century, everybody thought that the rules of arithmetic worked in all situations. But once Newton had discovered the binomial theorem there (seemingly) arose fundamental problems with arithmetic. For example, it was/is possible to write $(1-x)^{-1} = 1 + x + x^2 + x^3 + \cdots$, but on substituting $x = -1$ into this people obtained ½ = 1 + 1 + 1 + ... In other words, ½ = ∞. Similarly, substituting $x = 2$ gave $-1 = 1 + 2 + 4 + 8 + \cdots \to \infty$. On the other hand, substituting x = ½ gave $2 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$ which is indeed correct. So it seemed as if arithmetic failed in certain circumstances.
- Furthermore, once algebra had became firmly established (around the 17th century) as an accepted form of mathematical analysis, everybody thought that $a \times b = b \times a$ was always true. And, this is indeed true but only if a and b are real numbers. Since these were the only numbers known to exist at the time, it makes sense that mathematicians would hold such a belief. But by the 19th century new numbers were being created where this commutativity was not true. And even with real numbers there were cases where new mathematics was being developed such that this commutativity did not hold. For example, in terms of matrices we have

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 4 & 2 \end{pmatrix} \neq \begin{pmatrix} -1 & 0 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

And in terms of vector we have $\boldsymbol{a} \times \boldsymbol{b} \neq \boldsymbol{b} \times \boldsymbol{a}$, i.e. vector cross products are generally not commutative.

This lack of understanding of how numbers and arithmetic worked was due to a lack of definition and structure to numbers and arithmetic. It took from the time of Pythagoras (or even before, since the Egyptians and Babylonians were skilled at computation) to the 19th century and early 20th century for mathematicians to get a coherent handle on numbers and arithmetic, and for acceptable defintions to be developed.

The moral of the story is that, only after thousands of years of the study and usage of numbers, arithmetic and geometry have we come to construct mathematics fundamentally as arithmetic, not geometric. All axioms and definitions of modern (arithmetic) mathematics have come at the end of a process of study that took 2500+ years.

What was the key issue which forced a radical rethink of the nature of mathematics? Infinitesimals and infinities, particularly as these showed themselves in tangent and area problems (i.e. in what we now call differentiation and integration). These two areas of mathematics can be traced back to the ancient Greek mathematics of the sixth century B.C. but became prominent from the time of Newton (1642/1643 – 1727) and Leibniz (1646 – 1716) onwards. The problems and pitfalls encountered by mathematicians from then on proved to be the motivating factor which contributed to what we now call the arithmetisation of mathematics. In these notes we will look at certain key people and episodes which had a great influence in highlighting and resolving these problems.

2 The primacy of number and arithmetic: The Pythagoreans

2.1 An initial comment

Historians of mathematics record (at least) six major geographical regions, known as centers, where mathematics flourished over the last 4000 years. These are the Babylonian and Egyptian center, the Greek center, the Chinese center, the Indian center, the Arabian center, and finally the European center.

The European center of mathematics, which rose from about the 15th century, was influenced by the mathematics coming from Arabia, this itself being was influenced partly by Indian mathematics but mainly by Greek mathematics. Although the period of Babylonian and Egyptian mathematics (which came before the Greek period) did contain work on numbers and arithmetic, the result of the migration of mathematics from Arabia meant that the Europeans took on board the attitudes and approaches of the Greeks, namely that mathematics was all about geometry. Problems were set up geometrically and were solved using geometric methods (such as by comparison of angles and lines, similar triangles, trigonometry, circle theorems, etc.). As such, the mindset of European mathematicians of the 15th and 16th centuries was fundamentally geometric. And even though decimal numbers, arithmetic and algebra had become mainstream from the 17th century onwards, geometric thinking and assumptions remained core to mathematical thinking up until the early 19th century. However, problems in geometric reason had become apparent from the 16th century onwards, leading to a concerted effort by a number of mathematicians during the 19th century to base mathematics on number and arithmetic only.

It should be noted that there is some controversy among historians of mathematics as to who invented what, and when. It has been said that new generations of historians perpetuate false attributions by referencing previous historians who made false attributions. For example,

- Pythagoras' theorem was not discovered by Pythagoras (circa 570 B.C. – 495 circa B.C.) since this was already known to the Babylonians at least 1000 years before him;
- Pythagoras is unlikely to have discovered the irrationality of $\sqrt{2}$. Some say it was Hipassus of Metapontum (circa 530 B.C. – 450 circa B.C.) who was part of the school created by Pythagoras.
- L'Hospital's rule is named after Guillaume de L'Hospital (1661 – 1704) but was actually discovered by Johann Bernoulli (1667 – 1748);
- Maclaurin series is named after Colin Maclaurin (1698 – 1746), but he never claimed to have discovered it;
- The Argand diagram for representing complex numbers in a coordinate system was first invented by Wessel (1745–1818) before Argand discovered it himself.

I am not a historian of mathematics so all I can do is to recount what happened based on those historians whom I believe have properly studied the primary sources. I leave it to you to investigate this further (hence the bibliography at the end of this essay).

2.2 The general attitude of the Pythagoreans towards nummbers and arithmetic

During the ancient Greek period of mathematics Pythagoras created a school whose curriculum consisted of numbers and arithmetic, geometry, music and astronomy. The people who adopted

Pythagoras' view of the world, and of mathematics, were known as Pythagoreans and this movement covered the periods of the sixth to fourth century B.C. as well as the first to third century A.D.

There is much debate among scholars to this day about how much of the mathematics attributed to Pythagoras was actually developed by him and how much was developed by his followers. Subsequent authors may have projected their own views of what constituted Pythagorean thought into their history of Pythagoras and his followers. Hence, what follows is a summary of what is considered to be the accepted approach to numbers and arithmetic taken by Pythagoreans. More can be found in [53] and [89].

Over the course of the sixth, fifth and fourth centuries B.C. Pythagoras and his followers came to believe that everything in the universe could be reduced to numbers and the study of numerical relationships. But more than this they believed that the only numbers which existed were what we today call the natural number ($\mathbb{N}$). The primacy of natural numbers over everything else was based on the mystical belief that numbers could not only describe nature but also life and social affairs. As W. R. Knorr [53] says "The Pythagoreans took numbers to be their universal principle". For example, in life and social affairs, the number 1 represented unity and the origin of all things, the number 2 represented the feminine, and the number 3 represented the masculine. The number 3 also represented the "whole" because it could be separated into three units representing "beginning–middle–end". The number 4 represented justice. The number 5 represented marriage because it was the result of summing what was considered as the first two numbers: $2 + 3 = 5$, and so on.

As another example, it was found that if the string of a musical instrument was plucked to give a pleasing sound then shortening or lengthening the string by a certain amount would give another pleasing sound only when the shortening or lengthening was a ratio of integers. Hence, numbers came to be seen as the way to analyse musical notes. Finally, as we shall see in section 2.6, the Pythagoreans found that number, arranged in certain patterns, could also describe geometry. From this perspective numbers came prior to geometry (as well as astronomy since, in two dimension this was represented geometrically, as when plotting stars and constellations on paper). The Pythagoreans therefore believed that the world was fundamentally composed of numbers, and the belief arose that numbers were fundamental to all mathematics.

In terms of mathematics, most historians agree that Pythagorean mathematics was fundamentally arithmetic. Although they did study elements of geometry this was only on the basis that arithmetic problems could be represented geometrically. The classic example of this the figurate numbers, i.e. numbers which can be arranged in such a way as to represented geometric shapes (see section 2.6). The Pythagoreans saw the form and structure of geometric shapes (mainly polygons) as being created by numbers. They would then use these geometric configurations as a tool for performing arithmetic analysis (for more on this see [78]). "Pythagoras brought geometry to perfection … most of all he devoted himself to the arithmetic form of it" ([53], p133).

2.3 Defining "1" and numbers

Article 1: In the Pythagorean school of thought "1" represented the origin of all things in the sense that every other number could be constructed from it, for example: $1 + 1 = 2$, $1 + 1 + 1 = 3$, $1 + 1 + 1 + 1 = 4$, etc. Because of this they interpreted "1" not as a number but as the principal element or originator of the numbers 2, 3, 4, 5, etc. Also, since negative numbers did not yet exist all numbers could be made to decrease by subtraction except for "1", hence "1" not being consider as a number.

Article 2: The Pythagoreans had multiple definition for "1". As we have seen above, one definition was that "1" was not a number. This from Nichomacus of Gerasa (~100 AD) who was a proponent of the ancient Greek mathematics,

> "Unity, then, occupying the place and character of a point, will be the beginning of intervals and of numbers, but not itself an interval or a number, just as the point is the beginning of a line, or an interval, but is not itself a line or an interval." (Nicomachus[Nic52], II.6.3, p. 832)

Another thing which defined "1" as being distinct from the numbers was that it was immutable under multiplication. In other words, $1 \times 1 = 1$, $1 \times 1 \times 1 = 1$, etc. No amount of multiplying 1 by itself would ever change it, or make it become a number (today we say that 1 is the multiplicative identity).

Yet more definitions of "1" were that it was

- the least number (remembering that all numbers were only natural numbers);

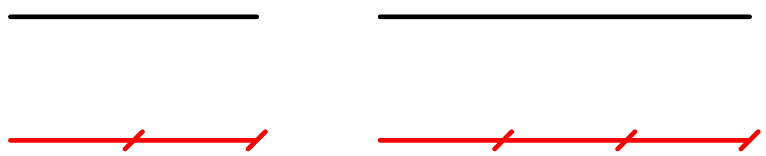
- that from which all other numbers came, i.e. it was the originator of all other numbers. Hence "1" itself could not be a number;
- that thing which is common to all numbers 2, 3, 4, 5….

As we can see the Pythagoreans were attempting to understand the nature of numbers, and doing so from a metaphysical perspective. The above illustrates their attempt at bringing a logic to the nature of "1". As we shall see momentarily the Pythagoreans also attempted, in their own way, to bring a logic and order to the numbers, 2, 3, 4, …

Article 3: Moving on to the numbers themselves, The Pythagoreans had various definitions of these over the ages. So, numbers (i.e. 2, 3, 4, …) were seen as

- a collection of units, i.e. 1s;
- a progression from 1 through all the multiples of 1, where 1 was not seen as a number since it generated all the other numbers which came after it;
- other definitions being of a purely mystical/spiritual nature (see section 2.1 above)

Article 4: Furthermore, the Pythagoreans conceived of numbers as being able to be split into discrete parts, for example, $6 = 3 + 3 = (2 + 1) + (2 + 1) = (1 + 1 + 1) + (1 + 1 + 1)$, or $6 = 4 + 2 = (1 + 1 + 1 + 1) + (1 + 1)$, etc. Since the ancient Greeks only had ratios of (integer) numbers they could only form things like $2 : 3$ or $5 : 8$. These ratios were not fractions as we understand this today. Instead, 2:3 meant two wholes of something (such as a line) compared to three wholes of something else (such as another line). This is illustrated below where the two black lines can be compared as being in the ratio 2:3 as shown by the red lines.



So for the Pythagoreans ratios were literally just a comparison of integer numbers. Since "1" had no integer parts of its own it could not be arithmetically split in the way numbers could, so "1" was not a number.

2.4 Defining even and odd numbers

Article 5: The Pythagoreans studied what could be called the structure and properties of numbers, the first property of which was the classification of numbers as either even or odd. They used this as their starting point for the study of numbers. Other classifications included the distinction between limited and unlimited, and one and many. They even put these three classification in order of priority: i) Limited – unlimited, ii) Even – odd, iii) One – many. Here we see the earliest attempts at systematically and rigorously (for the period) defining types of numbers.

In terms of mathematics, "[the] elements of number [are] the elements of all things ... and the elements of number [are] the odd and the even". And in terms of arithmetic, "What is the subject matter of the art of arithmetic? It is the odd and the even, regardless of the quantity of either." (both quotes from [53], p134).

Article 6: The standard definition of an even number was a number capable of being split into two equal parts without needing to be separated by a 1 in between. On the other hand an odd number was a number not capable of being split into two equal parts without needing to be separated by a 1 in between.

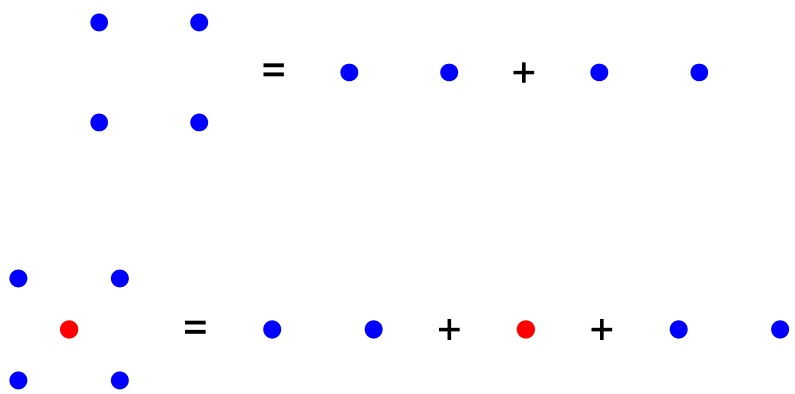
For example,

$$4 = 2 + 2 \text{ but } 5 = 2 + 1 + 2,$$

or

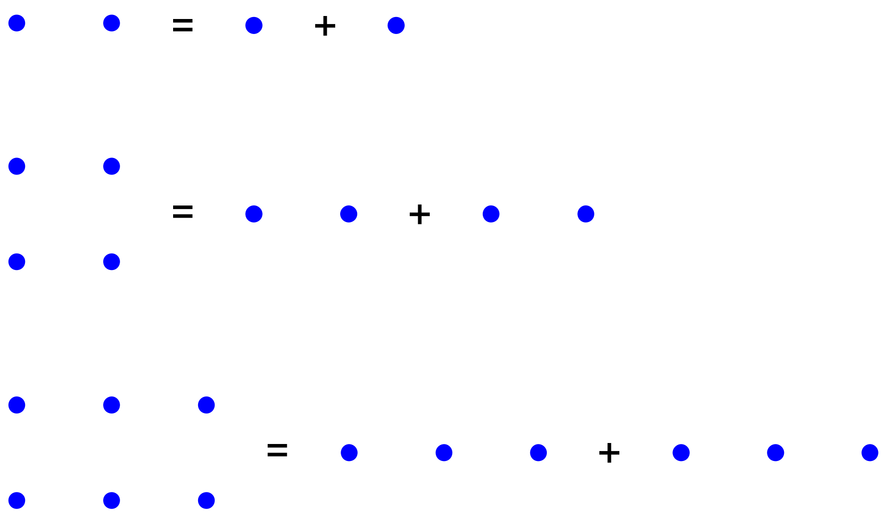
$$6 = 3 + 3 \text{ but } 7 = 3 + 1 + 3.$$

Such a distinction between even and odd numbers can also be represented geometrically. The Pythagoreans used pebbles on the sand to represent numbers, so that when time came to write on paper they used dots to represents the pebbles. As such they represented even and odd numbers as follows (where I have added the colour as a distinguishing feature):



However, the Pythagoreans themselves defined even and odd numbers differently. They defined even numbers in terms of the *structure* one could obtain from splitting such a number into two parts. For example, 6 is an even number because it can be split into two parts of the same size, namely 6 = 3 + 3. Certainly 6 can be written as 2 + 2 + 2 but the Pythagoreans were interested in splitting numbers into only two parts. Thus they described the feature of an even number as being even by saying that no other split into two parts produced a number with the greatest space (because pebbles take up space in the sand) whilst at the same time being the least in quantity (only two groups of three pebbles). This is equivalent to saying that numbers are even when such numbers, upon dividing by 2, have the largest quotient as well as having the least number of these quotients. By this definition, no part is greater than half of the original number 6, and there is no other split which produces a quotient greater than 3 when dividing by 2.

The diagram below shows that the numbers 2, 4, and 6 are even because they can be split into two parts having the greatest space and least quantity. This is basically a minimax problem: to find the minimum number of terms of greatest value.



However, this is not true for an odd number. For example, $5 = 2 + 3 = 1 + 4$. Here we have two parts, but for each of the two types of splits one of the parts is greater than half of the original number 5. In order to have parts which would be less than half of 5 we would have to write $5 = 2 + 2 + 1$ and now we have three parts to the numbers, but this is not the minimum of parts to the number 5.

Further properties of even and odd numbers were derived from such a way of thinking. For example,

- an even number could be split into two even parts or two odd parts but not an even and odd part, whereas an odd number could only be split into an even and odd part, and

- an even number differs by 1 from an odd number. Similarly, an odd number differs by 1 from an even number.

2.5 Further studies in evenness and oddness

Article 7: The Pythagoreans continued their study of numbers by further categorising even numbers, the first of which was *evenly-even* numbers. Such numbers were those which could continually be split into two parts without 1 intervening in the middle, as illustrated below.

$$2 = 1 + 1 \qquad\qquad 4 = 2 + 2 = (1 + 1) + (1 + 1)$$

$$8 = 4 + 4 = (2 + 2) + (2 + 2) = (1 + 1) + (1 + 1) + (1 + 1) + (1 + 1)$$

etc. So, evenly-even numbers are such that they can be continually divided by 2 all the way down until we reach 1 as a factor. In general, (and in order to contrast with unevenly-even numbers later on) one can say that evenly-even numbers are generated by multiplying one or more even numbers by a power of 2: 2×2^n, 4×2^n etc.

Article 8: On the other hand the *evenly-odd numbers* can only once be split into two equal parts. In modern terminology, evenly-odd numbers only have one factor of 2. For example,

$$6 = 3 + 3$$

or

$$18 = 9 + 9.$$

Other evenly-odd numbers include 10, 14, 18, 22. Further analysis of evenly-odd numbers revealed that such numbers would either have a part which would be an odd number if you were splitting into two (or other even) parts, or a part which would be an even number if you were splitting into three (or other odd) parts, viz

- If 18 is split into two parts then 18 = 9 + 9. So we have divided by an even number and we obtain at least one part (in this case two parts) which is an odd number.
- If we then divide 9 into three parts we have 9 = 3 + 4 + 2. So we have divided by an odd number of parts, and we obtain at least one part which is an even number.

Article 9: The third class of even numbers was the *unevenly-even* numbers. These are similar to evenly-even numbers in that they can be split into two equal parts on more than one occasion. But at some point we will obtain an number which cannot be split into two equal parts. For example, 12 is an unevenly-even number because 12 can be split into two equal parts, namely two 6s. The 6s can then be split into two equal parts, namely two 3s. But the 3s cannot be split into two equal parts. Numerically this is illustrated as

$$12 = 6 + 6 = 3 + 3 + 3 + 3$$

Similarly 20, 24, 28, etc. are unevenly-even numbers because we can break these numbers down multiple times into two equal parts but not all the way down to a factor of 1. Unevenly-even numbers therefore contain more than one factor of 2 but are not all powers of 2.

In modern terms unevenly-even numbers are those of the form 3×2^n, 5×2^n etc., or 9×2^n, 15×2^n, etc. (contrast this with evenly-even numbers). By the way in which these numbers are perceived, unevenly-even numbers can be said to be structured in a manner in between that of evenly-even numbers and evenly-odd numbers.

The classification of even numbers presented above illustrates the Pythagoreans approach to identifying the structure of even numbers in order to better understand their nature and behaviour. It is this desire to understand the nature and structure of an aspect of mathematics which lies at the heart of mathematics to this day. It is just that the Pythagoreans approached their analysis of numbers according to the knowledge and attitude of their day.

All of the above can today be subsumed into one theorem, namely the fundamental theorem of arithmetic, which says that any positive integer N can be uniquely factored into powers of primes, viz

$$N = 2^{n_1} \times 3^{n_2} \times 5^{n_3} \times \ldots$$

where $n_1, n_2, n_3, \ldots$ are positive integers.

Article 10: The same can be said about the way in which the Pythagoreans thought about odd numbers. These they also classified into three categories: first and incomposite, second and composite, and second & composite numbers which are first & incomposite with respect to another number.

The first and composite numbers might be said to be the numbers most odd, and these equate to the prime numbers 3, 5, 7, 11, 13, 17, … These are incomposite because they are not composed of any other parts (factors) save themselves and 1. It is because of this that they are considered the first among all odd numbers.

The second and composite numbers are those odd numbers which have factors other than 1 and themselves. For example, 9 is second-and-composite because it has 3 as a part (factor). It is because of this that such numbers are called composite, and they are called second (as opposed to first) because that have parts other than 1 and themselves. Similarly, for 15, 21, 25, etc.

Finally, the second & composite numbers which are first & incomposite with respect to another number are, in modern terms, numbers which are relatively prime. So, 9 and 25 are second & composite numbers, but are first & incomposite with respect to each other since they share no common parts with each other, i.e. 9 and 25 are relatively prime since they have no factors in common.

Article 11: The Pythagoreans also went on to study further classifications of numbers (for example, perfect numbers, amicable numbers, abundant numbers), the ordering of numbers (in terms of < and >), and number patterns and arithmetic. See [Thomas Taylor (1812)] for more.

2.6 On numbers represented geometrically: The figurate numbers

Article 12:

The Pythagoreans represented numbers by dots or points, this being the first and most important geometric representation of a number. These dots/points were designed to illustrate pictorially the physicality of the pebbles they used when manipulating numbers and doing arithmetic (were arithmetic was performed by moving pebbles around). Since each pebble was a single pebble which could not be divided into smaller pebbles, so each point was seen as an indivisible atom or unit.

Article 13: Just as “1” was the origin of all numbers the point was the origin of all other geometric objects. For example, the Pythagoreans believed that lines were created by the motion of a point (a belief also held by many other mathematician and philosophers from ancient times up to and beyond Newton (1642 – 1726/1727), and which formed the foundation for Newton’s creation of calculus). So, the point was considered the principal element from which lines could be drawn. The first geometric figure was therefore the dot, this representing “1” or unity, and the second geometric figure was the line, formed by the joining of two dots, as shown below. Here, the line represented the number “2” because the line had two extremities.

Similarly, the line was the principal element from which surfaces or plane figures could be drawn since it was the motion of a line which created such plane figures (a view also held by renaissance mathematicians). The numbers 1 to 4 were then identified with point, line, plane and solid, as illustrated below (taken from chapter 2, [75]).

Article 14: Even numbers could be represented in a figurate manner as being given by a sequence of dots which could be split in two equal parts without being separated by a dot. On the other hand odd figurate numbers were ones which would contain an extra, central, dot separating two equal parts, viz

Article 15: The Pythagoreans used polygonal figures to represent numbers and number patterns. So, just as "1" extended into numbers, the point extended into lines. In doing this the Pythagoreans had started the process of geometrising numbers by assigning a geometric figure to each number. This is the opposite of what Descartes and Fermat did with their numericalisation/algebraicalisation of geometry.

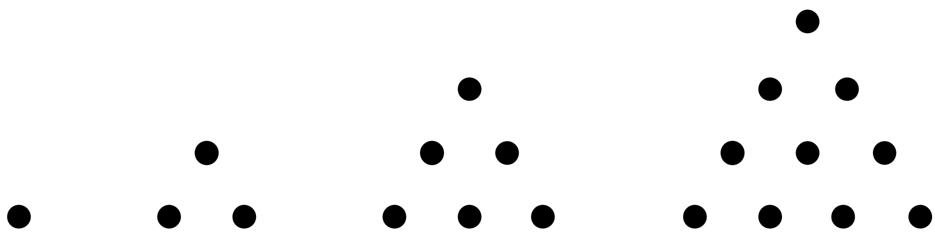
Article 16:

But the dot and the line were not the only geometric representations numbers. Other geometric figures such as triangles, squares, pentagons, etc., were used to represent the numbers 3, 4, 5, etc. Such figures are known as *figurate numbers*, and were constructed by arranging dots in certain ways so as to produce geometric figures. By doing this the Pythagoreans were able to study geometrically the essence of numbers.

The first set of figurate numbers was the triangular numbers, and the first triangular number was 3 since three dots could be arranged in the form of a triangle. The next triangular number was 5 for the same reason. These two figurate numbers are illustrated below.

Article 17:
We now have a problem, namely trying to classify "1" as a figurate number. Clearly "1" does not have a triangular shape, but the process of generating triangular numbers starts from "1". This, along with wanting consistency, made the Pythagoreans include "1" as the first triangular number.

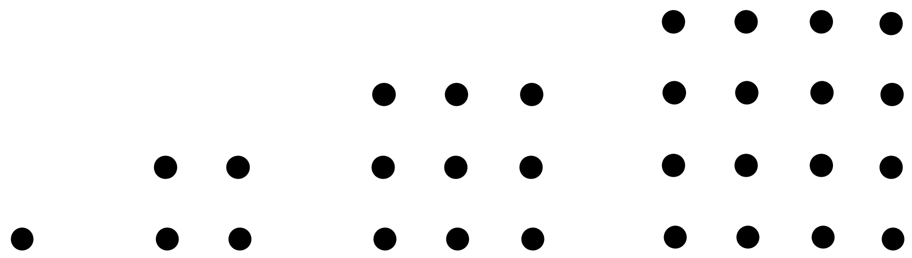
Therefore, starting from unity, we see that the triangular numbers are numerically created by repeatedly adding the next number along in the sequence of numbers. Geometrically speaking this translates to adding a new row of dots where each new row consists of an extra point compared to the previous row.



1   1 + 2 = 3   1 + 2 + 3 = 6   1 + 2 + 3 + 4 = 10

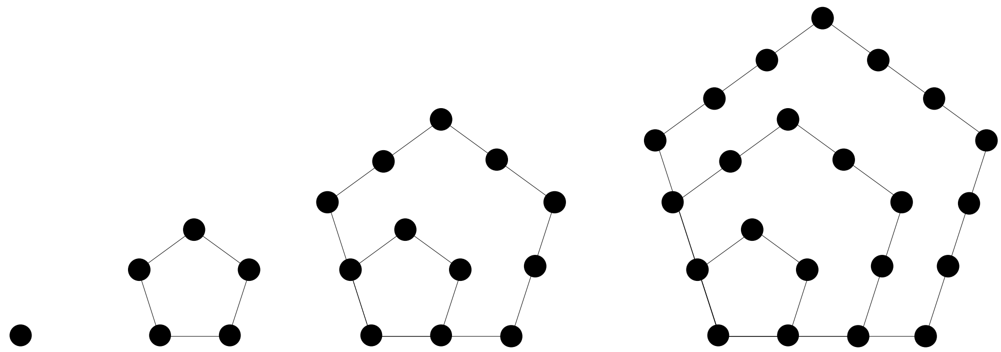
Similarly the subsequent triangular numbers are 15, 21, 28, etc. In analysing this geometric pattern the Pythagoreans found that triangular numbers formed the sum of the natural numbers.

Article 18: Number can also be used to represent other rectilinear geometric figures, such as square numbers, pentagonal numbers etc. For example, the numbers 1, 4, 9, etc., can be represented by squares as illustrated below.



1 1 + 3 = 4 1 + 3 + 5 = 9 1 + 3 + 5 +7 = 16

Thus the square numbers can be seen to represent the sum of odd integers. Dots can be similarly arranged as pentagons, hexagons, etc. thus giving rise to pentagonal numbers, hexagonal numbers, etc. An example of pentagonal numbers is illustrated below



1 1 + 4 = 5 1 + 4 + 7 = 12 1 + 4 + 7 + 10 = 22

where this can be seen as the sum of alternate odd and even numbers.

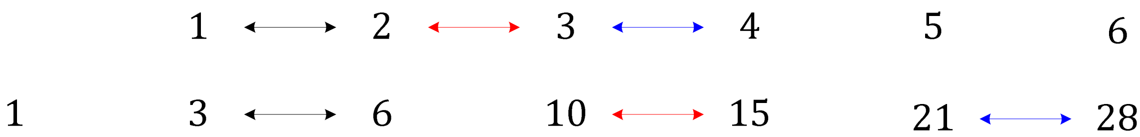
So, we see that, by their geometric representation of numbers, the Pythagoreans were able to analyse, create coherent and logical patterns of numbers, and classify these. This is an example of type of mathematical investigation the Pythagoreans were conducting into the nature and behaviour of numbers.

Article 19: Returning to the issue of classifying "1" as a figurate number, more confusion arises. For example, if the unit is a figurate number how do we distinguish it from being triangular, square, pentagonal, etc.? In all cases of figurate numbers the unit would have to be triangular and square and pentagonal etc., assuming, of course, that that "1" is a number. If it is not a number then what is the starting point/number for figurate numbers?

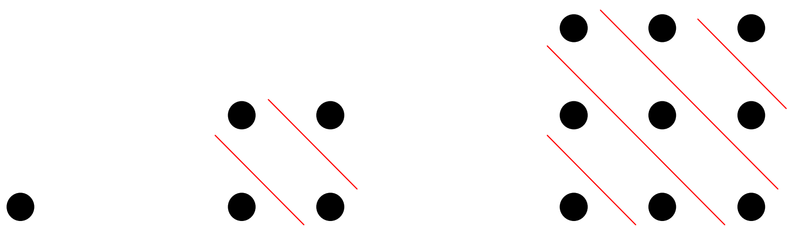
If it is a number then one particular way around the problem was suggested by Nicomachus (circa 60 A.D. – circa 120 A.D.) and Theon (circa 70 A.D. – circa 135 A.D.) who said that "1" was a number or figurate number "in potency", whereas all numbers and figurate numbers were numbers "in actuality". In other words, "1" had the potential to become a number and a figurate number, and this was realised in actuality by 2, 3, 4, 5 and by the triangular, square, pentagonal, etc., numbers.

Article 20: Numerical patterns can be found via the figurate numbers. For example,

- triangular numbers come in odd and even pairs: 1 and 3, then 6 and 10, then 15 and 21, then 28 and 36, etc.
- triangular numbers taken in pairs from a certain position have the same ratio as natural numbers taken in pairs from a certain position. The diagram below illustrates this with the natural numbers as the top row and the triangular numbers as the bottom row. Here we see that the ratio of numbers associated with black arrows is the same, the ratio of numbers associated with red arrows is the same (i.e. 2 : 3 has the same ratio as 10 : 15), the ratio of numbers associated with blue arrows is the same, etc.



Similarly for square numbers alternative number patterns can be found. One example is generated by dividing the square into diagonal sections, viz:



$1^2 = 1$ $\quad$ $2^2 = 1 + 2 + 1$ $\quad$ $3^2 = 1 + 2 + 3 + 2 + 1$

In general, we see that

$$n^2 = 1 + 2 + 3 + \cdots + (n-1) + n + (n-1) + \cdots + 3 + 2 + 1$$

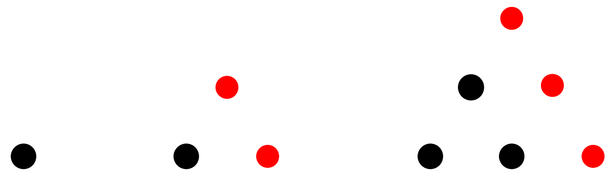
which leads to the usual formula for the sum of positive integers

$$\frac{1}{2}n(n-1) = 1 + 2 + 3 + \cdots + (n-1)$$

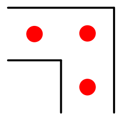
This sum was known to the Babylonians and, later, the Greeks (although none of them presented it in this form).

Article 21: All figurate number were constructed using something called a *gnomon*. This consisted of a set of points, appropriately configured, and designed to fit around the preceding figurate number so as to create the next figurate number in the sequence.

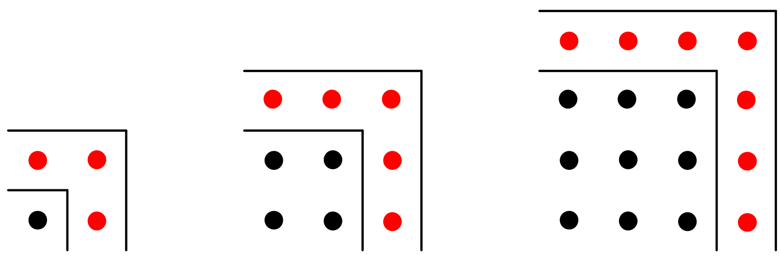
So, to construct the triangular number 3 from the triangular number 1, and then the triangular number 6 from the triangular number 3, one simply adds a diagonal of an appropriate number of dots, viz,



etc. As for the square numbers, they can be generated as follows: construct a gnomon consisting of a right-angled set of points. So, to the single point add the three points enclosed by right-angled lines as shown below:



This allows us to constructs the figurate number 4. We then add to these five points similarly enclosed between right-angled lines, and so on.



In terms of standard arithmetic we can describe this process as the sum of the series of odd numbers:

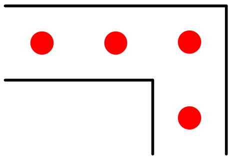
$$1,\ 1+3,\ 1+3+5,\ 1+3+5+7,$$

etc. In this case the added points are 3, 5, 7, etc., instead of 1, 2, 3,… for the triangular points.

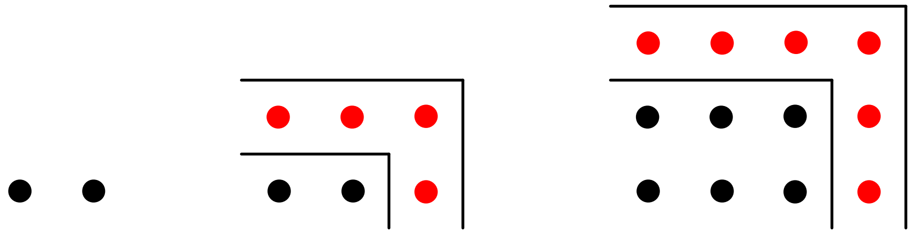
Article 22: From this we can see that the gnomons of triangular numbers successively increase by one unit, gnomons of square numbers successively increase by two unit, the gnomons of pentagonal numbers successively increase by three unit, etc. In fact, so important was the idea of a gnomon that it was actually made into a definition by one Pythagorean called Iamblichus (circa 242 A.D. – circa 325 A.D.)

> **Definition 1**: A *gnomon* is that number which, when added to a term in a given class of consecutive figured numbers, produces the next terms in that class. [p143,53]

Article 23: A next logical step is to generate rectangular numbers. Instead of starting with one point we start with two points. Then by the process of "enclosing by right-angled lines" we can set up many different rectangular numbers. For example, using the gnomon



and its extended versions, we can generate a sequence of rectangular figurate numbers such as



Arithmetically speaking the series generated is now 2, 6, 12, 20, 30, ...

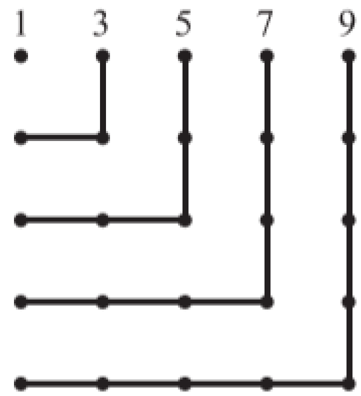
2.7 Number patterns and arithmetic on figurate numbers

Article 24:

The Pythagoreans showed that figurate numbers could be used for performing arithmetic. This is not to say that they replaced arithmetic on numbers by arithmetic on the geometric form of numbers. It is simply that, by arithmetic on figurate numbers, they were able to obtain particular number patterns and demonstrate the geometric transformations which gave such number patterns. Such arithmetic could be interpreted as a geometric arithmetic.

Article 25

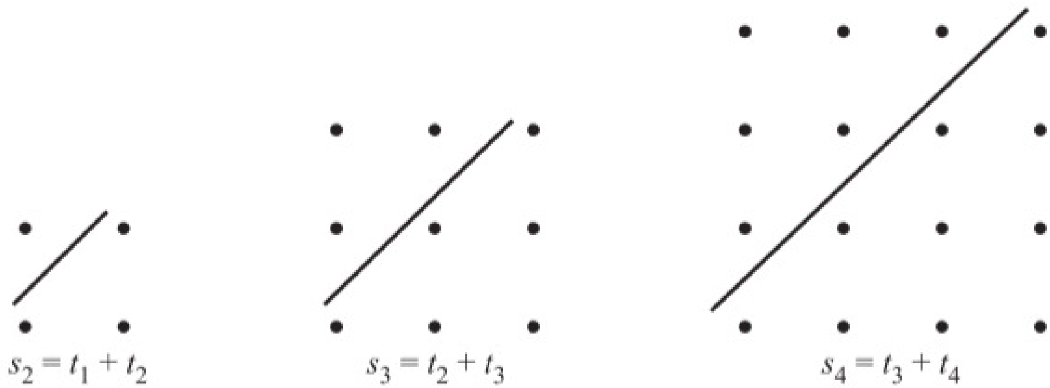
The first basic number pattern which can be identified from figurate number is one obtained from square numbers. The Pythagoreans knew about such a result but it is not known exactly how they derived it. One possible approach is to look at any square number as a collection of individual gnomons as illustrated below.



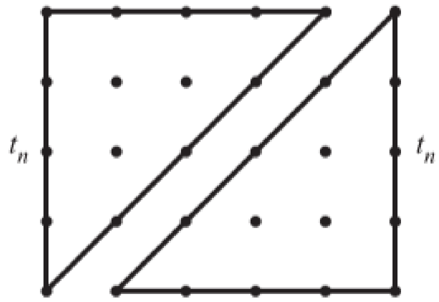
Here we have the square number 5 which can be seen to be composed of the sum of points on each gnomon: $1 + 3 + 5 + 7 + 9 = 5^2$.

Article 26

Another basic number pattern which can be identified from figurate number is the connection between triangular numbers and square numbers. By drawing a line appropriately across a square number we see that a square numbers is the sum of two consecutive triangular numbers. This is illustrated in the diagram below where s_2, s_3, s_4 are the second, third and fourth square numbers, and t_1, t_2, t_3, t_4 are the first, second, third and fourth triangular numbers.



Now that we have derived triangular numbers from square numbers, consider putting together two triangular number t_n. For the case where $n = 5$ we have



In modern notation we have $2t_n = n(n+1)$ implying

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

Since a square number is the sum of two consecutive triangular numbers we have

$$s_n = t_n + t_{n-1} = \frac{n(n+1)}{2} + \frac{n(n-1)}{2} = n^2 .$$

But the triangular number are created by adding an extra row of n dots to t_{n-1} to obtain t_n. So

$$\begin{aligned} t_n &= t_{n-1} + n \\ &= t_{n-2} + (n-1) + n \\ &= \dots . \\ &= t_1 + 2 + 3 + \cdots + (n-1) + n \\ &= 1 + 2 + 3 + \cdots + (n-1) + n \end{aligned}$$

Hence we obtain the well known formula

$$1 + 2 + 3 + \cdots + (n-1) + n = \frac{n(n+1)}{2}.$$

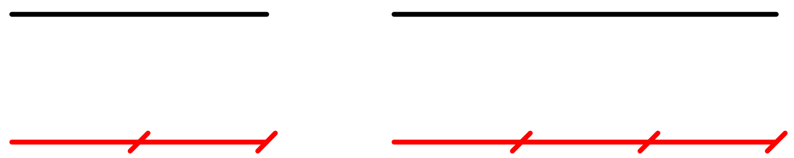
Again, the Pythagoreans knew about such a result but it is not known exactly how they derived it (see [12] and [78] for more).

2.8 On ratios

Article 27: The Pythagoreans also dealt with ratios in their numerical work. However, their motivation for, and definition of, ratios was different to ours. Their study of ratios was born from a study of musical tones. And their definition of ratios was that of only comparing whole numbers. They had no conception of fractions as we do now, so that a ratio of 2:3 could not be interpreted as the fraction 2/3 or 0.66666666... (in fact, decimal fractions did not exist in the days of the Pythagoreans).

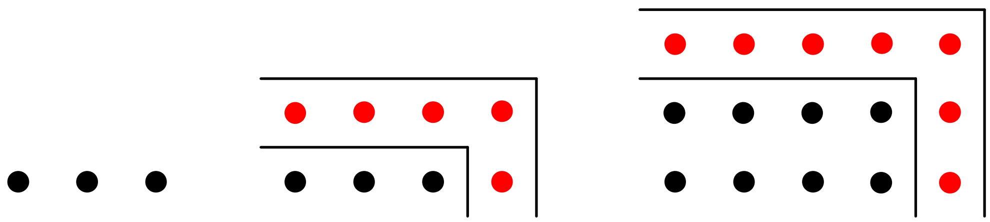
Today we say that 2/3 represents two parts out of a whole consisting of three parts. But such a conception would be anathema to the Pythagoreans since, to them, every number was a whole number and could have no parts. A line one unit long could not contain two parts, each being one half of the line. All that would have happened is that one would have split the line into two shorter lines.

In that case the Pythagoreans would use ratios as a means of comparing two wholes, as when comparing lengths of lines. In other words, the ratio 2 : 3 meant that one line being twice the length of a unit line was being compared to another line being three times the length of a unit line. This is illustrated diagrammatically below where the two black lines can be seen by the red lines to be in the ratio 2 : 3.

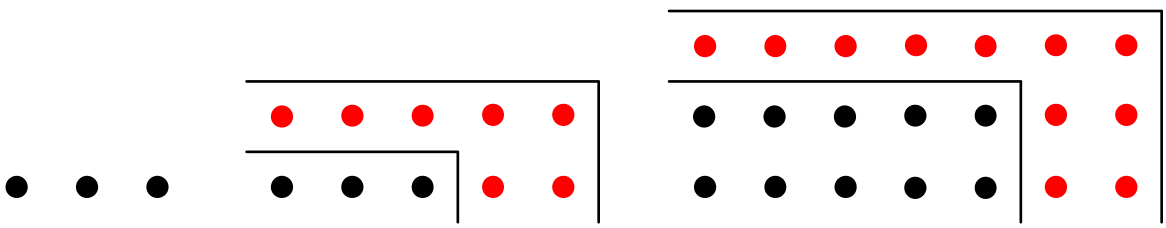


The Pythagoreans then used ratios to compare sides of rectangles. As such we can say that the ratio of the rectangle's shortest side to its longest sides of article 27 is 3:2 and 4:3. Clearly, we can generate many different type of rectangular figurate numbers depending upon how many points we start with and what gnomon we use.

Up to now we started with two points and a gnomon which successively added one more point to the base of the rectangle. But we could have started with three points and a gnomon which successively added one point or two points to the base of the rectangle, as illustrated below.



or



This implies that each rectangular figurate number has its own ratio of sides. For the rectangular numbers above we have the sequence

$$1:3,\ 2:4,\ 3:5,\ 4:6,\ \ldots$$

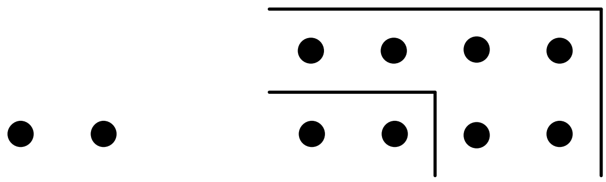
and

$$1:3,\ 2:5,\ 3:7,\ 4:9,\ \ldots$$

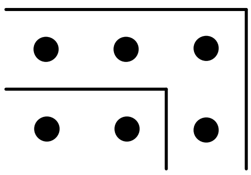
etc.

Article 28: The Pythagoreans managed to classify all possible ratios as falling into one of six categories as follows:

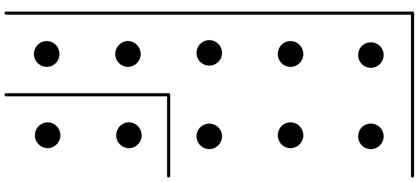
- Equal ratios relate to the ratio of the sides of square numbers.
- Multiple equal ratios relates to the ratios of one side of the square to a multiple of the other sides of the square. These are rectangles with sides a and b such that b is an integer multiple of a. The diagram below we see the case when $a = 2$ and $b = 4$.



- Epimore ratios are ratios where one side of the rectangle differs by 1 from the other side of the rectangle. For example, the rectangular number illustrated below forms an epimore ratio 2:3.



- Multiple epimore ratios are those where the larger term of a ratio is a multiple of the smaller terms plus 1. So, given the rectangular number illustrated below, whose ratio of sides is 2 :5, the multiple epimore ratio is 2 : (2×2+1). where the "+1" is the epimore aspect, and the "2×2" is the multiple aspect.



Similarly, 3:10 is multiple epimore since we have 3 : (3x3 + 1).

- Epimere ratios extend the idea of epimore ratios by adding not just 1 but any number m. So, these are ratios of the form n: $(n + m)$ where $m > 1$. Hence 3:5 is 3:(3 + 2), 4:7 is 4:(4 + 3), and 5:9 is 5:(5 + 4). The reason for needing epimere ratios is due to the fact that none of the epimere ratios can be expressed as epimore ratios, i.e. in the ratio 5:8, 8 cannot be expressed as an integer multiple of 5. In modern terms epimere ratios are ratios where the integers are co-prime (i.e. have no common divisor).
- Multiple epimere numbers follow the same logic as multiple epimore numbers.

2.9 Conclusion

The Pythagorean classification of numbers, their analysis of number patterns, and the arithmetic of figurate numbers presented above illustrate a strong arithmetic mindset. As such we could properly call their work arithmetic analysis. The Babylonians and Egyptians (who came before Pythagoras) certainly used numbers and arithmetic but only in a practical sense for commerce, surveying, geodesy, etc., and did not pursue a more systematic or theoretical study of these (see [68] for more). In this sense we might say that the Egyptians and Babylonians were applied arithmeticians. On the other hand, the Pythagoreans seem to be the first people of record to make numbers and arithmetic an actual subject of study by studying and organising numerical patterns and finding arithmetic in these patterns. So, they may well have been the first pure arithmeticians.

3 The primacy of Geometry: Euclid's elements

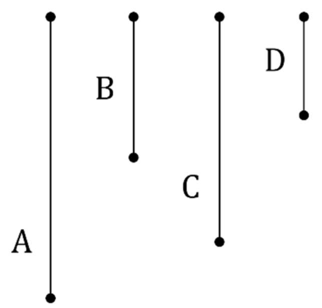
3.1 Introduction

It is hard to underestimate the influence the *Elements* of Euclid (anywhere between 347 B.C. – 212 B.C.) have had, for good and for bad. Since its publication circa 300BC it has been regularly reproduced and commented on to this day. Its content has been included as part of the school curriculum also to this day.

Euclid's *Elements* is a collection of the known geometry to that time. Much of the content of the *Elements* is not Euclid's. But what is his is the systematisation of the subject of geometry, starting from definitions and axioms, to building geometry from the simple to the complex through a deductive chain whereby new theorems were based on previous theorems and axioms. This is what was new and paradigm shifting in mathematics.

Euclid, just like the Pythagoreans, defined geometric things such as points, lines, planes, angles, circles (and their properties such as diameter), triangles, quadrilaterals. He also defined numbers and arithmetic (Books VII to IX). And, just as Euclid's Book I starts with definitions of geometry, his book VII (the first of his three books on numbers and arithmetic) starts with definitions. It should be noted, however, that Euclid does not deal with pure numbers. His numbers always refer to lengths/magnitudes of lines. Effectively he is numericalising geometry. For example, Book VII, proposition 13 starts with a statement of the problem as:

> Let the four numbers A, B, C, and D be proportional so that so that A is to B as C is to D.



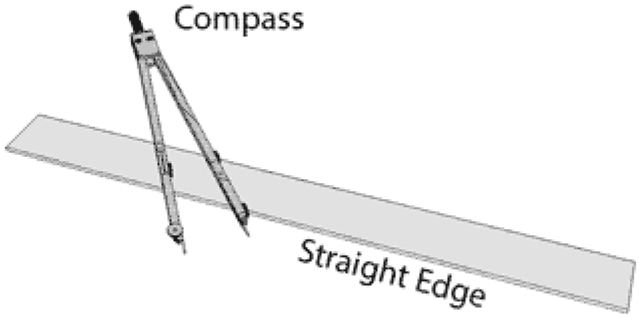
Apart from definitions, Euclid also states axioms which he categorises into two types, one type called *postulates* and another type called *common notions*. Today we don't make such a distinction, but in Euclid's time common notion were self-evident truths which were universal, whereas postulates were self-evident truths relating only to geometry. In which case, the axiom "If A equals C and B equals C then A equals B" would be classified as a common notion because it applies to any comparable objects A, B, C: lines, numbers, apples, chairs, etc. On the other hand, the axiom "A straight line can be drawn between two points" would be classified as a postulate since this statement relates only to geometry. For our purpose we will use the modern understanding of the word axiom to cover these two types.

It should be noted that although Euclid states axioms for geometry, neither he not the Pythagoreans ever stated axioms for the construction of these numbers, nor did they state axioms for arithmetic such as $a \times b = b \times a$ or $a(b + c) = ab + ac$. Today we have what is known as either set theory or the Peano axioms for the construction of the natural numbers, and Dedekind cuts or Cauchy equivalent classes of sequences for the construction of real numbers, as well as axioms of arithmetic.

It can be theorised that geometry, as organised by Euclid, became the paradigm of rigour, and hence the foundation of mathematics. This is due to the systematic approach of using definitions, axioms, proofs based on definitions and axioms, and then proof based on definitions, axioms, and previously proved conjectures to build a coherent approach to analysing geometric structures and relationships. But, as seen by proposition 13 above, he did not have purely numerical/arithmetic definitions and axiom for number and arithmetic. Instead he relied on associating numbers with geometric elements such as lines. Furthermore, the devices used to draw geometric figures, the straight edge and pair of compasses (see next section), proved the constructability of such figures. The proof is in the fact that the final diagram has been drawn by ruler and compass.

3.2 On the construction of geometric objects

Euclid stated the basic axioms of geometry two of which are that a straight line can be drawn between any two points, and a circle can be drawn from any given centre and with any given radius. Points, lines and circles are the most fundamental starting points of geometry. They describe the theoretical objects which are taken as the most basic shapes. However, geometry is a discipline in which we *actually draw* points, lines, and circles (as well as other objects deriving from these). So, what could the most basic, elementary mechanical instruments which could be used to draw these objects? The answer is a straight edge and pair of compasses (known hereafter simply as a compass) as seen in the diagram below. There are no simpler types of instruments which can construct points, lines and circles. And, from the time of Euclid through to the 19th century, all geometric constructions had to be performed using only these two instruments.



As a result of this practical restriction curves which could be drawn with these two instruments were called *geometric curves*. Other curves were known to the ancient Greeks particularly the conic sections such as the ellipse, the parabola, the hyperbola but these could not be constructed by straight-edge and compass.

Similarly, only certain geometric operations were possible using the two tools above. Some of these included adding and subtracting lines, transferring lines from one position to another, multiplication and division of lines, and even square rooting of lines. More complicated operations such as trisecting angles, transforming a circle into a square of the same area, or doubling the volume of a given cube were not possible (although such operations were possible using more complex devices as constructed by some of the ancients such as Nicomedes, as well as Galileo, Descartes, and others in modern times).

So why did ancient geometers place such the restriction of the type of devices which could be used? Because of the philosophy of the day. This philosophy, originated by Plato (428-427 B.C. – 348-347 B.C.), was based on the concept of ideal forms. The realm of ideal forms was a realm where perfect objects existed. There were perfect lines, circles, triangles, polygons, etc. The circle was considered the perfect geometric form reflecting unity and wholeness, and the equilateral triangle represented balance and harmony.

In the case of a line, the ideal form of a perfect line is one which has absolutely no kinks, no curvature, and absolutely no change in direction. This compares with the imperfect line drawn on paper which is a corrupted version of the ideal line because it is subject to all sorts of unevenness due to the physicality of the paper, the unevenness of the straight edge, and the nib of the pencil or pen. I might go so far as to say that even a laser beam, which can be directed with incredible accuracy and straightness would be seen by Plato as flawed compared to the ideal line. Similarly, the points, circles, triangles, etc we draw are not perfect.

The ancients wanted to wanted to emulate the exactness associated with geometric ideal forms. But, in constructing their geometric figures they needed to use certain instruments. The use of such instruments in constructing an ideal line or circle in the real world would “corrupt” said ideal line or circle. So, in order to minimise such corruption they specified that the instruments had to have minimal mechanisms. And the instruments with minimal mechanisms for drawing lines and circles were the straight edge and the compass (note that the straight edge is not a ruler since a ruler has graduation marks on it which may not be perfectly located, and since errors in measurement are made when using the gradations. This would therefore add another layer of corruption or degradation to the ideal form of a line).

Corruption of the form of ideal geometric objects becomes worse when dealing with even more complicated geometric objects such as ellipses, parabolas, hyperbolas, conchoids, spirals, cycloids, etc. These latter objects are curves whose curvature changes along the path of the curves or have abrupt changes in direction. For example, a circle has a constant curvature since the end point of the circle meets the starting point of the circle. But a spiral is a curve where its end point does not meet is starting point because its curvature is always changing. So, the ideal perfect spiral will exhibit an ideal change of curvature along its path. But, whatever device we use to physically draw the ideal spiral on paper, there is the possibility of ever accumulating corrupting because of the continual change in curvature.

> “The philosophers wanted to free their reasoning from the imperfections of the world so that they could understand the true, underlying "essence" of the ideas they were contemplating.
>
> Geometric construction was an attempt to implement these pure ideas. A straightedge and compass are relatively pure instruments—a straightedge just has to be straight, and a compass just has to be able to keep a constant radius as you draw circular arcs. These instruments are not very far from pure, perfect ideals, because they are simple. There aren't a lot of complications that can introduce imperfections into the work.
>
> In contrast, a ruler, in addition to being straight, has to have a lot of little marks, and those marks have to be positioned in precise locations. (Likewise for a protractor.) If you use a ruler to draw something, you have to trust that the maker of the ruler not only made it straight but also put those marks in the right places. That requires more faith in an inherently imperfect object than using a straightedge does. This was philosophically unsatisfactory for the Greeks, who wanted to approach a pure ideal as closely as possible, unpolluted by imperfection.
>
> So the Greeks stuck to the purest tools they could use—an unmarked straightedge and a compass. They also formulated rules about what is and is not allowable in a construction, to limit how much imperfection can creep in. For example, you can use the compass to draw a circle if you have already constructed the point at the center of the circle and a point on the circumference. But you cannot measure a distance at one location, pick up the compass, and draw a circle of that radius at another location, because the act of picking up and moving the compass could change the radius slightly, and that kind of imperfection should be avoided” (From

https://www.reddit.com/r/askmath/comments/28adsd/why_did_ancient_mathematicians_focus_so_much_on/)

So it was that the circle and the straight line were considered the only acceptable geometric curves when it came to constructing geometric figures. Thus, any solution to a geometric problem had to be constructable by straight-edge and compass alone (the reason for this goes back to Plato as explained on p29). The conic sections were known and used but did not have the importance and precedence of lines and circles. Exotic curves such as spirals, quadratrix, conchoids, etc., had no place in Greek geometry because there was no way of constructing them by straight-edge and compass.

So, what is it that can be drawn, and where do we start? A view held by Aristotle, and one which was carried forward for about 2000 years, was that points where either created by the intersection of two lines or were the extremities of a finite line, known as a line segment. Since a line segment can be drawn only when we have two points we start with points, specifically two points, as already given to us (this contrasts with the axioms of arithmetic where we start with only one number as given, either 0 or 1). To see this, think of placing pencil on a piece of paper. This placement first creates a point. Then by using a straight edge the motion of the pencil along the straight edge creates a line. Then when the line terminates it creates another point to give a line segment. Without the terminating point there is no line segment.

We then use points, lines and circles as a basis for constructing other geometric objects such as

- triangles, squares and other polygons,
- perpendicular lines, angles or tangents,

or

- new points, these coming from the intersection of lines and/or circles.

Hence we can set up two axioms which are, in fact, those given by Euclid:

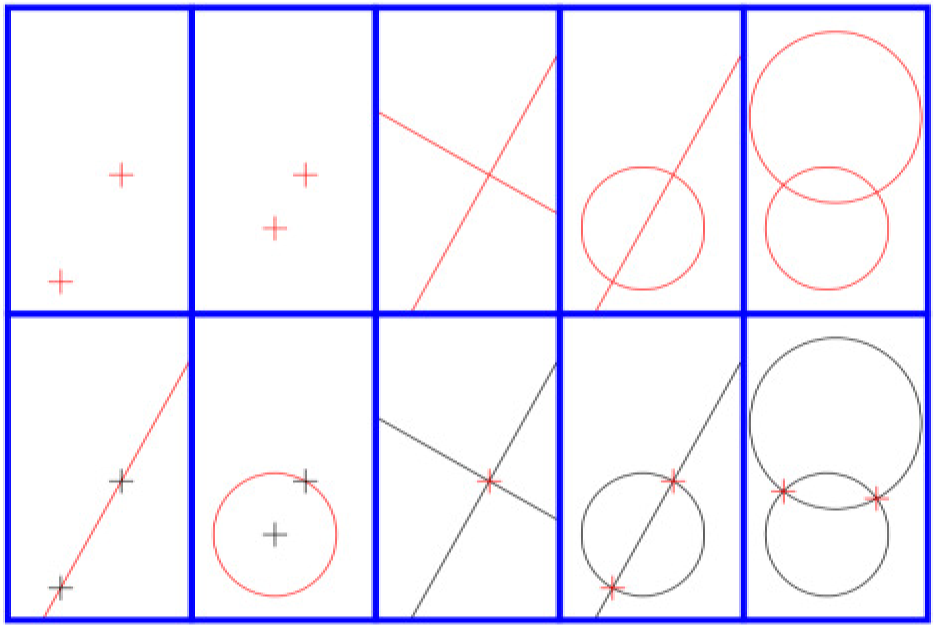
- A line segment can be drawn between any two points.
- A circle can be drawn through one point with another point as its centre.

There are then five basic constructions that can be performed using straight-edge and compass:

- Constructing a line through two points;
- Constructing a circle through one point with another point as its centre;
- Constructing a point which is the intersection of two non-parallel lines;

- Constructing a point which is the intersection of a line and a circle (if they intersect);
- Constructing a point which is the intersection of two circles (if they intersect).

So for any geometric problem we have initial objects (points) from which other basic objects can be constructed (lines and circles) and from which even more complex objects can be constructed (polygons, angles, perpendicular bisectors, etc.).



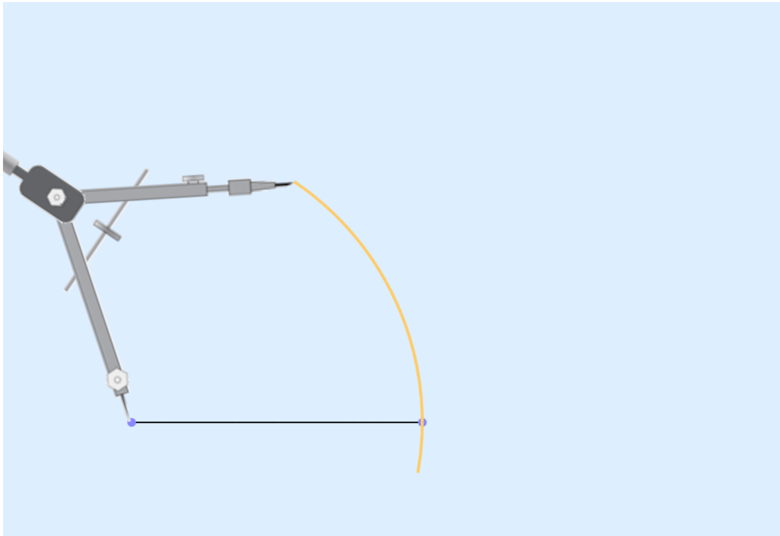
3.3 Examples on how to construct geometric objects

So, one question we can ask is, How do we know that an equilateral triangle exists? Because it can be constructed as follows (All images to follow are taken from https://www.mathsisfun.com/geometry/construct-equitriangle.html. See this URL for many more animated geometric constructions):

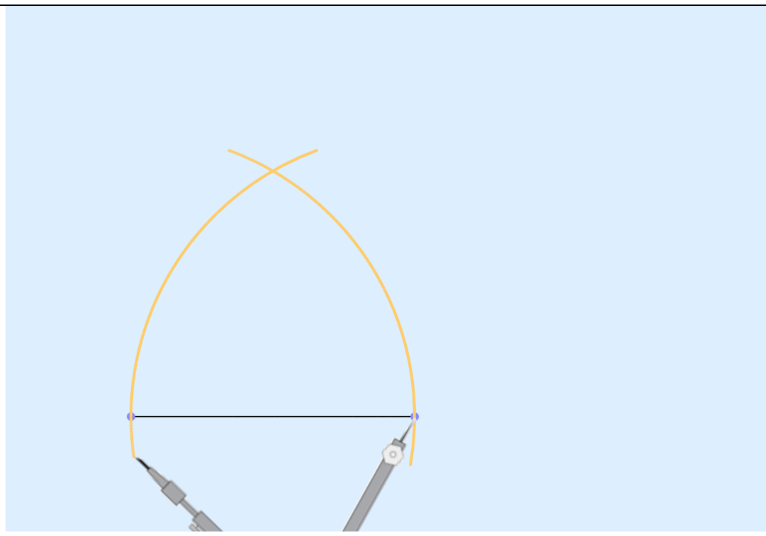
Draw two points then use a straight-edge to draw a line joining these two points

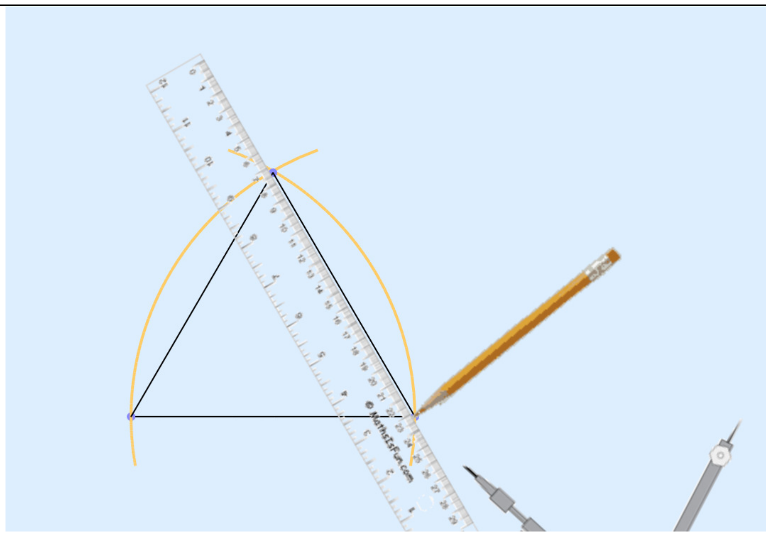
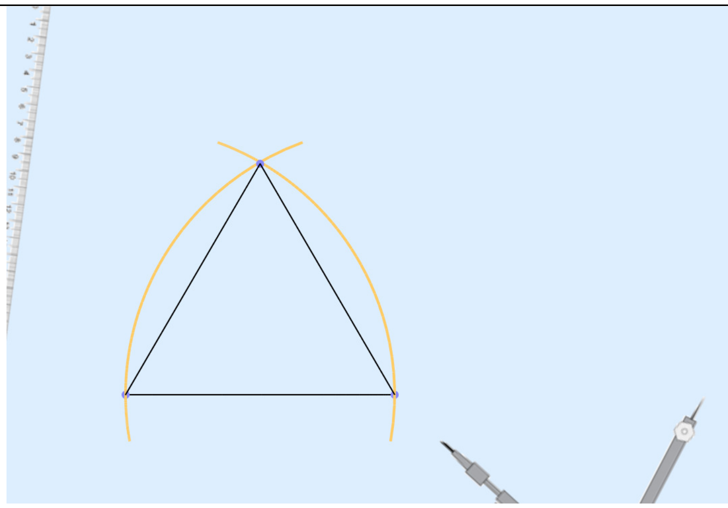
Next, place one end of your compass at the starting point of the line, and open the compass to extend to the end point of the line.

Now draw an arc to pass through the end point of the line.

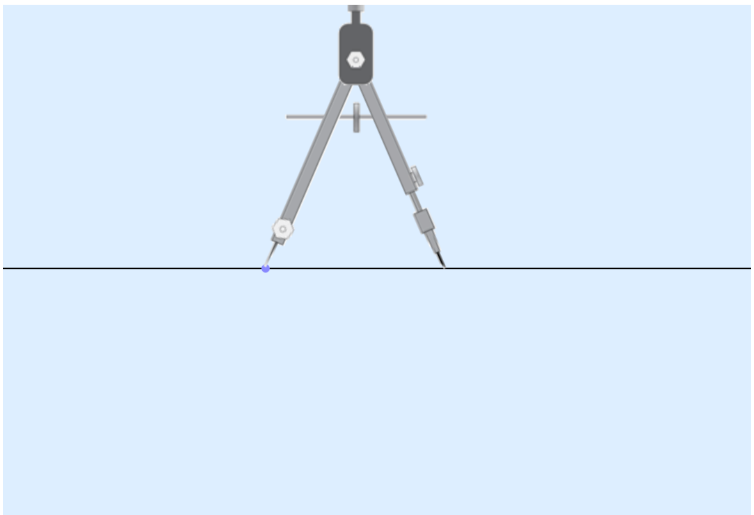


Repeat this process by drawing two arcs, each to pass through the starting points of the line.

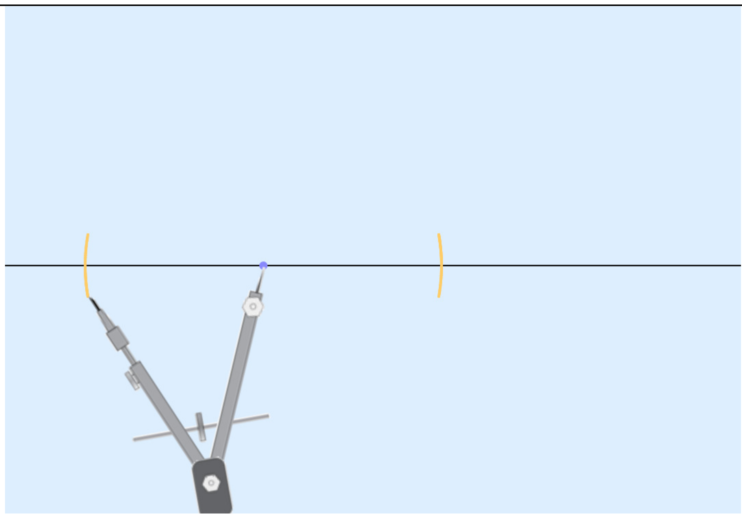


The two arcs intersect, thus constructing a new point. Now use the straight-edge to join the point of intersection of the arcs with the end points of the line	
We now have an equilateral triangle	

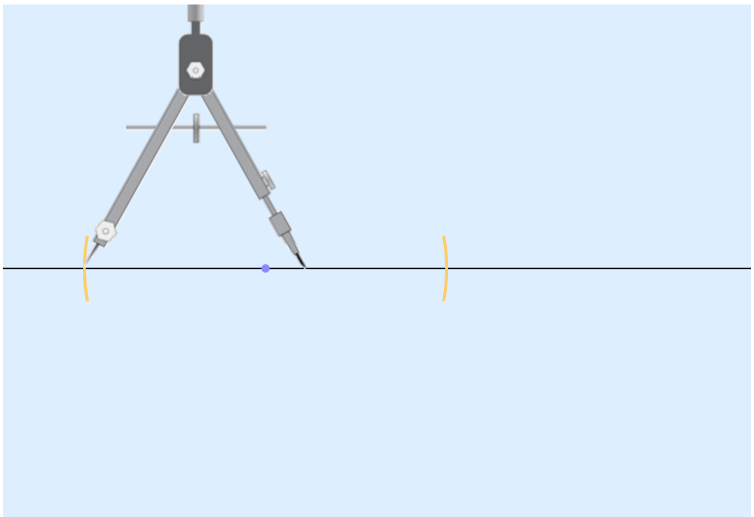
We are also able to use straight edge and compass to construct perpendicular lines:

For a perpendicular, assume we have already constructed a line segment. Now, assume a point given anywhere on this line segment. Take your compass, and from your point open up the compass to any amount	

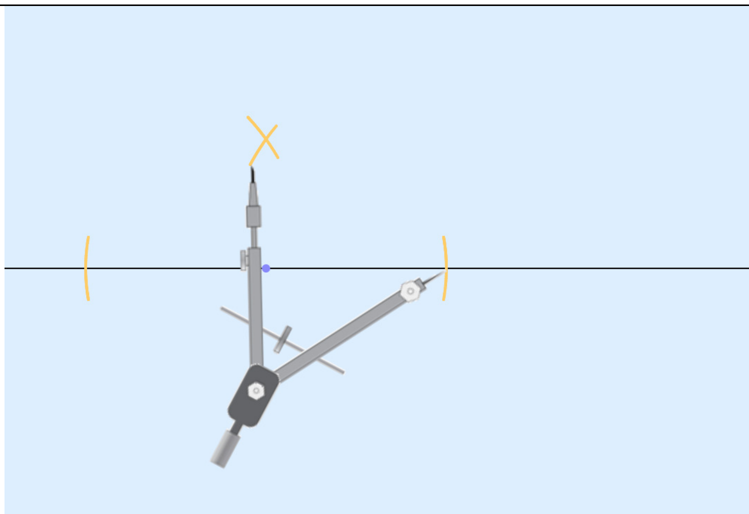
Now draw an arc which crosses the line on one side of the point, then separately on the other side of the point



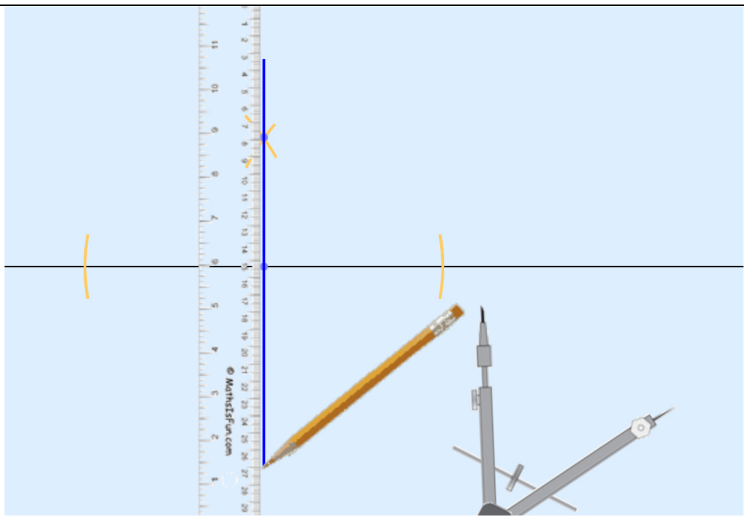
The arcs can be seen to intersect the line segment (thus constructing new points), and the original given point can be seen to be half way between the arcs. Now place your compass at one of the intersection points and open the compass to any length beyond the half-way mark.



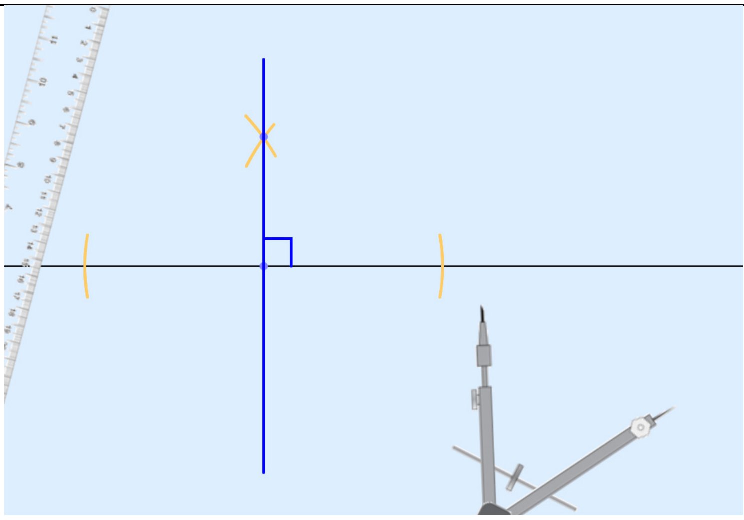
From the arcs on the line draw two other arcs above the central point.



The intersection of the two new arcs creates a point. Use your straight-edge to join this new point with the original point on the line.



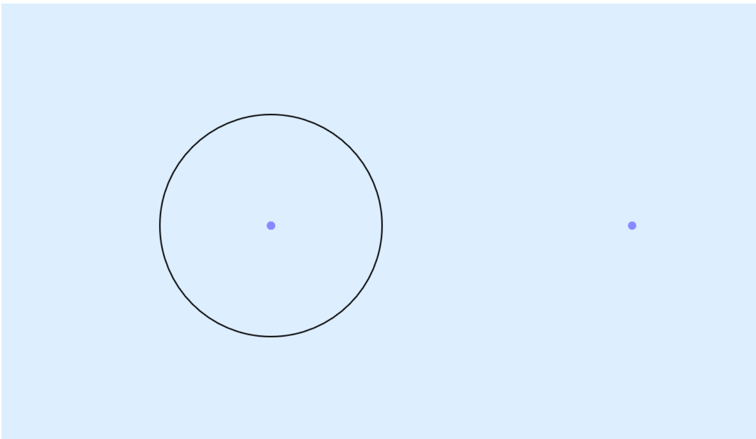
We now have a perpendicular to the horizontal line.



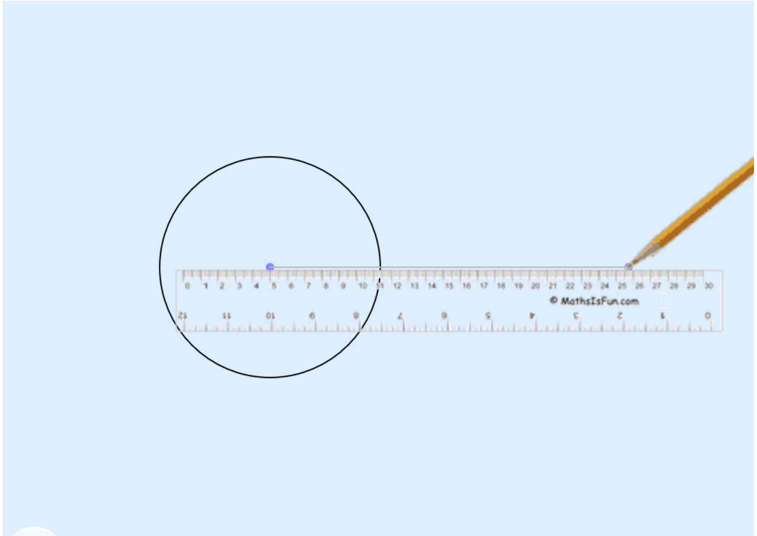
Knowing how to construct an equilateral triangle as well as perpendicular lines allows us to can bisect a line segment. Refer to Euclid I.10 for more.

And finally, tangent lines exist because they can be constructed as follows:

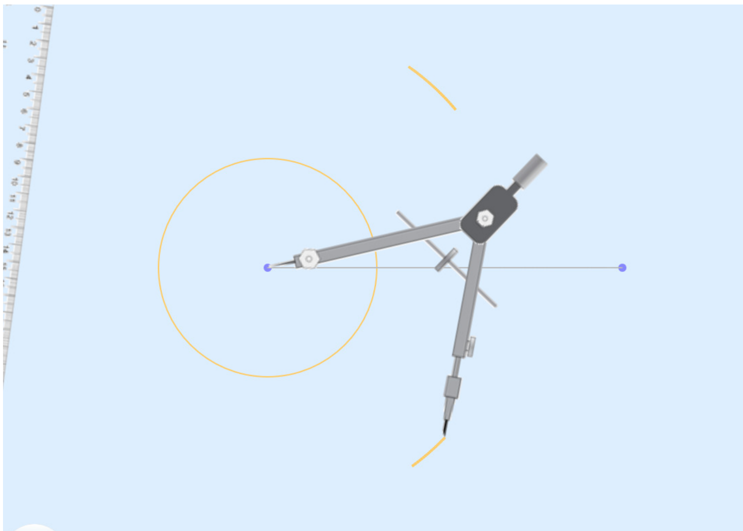
Construct a circle and draw another (given) point outside the circle level with the centre of the circle. This given point can in fact be anywhere outside the circle.



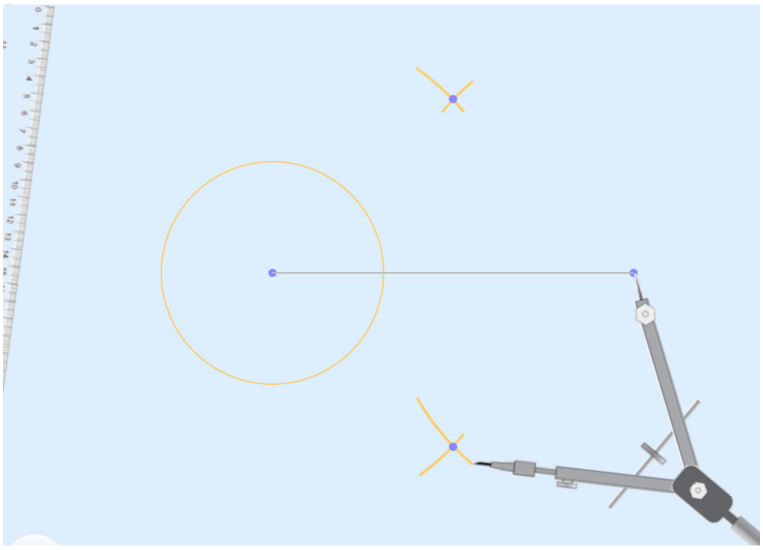
Draw a line to join the two points.



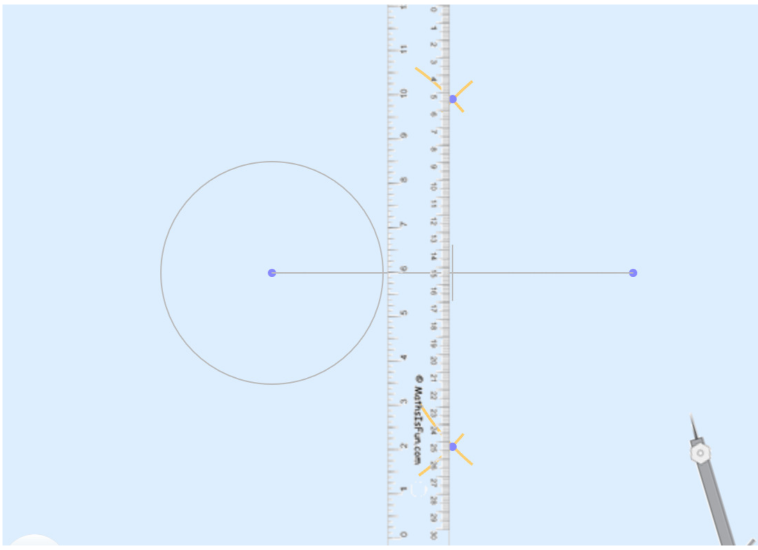
Take your compass, open it up to an amount greater than the radius of the circle but less than the length of the line; Place the compass at the centre of the circle and draw two arcs as shown, one above the line and one below the line.



Now take your compass to the point at the other end of the line and draw two arcs to cross the previously constructed arcs. The intersection of these arcs constructs two new points.

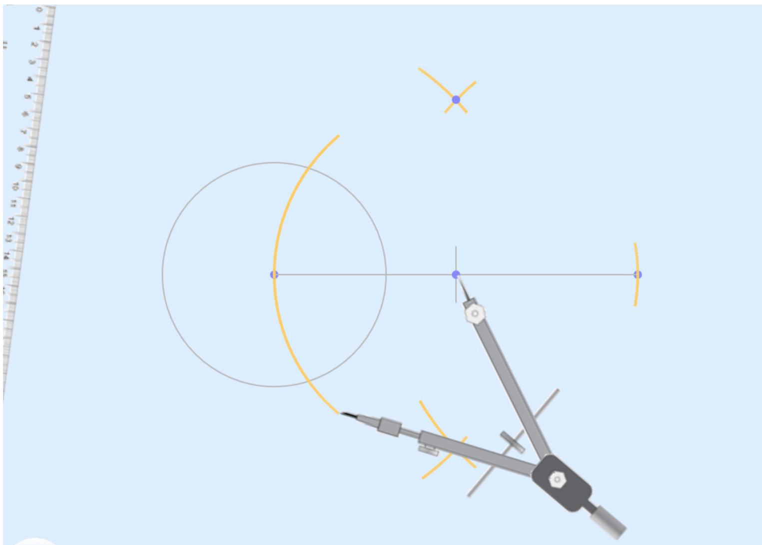


Use the straight edge to construct a perpendicular line segment which cuts the horizontal line. The intersection of the vertical and horizontal line constructs a new point (which happens to be midway between the two points on the horizontal line).

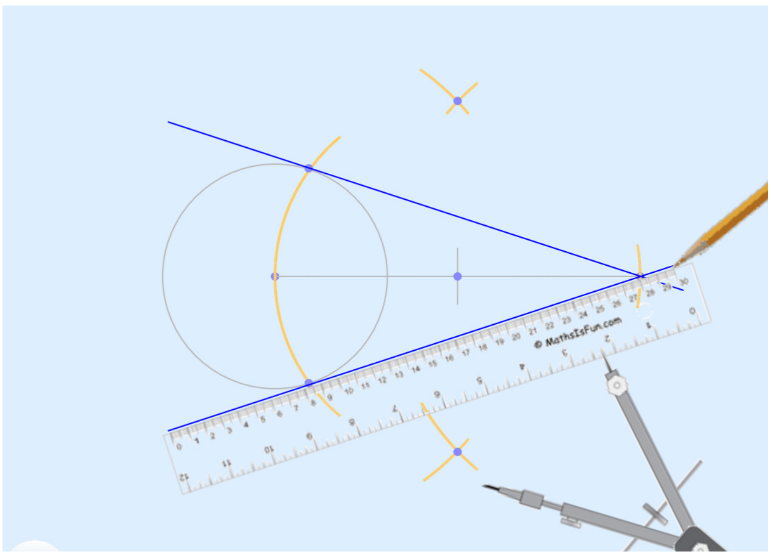


Place your compass on your newly constructed point and open it up to the end point of the horizontal line. Then draw an arc through this end point.

Also, draw an arc through the centre of the circle. This constructs two new points on the circumference of the circle.



Use your straight edge to construct a line joining one of these circumference points to the point at the end of the line. Then, repeat for the other point on the circumference. These two newly constructed lines are tangent to the circle at the points on the circumference.



But this is not all we can do with geometric construction. We can also perform arithmetic on these lines, curves, etc. This is called geometric arithmetic. We will see more of this in section 7, but it is possible to use straight edge and compass to add, subtract, multiply, divide and even take square roots of line segments, the result of this being line segments and curves.

3.4 A simple Euclidean geometric proof

(to come)

4 Selected rare moments of arithmetic study from Euclid onwards

Mathematicians subsequent to Euclid lived in a time when, due to Euclid's work, geometry had become the paradigm of mathematics. As a result, very few of them were arithmeticians, i.e. mathematicians studying arithmetic independently from geometry. But some did, such as Euclid himself, Heron of Alexandria (anywhere between 1st century B.C. to 3rd century A.D.), Nichomachus of Gerasa (circa 100 A.D.) and Diophantus (1st century A.D. or 2nd century A.D.).

4.1 Euclid

Euclid's thirteen books of the *Elements* is remembered generally for its geometry. But books VII to IX deal with arithmetic. In book VII Euclid gives definitions for things such as

- Unit:
- Number:
- Part:
- Multiple:
- Even and odd numbers:
- The product of even and odds amongst each other:
- Prime numbers:
- Relatively prime numbers:
- Square and cube numbers:

and

- Perfect numbers:

But Euclid gives no axioms of arithmetic. There is therefore no comparison between book VII and book I in the sense that arithmetic is not systematised by starting with definitions and axioms, and then building arithmetic algorithms and procedures from the simple to the complex through a deductive chain whereby new theorems are based on previous theorems and axioms (exactly what is done today in modern analysis textbooks).

In book VII we find the well-known division algorithm for finding the greatest common division of two numbers (since the *Elements* is effectively an organised collection of the mathematics up to that time it is likely that the division algorithm was invented by someone before Euclid. Current thought is that the algorithm came from Eudoxus (395-390BC – 342-337BC)). Book VII illustrates the algorithm by use of line segments to represent numbers

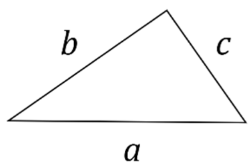
Eudoxus' procedure represents a systematic and generalised nature to the process of division. This goes far beyond what the Egyptians and Babylonians did since neither of these had a systematic algorithm for division. As a result, one might consider the division algorithm as illustrating a proper analysis of arithmetic.

There are two things to note about this algorithm:

1) It is not constructed from previous number or algebraic axioms, definition or theorems. In modern language, it is not deduced logically or arithmetically from prior mathematics which itself has been constructed logically or arithmetically. It seems to simply be a record of the steps needed to work out a division.
2) Euclid's arithmetic is not a true pure arithmetic. As seen above Euclid represents numbers as straight lines with letters marking the ends of the lines. When he wants to do arithmetic on numbers he geometrically manipulates the line segments appropriately. So it might be said that his arithmetic is a geometric arithmetic not a number arithmetic;

4.2 Heron of Alexandria

Heron of Alexandria was a Greek mathematician and engineer who lived circa 100 A.D. He was interested in mensuration, namely the measurement of length, area and volume of geometric figures, and is most remembered for the formula for the area of a triangle based solely on knowing the sides of the triangle:



$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

where a, b, c are the sides of the triangle, and $s = \frac{1}{2}(a+b+c)$ (this formula was probably known before Heron but he is the first to have recorded it, hence it is attributed to him).

Although principally a geometer (his derivation of the formula above is purely geometric, without any trigonometry) Heron did arithmetic work, particularly in finding approximations to roots, this being necessitated by his area formula above.

As we shall see later, the ancients worked with ratios and proportions. This means that they compared line lengths of integer magnitudes. For example, given two lines, if we can find a common measure shared by both lines then we might say that the first line can be assigned a length of 2 units and the second line can be assigned a length of 3 units. Hence the lines are said to be in the ratio $2 : 3$ (note that such a ratio would not represent two-thirds of a line). Similarly, lines could be in the ratio $17 : 27$ or $14688 : 4673$, etc. A problem then arises when one meets what we today call $\sqrt{2}$. This cannot be expressed as a ratio $a : b$, where a and b are integers, so all one could do is to find the ratio which came closest to $\sqrt{2}$ such as $99 : 70$ or $239 : 169$ or $577 : 408$, etc. (if, today, we were to write $\sqrt{2}$ as a ratio we would simply write $1.41421356 \ldots : 1$. But the ancients had no conception of irrational/real numbers as we now do). The implication of this is that most ancients after the period of Thales and Euclid did not actually calculate square roots.

But Heron did not work with ratios. He actually went about calculating/evaluating square roots. His process for finding roots was known by Archimedes (c. 287 B. C. – c. 212 B.C.) before him, and the Babylonians (c. 1800 B.C. – c. 1600 B.C.) before that. But there is an interesting point about the way Heron calculated square roots whereby he consolidated the process used by the Babylonians.

The Babylonians used what might be called a two-step process for calculating roots. So, if they wanted to find $\sqrt{n}$, where n was not a perfect square, they would make an initial guess as to its value. Call this initial guess a_1. They would then form $b_1 = n/a_1$. If a_1 were too small compared to $\sqrt{n}$ then b_1 would be too large. If a_1 were too large compared to $\sqrt{n}$ then b_1 would be too small. In either case the average (arithmetic mean) of a_1 and b_1 would be a better approximation to $\sqrt{n}$. The process would then be iterative:

Stage 1	Stage 2
a_1 is given	$b_1 = n/a_1$
$a_2 = \frac{1}{2}(a_1 + b_1)$	$b_2 = n/a_2$
$a_3 = \frac{1}{2}(a_2 + b_2)$	$b_3 = n/a_3$

etc.

The general iterative formula would be: a_1 as given, $b_i = n/a_i$, $a_{i+1} = \frac{1}{2}(a_i + b_i)$. For $\sqrt{2}$ with a starting value of $a_1 = 3/2$ the Babylonians obtained the following:

i	a_i	b_i
1	$\frac{3}{2}$	$\frac{4}{3}$
2	$\frac{17}{12}$	$\frac{24}{17}$
3	$\frac{577}{408}$	$\frac{816}{577}$

etc. Heron's approach ends up consolidate the two steps of finding a_i and b_i into one step by doing $a_{i+1} = \frac{1}{2}(a_i + n/a_i)$. Heron describes his process via concrete examples. So, having used his formula above to find the area of a triangle to be $\sqrt{720}$ he then proceeds as follows:

> "Since 720 has not its side rational we can obtain its side within a very small difference as follows. Since the next succeeding square is 729, which has 27 for its side, divide 720 by 27. This gives $26\frac{2}{3}$. Add 27 to this making $53\frac{2}{3}$, and take half of this or $26\frac{1}{2}\frac{1}{3}$. The side of 720 will therefore be very nearly $26\frac{1}{2}\frac{1}{3}$. In fact if we multiply $26\frac{1}{2}\frac{1}{3}$ by itself, the product $720\frac{1}{36}$, so that the difference (in the square) is $\frac{1}{36}$.
>
> If we desire to make the difference still smaller than $\frac{1}{36}$, we shall take $720\frac{1}{36}$ instead of 729 [or rather we should take $26\frac{1}{2}\frac{1}{3}$ instead of 27], *and by proceeding in the same way* we shall find that the resulting difference is much less than $\frac{1}{36}$."
> (emphasis added) [p324, 36]

Heron's describes what today we would write symbolically as $a_{i+1} = \frac{1}{2}(a_i + n/a_i)$, where $a_1 = 27$, and $n = 729$ (note Heron's phraseology of 720 being compared with sides not being rational. This relates to the fact that any square of area x^2, where x is integer or a ratio, has side x. Then we can speak of an area of 16 units^2 as having rational sides of 4 units, and an area of 17 units^2 as not having rational sides. When Heron says "Since 720 has not its side rational" he means $\sqrt{720}$ not being rational).

The fact that Heron says "by proceeding in the same way" means that he has an algorithmic process in mind, and the fact that he says "take $720\frac{1}{36}$ instead of 729" means he is thinking iteratively. So, as with Euclid, we might say that Heron's approach qualifies as arithmetic analysis, just as the real analysis of today qualifies as analysis.

Other problems Heron dealt with are equivalent to solving quadratic equations. An archetypal example would be (see [p344, 36], for actual examples considered by Heron), given a square such that the sum of its area and perimeter is 12 find the sides of the square. In modern notation this is equivalent to solving $x^2 + 4x = 12$. Both the quadratic formula and the process of completing the square were known in Heron's day, so that the answer would be $x = 2$, where the negative answer would have been ignored since no geometric object could have sides of negative length. It should be noted that in this era mathematical solutions were not presented as algebraic steps (as we do today) but were described in prose. For example, the solution to the above quadratic by the process of completing the square could be described as,

> Firstly, remove 12. Thus, the result gives you nothing. Then add and take away 4. The addition of the 4 can be combined with the first two elements to form a square, and the combined removal of 4 and 12 remove 16 from this square. Reinclude 16. This is the square of 4. Now remove 2 to obtain the side of the square to be 2.

This is certainly a convoluted way of describing the algebraic process, but this is exactly the type of style used before the advent of algebra. This compares with the standard way of showing the solution, namely

$$\begin{aligned} x^2 + 4x &= 12 \\ \Longrightarrow \quad x^2 + 4x - 12 &= 0 \end{aligned}$$

Completing the square

$$\begin{aligned} x^2 + 4x + 4 - 4 - 12 &= 0 \\ \Longrightarrow \quad (x+2)^2 - 16 &= 0 \\ \therefore \quad (x+2)^2 &= 16 \\ \Longrightarrow \quad x + 2 &= 4 \\ \therefore \quad x &= 2. \end{aligned}$$

The point to note is that the descriptive solution does not use any form of geometric reasoning. Certainly it uses terms such as "square" which is a geometric term, but such terms (which were used freely in solutions) would only be used to refer to this or that unknown needing to be found. Furthermore, there is no need to construct the geometric solution to the answer. Therefore, such an arithmetic/algebraic analysis as above would be seen to stand on its own two feet, independent from geometry.

It should be noted that, when geometry has become the paradigm in the study of pure mathematics, geometers only used numbers and letters to refer to elements of geometric figures. A line might be said to be of unit length only so that it could be compared to another line double the length. Such a comparison could then be written as the ratio 1 : 2. Not only is this ratio not a fraction (it does not represent ½), it isn't even comparing two integers. It simply represents the comparison of two lines, with the numbers acting as proxies for the lines. And when algebra started being developed by Vieta (1540 – 1603), Descartes (1596 – 1650) and Fermat (1601 – 1665) in the 16^{th} and 17^{th} centuries, such an attitude carried over to algebra. Any equation using letters would only be a representation of a geometric configuration. The numbers and letters would have no meaning in and of themselves, independent of the geometric configuration. So, there was no such thing as the number "2" in our modern sense. Certainly, there would be 2m, 2kg, 2 seconds, 2 litres, etc., but no "2". The same goes for the use of letters such as *x* or *y*. Today we conventionally use these to represent variables. But in the days of the ancients, and even in the 1500s, there was no conception of treating letters as independent mathematical objects which could be studied, manipulated or analysed. Any algebraic manipulation effected on equations would be a shorthand for the process of geometric transformation by the use of a straight edge and compass (see section 7.8 for more on this). Any solution to an equation obtained by algebraic manipulation would then only be seen as valid if it could be constructed geometrically.

In this context one might say that the work of Heron of Alexandria on generalising the process of finding roots to a number (i.e. having an iterative algorithm) might be considered to be quite a paradigm shift. But, although he studied arithmetic as arithmetic (not as a proxy for geometry) he was still principally a geometer.

4.3 Nicomachus of Gerasa

Nicomachus is one of the few ancients we know of to have presented a systematic and organised study of arithmetic independent of geometry. He wrote a book called *Introductio Arithmetica* which was as influential as Euclid's *Elements* not for its mathematical originality but for its long standing use, to the extent that it was a standard for more than 1000 years, with many commentaries made about is over that period (one such well known commentator was

Boethius (circa 480 – 524 A.D.). See [80] for more). It should be noted that (based on extant knowledge) that Nicomachus was a philosopher and commentator of mathematical works rather than a mathematician.

The mathematics of *Introductio Arithmetica* is Pythagorean in its approach. Illustrative of its content is the following:

- since Nicomachus was a Pythagorean he speculated in mystical as well as mathematical interpretations of numbers. As such he distinguished two types of number: the conceptual numbers which he called “divine numbers” (p28, [12]), and the practical numbers used for measurement and other physical, material uses, which he called “scientific numbers”. I might say that Nicomachus’ distinction is something like that of the number 1 as a pure mathematical object compared to 1m as an applied mathematical object prone to human error in its measurement. This suggests that there is an exactness to a pure number which does not exists in the applied case. And such a distinction links back to Plato’s ideal forms on p29, where the perfect, eternal and unchanging object exists, and the material Earthly world where a form is instantiated for practical use, leading to inaccuracies and errors of measurement;
- defining what we today call the number 1: "Unity, then, occupying the place and character of a point, will be the beginning of intervals and of numbers, but not itself an interval or a number, just as the point is the beginning of a line, or an interval, but is not itself a line or an interval." (Nicomachus[Nic52], II.6.3, p. 832. Nicomachus[Nic52], 1.7.1, p. 814);
- the classifying of numbers as even and odd, then subdividing these as i) evenly even, i.e. numbers of the form 2^n, ii) even-odd, i.e. numbers of the form $2(2n + 1)$, and iii) odd-even, which in this case mean numbers of the form $2^{m+1}(2n + 1)$. To our modern eyes, classification iii) is the same as classification ii), but this is what Nicomachus presented.

  Odd numbers are then subdivided into i) prime and incomposite (i.e. prime numbers), ii) secondary and composite (i.e. products of primes), and iii) “that which is secondary and composite, but in relation to another is prime and incomposite” (p100, [36]), in other words two numbers which are composite but are relatively prime, such as 25 and 49;

- A description of perfect numbers these being numbers which are equal to the sum of their divisors (i.e. $1 + 2 + 3 = 6, 1 + 2 + 4 + 7 + 14 = 28$, etc., where the number itself is not included as a divisor). Such numbers were studied by Pythagoras and Euclid. But here Nicomachus goes onto to study the only other two types, namely deficient numbers and super abundant numbers. The former are numbers of where the sum of the divisors is less than the number itself, and the latter are numbers where the sum of the divisors is greater than the number itself;
- a description of the sieve of Eratosthenes which is a procedure for finding prime numbers. This is quite a simple and elegant approach which consists firstly of listing all the odd numbers starting from 3:

  3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 ,31, …

  Now, 3 is a prime number but multiples of 3 are not, so we remove all multiples of 3. Then 5 is a prime number but multiples of 5 are not, so we remove all multiples of 5. This process is then repeated for the next number in the sequence of remaining numbers, these being 7, 11, 13, …;
- figurate numbers. This has been discussed enough in 2.6, so we won't go into it anymore here (for examples of the figurate numbers considered by Nicomachus see p5-7, [88]);
- a version of the Euclidean algorithm. In this case Nicomachus but does not use the symbolic way of Euclid (of doing arithmetic on a line AB) but instead simply describes the process in prose;
- a study of ratios less than unity and greater than unity. Here Nicomachus classifies different types of ratios such as *multiples*, ratios that are doubles, triples, etc., and *submultiples*, where ratios are halves, thirds, quarters, etc. Other categories of ratios include those of the form $1 + \frac{1}{n}$, $1 + \frac{m}{m+n}$, and $m + \frac{1}{n}$;

Nicomachus then goes onto arithmetic proper by showing a rule for constructing such ratios. For example, suppose that $a : b = b : c$ equals one of Nicomachus' ratios. We now form two sets of three numbers

$$a \qquad a + b \qquad a + 2b + c \qquad (*)$$

and

$$c \qquad c + b \qquad c + 2b + a \qquad (**)$$

If we take $a = b = c = 1$ repeated use of (*) gives $(1, 2, 4)$ then $(1, 3, 9)$, then $(1, 4, 16)$ etc., illustrating the case of successive multiples. Use of (**) on $(1, 2, 4)$ gives $(4, 6, 9)$ which elements are in the ratio $3/2$. Similarly, from $(1, 3, 9)$ we obtain $(9, 12, 16)$ which elements are in the ratio $4/3$. Using the (*) on $(9, 6, 4)$ we obtain $(9, 15, 25)$ where the elements are in the ratio $1\frac{2}{3}$, one of the types of ratios identified in Nicomachus' classification. Using (**) again on $(9, 6, 4)$ we obtain $(4, 10, 25)$ where the ratio this time is 2½, another of the types of ratios in Nicomachus' classification.

There is a type of similarity here to modern analysis. In modern analysis we define axioms and procedures for the construction of number systems $(\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R})$. These sets act to classify different types of numbers just as Nicomachus is classifying different types of ratios. Then, in modern mathematics, we construct rules for the use of these numbers, i.e. the axioms of arithmetic, just as Nicomachus goes on to show procedures which gives the ratios he has categorised. I might say that such an approach could be seen as a pseudo-axiomatic presentation of arithmetic. But unlike the geometers from Euclid onwards, Nicomachus provided no general proof from the pattern above. However, what can be said is that Nicomachus' work did indeed illustrate an arithmetic way of thinking which classified as arithmetic analysis independent of geometry.

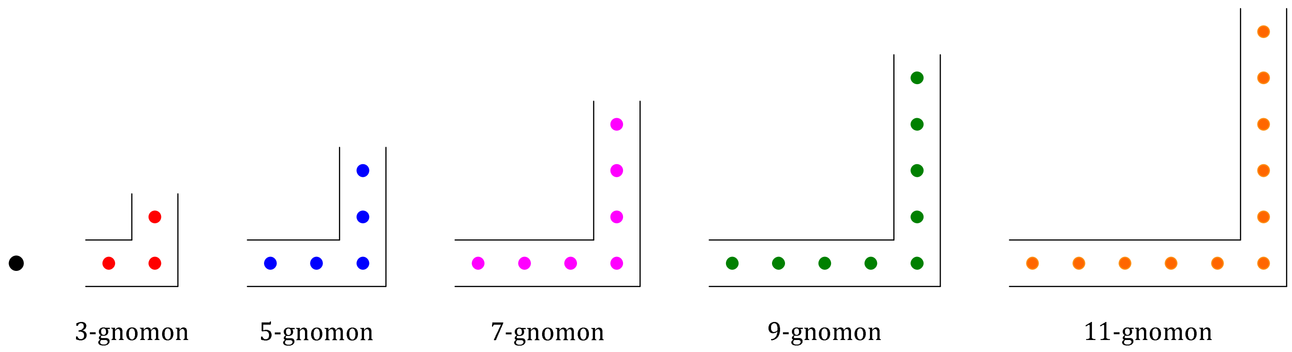
Although Nicomachus was a philosopher and mathematical commentator, he did produce one original piece of mathematical work, this being firstly that the sum of odd integers can be expressed as the sum of cubes of natural numbers, and secondly the connection between the sums of cubes of odd numbers and the square of the sum of integers. In the first case Nicomachus lays out the odd integers and then considers groups of these. So from

$$1, 3, 5, 7, 9, 11, 13, 15, 17, 19, \dots, n$$

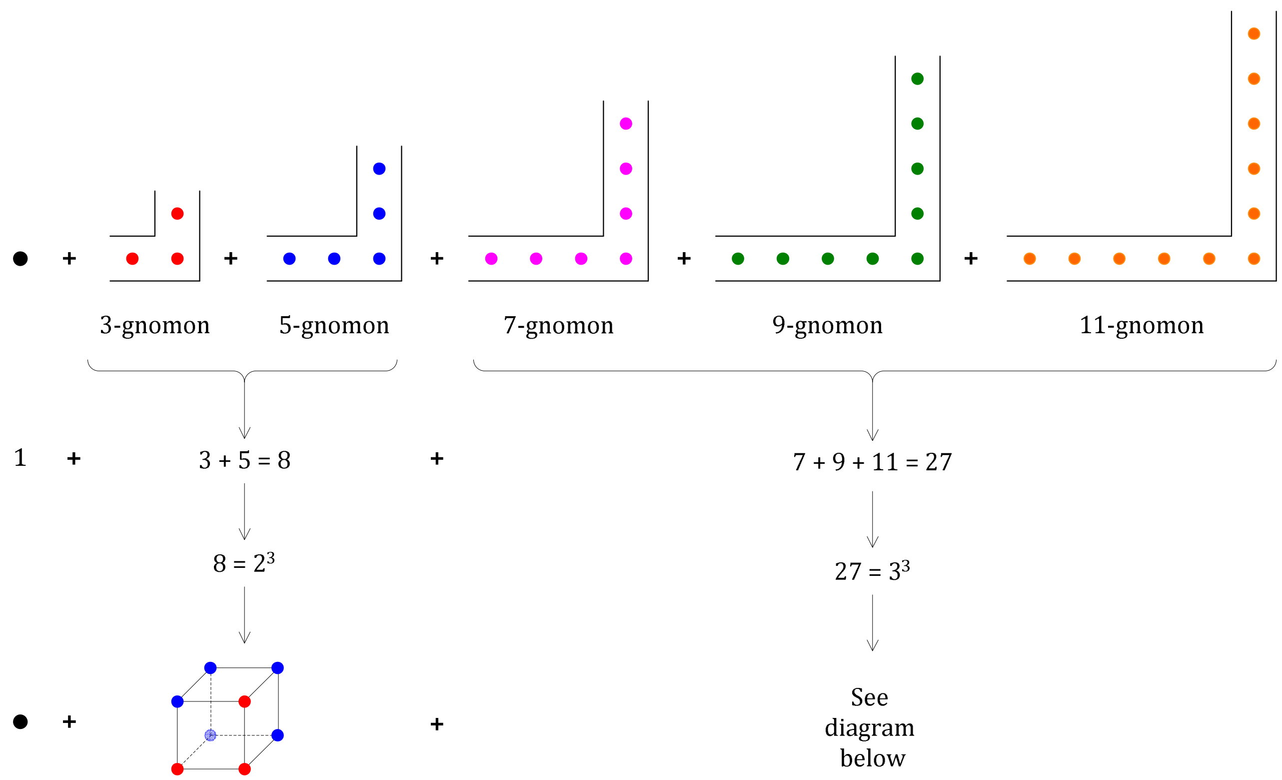
where n is an odd number, the first number, 1, is a cube, the sum of the next two number, $3 + 5$ is a cube, the sum of the next three numbers, $7 + 9 + 11$, is a cube, etc. So we have the series

$$1 + (3 + 5) + (7 + 9 + 11) + \cdots = 1^3 + 2^2 + 3^3 + \cdots$$

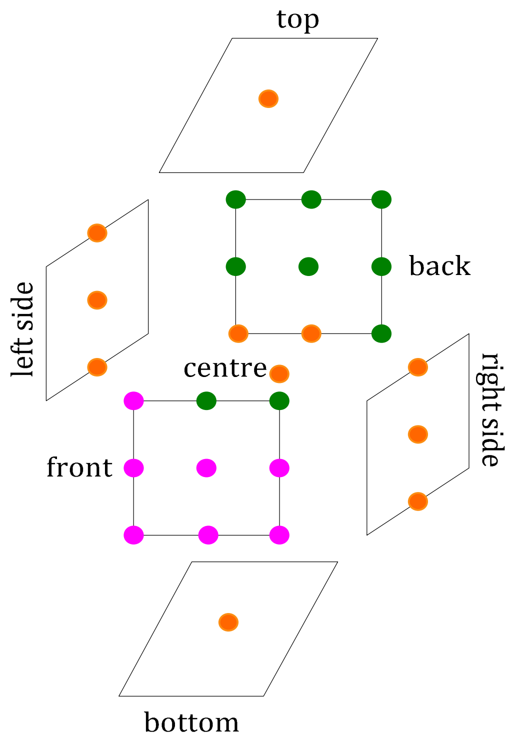
Nicomachus did not prove this result in general, relying instead on and assuming the generalisation based on individual examples. Given that Nicomachus was a Pythagorean, and was therefore versed in using gnomons to identify number patterns, a gnomonic demonstration could have been given. Recall the meaning of "gnomon" on p20. Let us call a single point (the number 1) a 0-gnomon, then the number 3 a 3-gonomn, the number 5 a 5-gnomon, etc., as shown below.



Then, adding the correct combination of gnomons we can obtain cubes of the natural numbers, thus:

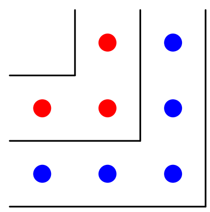


In order to more easily show the distribution of the dots of the 7-, 9-, and 11-gnomons together on one cube I have separated the faces of the cube, with the free-floating orange dot being at the geometric centre of the cube.

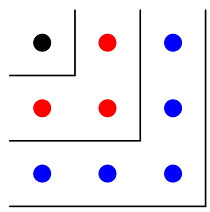


If you join the faces together to form a cube you will see three rows of horizontal and vertical dots whichever perspective you look from.

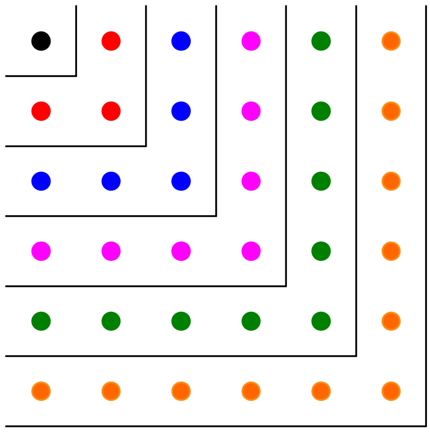
In the second case Nicomachus found that $1^3 + 2^3 = (1 + 2)^2$, $1^3 + 2^3 + 3^3 = (1 + 2 + 3)^2$, etc., which again has a gnomic demonstration, thus: joining the 3-gnomon with the 5-gnomon we have



which is equivalent to the number 8, which happens to be 2^3. If we include the 0-gnomon in this configuration we have



So we have performed the arithmetic of $1 + (3 + 5) = 1^3 + 2^3$. But this figurate configuration above is a perfect square representing 3^2. Hence $1^3 + 2^3 = 3^2 = (1 + 2)^2$. Extending this to include the next three gnomons we have



Here we have had to group the 7-, 9-, and 11-gnomon in order to form the perfect square 6^2. Note that $7 + 9 + 11 = 27 = 3^3$. Hence we have

$$1 + (3 + 5) + (7 + 9 + 11) = 1^3 + 2^3 + 3^3 = 6^2.$$

But 6^2 which can be written as $(1 + 2 + 3)^2$, hence

$$1^3 + 2^3 + 3^3 = (1 + 2 + 3)^2.$$

etc.

The generalisation of this pattern is clearly $1^3 + 2^3 + 3^3 + \cdots + n^3 = (1 + 2 + 3 + \cdots + n)^2$. Nicomachus did not prove this general result (again relying on assuming the generalisation based on individual examples). Instead, as is the want of commentators, he makes extensive commentaries about them as, for example, below.

> "(1) But there appears as a mean between these two kinds already considered, that is as it were opposed in the manner of extremes, the so-called perfect number which is found in the realm of equality. This is a number that neither makes the sum of its own parts greater than itself nor shows itself greater than the sum of its parts, but is always equal to the sum of its parts. Now that which is equal is always regarded as midway between the more and the less and is, so to speak, moderation between the excessive and the deficient, the harmonizing tone between that which is too high and that which is too low. (2) Whenever, then, a number neither exceeds in amount all its parts, after all that it may contain have been combined and added up and compared with itself, nor is surpassed by them in amount, then such a number is

properly called perfect, since it is the number that is equal to the sum of its own parts. For example, the numbers 6 and 28; for 6 can be divided into one-half, one-third, one-sixth, which are 3, 2, 1, and these added together make 6, which is equal to the original number, being neither more nor less. And so 28 has as its parts a half, a fourth, a seventh, a fourteenth, a twenty-eighth, which are 14, 7, 4, 2, 1, and these added together make 28. And so neither are the parts more than the whole, nor the whole greater than the parts, but the comparison results in equality, which is the peculiar character of the perfect. …" (p16-17, [44])

4.4 Diophantus – To come

4.5 Conclusion

It should again be noted that few mathematicians of the ancient period (say from Thales (###) through to Euclid through to the beginning of end of first century) studied mathematics from an arithmetic perspective. Certainly, Euclid's books VII to IX of the *Elements* deal with arithmetic but the main reason these books never came to define the nature of mathematics was partly because of the philosophical attitude the ancients had towards geometry as the example of exactness and perfection, and partly because these books lacked axioms. This contrasts with book I which lays down definitions and axioms of geometry from which every subsequent proposition is proved. For example, Euclid state the following definitions: i) a point is that which has no parts, ii) a line is a breadthless length, iii) the ends of a line are points, iv) a straight line is a line which lies evenly with the points on itself, etc., with subsequent definitions dealing with surface, angles, boundaries, circles and triangles. Then there are the main five axioms of Euclidean geometry: i) a straight line can be constructed between any two points, ii) any straight line can be extended continuously, iii) a circle can be constructed having a centre and a radius, iv) all right angles are equal to each other, and v) two straight lines are parallel if they do not intersect (this is the simplified version of Euclid's original description). Note that axioms i) and iii) are what might be called the primitive or axiomatic geometric figures which are needed for the construction of all other geometric figures thereafter. Hence straight lines and circles need to be stated as axioms since they themselves cannot be constructed from more basic geometric figures.

But book VII (the first book about arithmetic), although having definitions, does not state any axioms of numbers or arithmetic. There is nothing at all about what we today call the associative law of addition or multiplication: $a + (b + c) = (a + b) + c$ and $(ab)c = a(bc)$; there is nothing about the distributive law: $a(b + c) = ab + ac$, etc., there is nothing about the commutative law for addition and multiplication: $a + b = b + a$ and $ab = ba$. This last one is important to state since there are times in arithmetic when the commutative law does not hold (such as matrix multiplication of the vector cross product). So, it is important to state axioms of numbers and arithmetic since in order to know the boundaries within which we can work. For example, all real numbers have an order: $2 < 3$, $-1 < 0$, $\sqrt{2} < \pi$, etc. But there is no such thing as ordering the complex numbers (i.e. is $1 + 2i \gtrless 2 + i$?).

So, Euclid had laid down geometry in a systematic and organised manner on the basis that any statement could be proved from previous statements, these also having been proved from statements previous to these, etc., all the way back to statement considered as self-evident, namely the axioms. In other words, Euclid had shown that mathematics could be developed in deductive manner. No such organised and deductive approach could be said about books VII to IX. So it was that geometry was the model of mathematical rigour rather than arithmetic.

> "In view of the fact that as a consequence of the work of the classical Greeks mathematical results were supposed to be derived deductively from an explicit axiomatic basis, the emergence of an independent arithmetic and algebra with no logical structure of its own raised what became one of the greatest problems in the history of mathematics." (p144, [51])

However, Nichomachus' arithmetic work *Introductio arithmeticae* was, by all accounts, as equally important and influential in its day and beyond as Euclid's *Elements*.

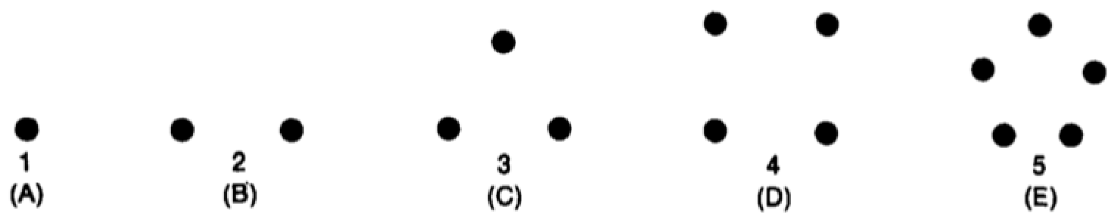
So why did arithmetic fall out of favour, and why did geometry become the paradigm of mathematical epistemology? One reason has been stated in the previous paragraph. Another reason (which we will come to in detail in the next section and provide an answer for in section 6.8) is based on the fact that, to the ancients, the only numbers that existed were integers. Even ratios were composed of integers since ratios were only intended as a means of comparing two integers, and therefore two lines. There was not even a conception of fraction as we understand these today, such as in ¾ meaning three parts out of a whole consisting of 4 parts. So when a geometric configuration was constructed which contained a line whose length could not be compared in integer terms to any other line (in other words the comparison of these two line could not be expressed as a ratio of integers) there was a doubt as to the extent to which numbers and arithmetic could be used as a means of analysing mathematics. This is where the idea of commensurability and incommensurability arise, as we shall now see.

5 Discovering incommensurability

5.1 Commensurability

As we have seen, the Pythagorean considered numbers (i.e. positive integers) as a foundational aspect of mathematics. Their philosophy that "all was number" meant that everything could be measured in terms of integers and their ratios. However, the discovery of incommensurable lines (to be discussed shortly) reversed this emphasis. Furthermore, as a result of Euclid's *Elements* later generations of Greek mathematicians would base all mathematics on geometry and geometric analysis. For these, and future, generations up to the mid 1600s, mathematics was all about geometry. Numbers were, of course, still used in commerce and for practical purposes, but this was not the field of inquiry of the Greek mathematician.

Central to Greek geometry was the concept of *magnitude*. For them there were two types of magnitudes: discrete and continuous. Discrete magnitudes where things that could be counted, such as a collection of pebbles or dots. Discrete magnitudes could also be set into a 1-1 correspondence with natural numbers (represented as figurate numbers). For example, triangle (C) below consists of discrete magnitudes (dots) which then equates to the natural number 3.

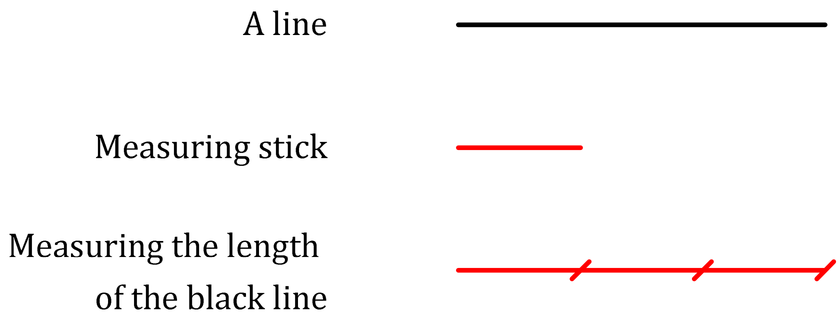


Similarly, one could count the number of lines or planes or solids one had, i.e. "I have 2 lines, 3 triangles and 4 pyramids"

But continuous magnitudes such as lines or planes or solids did not correspond to any number. rather, they dealt with length or area or volume or some other sort of span or extent. For example, although they would agree that the diagram below showed a line of a certain length they would never assign a number to describe its length. So they would never describe this lines as being (say) 3 metres long. It had no numerical measurement.

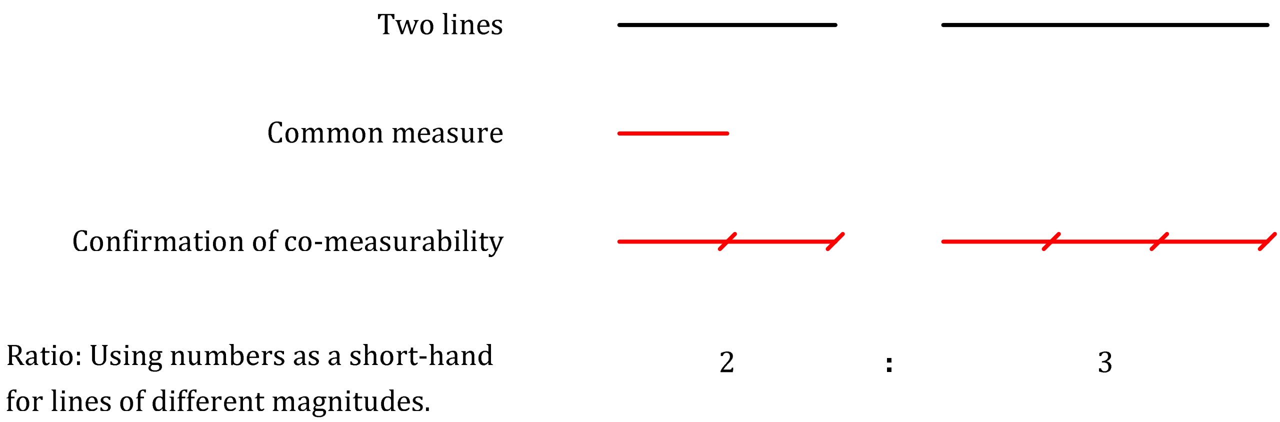
A line

Instead they would measure lines "geometrically" by means of a measuring stick. This measuring stick would act as a reference length against which all lines could be measured. They would then count the number of times the measuring stick would go into the line. Hence for the red measuring stick shown in the diagram below we see that it goes three times into the black line.



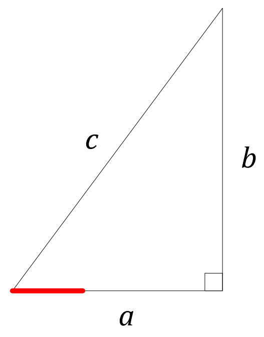
In terms of comparing magnitudes, i.e. the lengths of two different lines (or areas or volumes) the ancient Greeks would then speak of ratios. Today we write ratios numerically such as 2:3. But, to the Greeks this would mean that two lines were being compared not numerically but simply geometrically for their span/extension. In that case what we interpret as "lines in the ratio of 2:3" the Greeks would interpret as a line comprising twice a reference line of a given length, and another line comprising three times the reference line. In other words, the numbers in a ratio represented the comparison of spans or extent of lines (or other geometric objects).

And this is the crux of the matter: In order to compare the length of two lines a reference line was found whose length was common to the two original lines (this idea also applies to finding a common area or volume when comparing areas or volumes), as illustrated below.

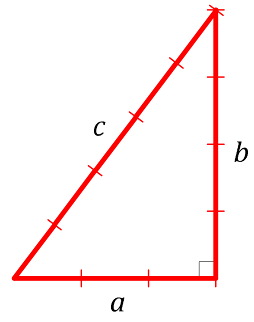


In comparing the ratio of two lines the ancient Greeks spoke in terms of lines being commensurable. The word commensurable comes from the latin: *co-* meaning "together with", and *mensura* meaning "a measuring". So, the two black lines above are commensurable because they have a common measure. Hence, the philosophy that "all is number", i.e. everything could be expressed in integers, implied that all lengths were commensurable.

As another example of commensurability consider the right-triangle shown below in (i):



(i)



(ii)

Let us take the red line of diagram (i) as our reference line or unit measure. Can this be used to measure all the lines of the triangle an integral number of times? Another way of asking this is, does our unit measure fit an integer number of times into the lines of length a, b, and c? By diagram (ii) above we see that it does. Hence, lines a, b, and c are commensurable.

How did the ancient Greeks find the relevant reference/unit line? Well, the idea of ratios arose from the pratice of land measurement. A given line segment s_1 would be measured by laying a unit segment l along it as many times as necessary to cover the whole length of s_1. The unit line l would be constructed so as to fit an integral number of times along s_1. Thus

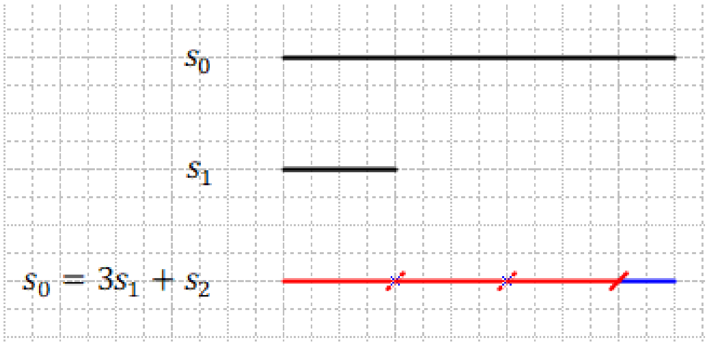
$$s_1 = \underbrace{l + \cdots + l}_{m\ times} = m.l$$

If a second line segment s_2 were such that

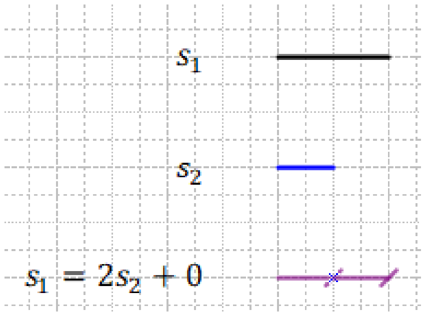
$$s_2 = \underbrace{l + \cdots + l}_{n\ times} = n.l$$

then the ratio $s_1 : s_2$ of the two line segment would be the same as the ratio $m:n$ of the two multiples m and n (see [27], p28 onwards for more). Such an approach might be said to be the arithmetisation of geometric incommensurability. But whereas geometric incommensurability can be represented by a line of finite length it cannot be represented by a finite number.

The process of finding out whether or not two lines were commensurable was by "repeated taking away". Given two line segments s_0 and s_1, where s_0 is longer than s_1, one takes away as many of s_1 from s_0. This then leaves a remaining line segment shorter than s_1. This can be seen in the diagram below where $s_2 < s_1$.



We then take as many of s_2 away from s_1. This may then leave a remaining line segment s_3 where s_3 is shorter than s_2:



One continues this process of taking away as many of the short segments from the longer segments until there are no segments left. The whole process is represented algebraically as

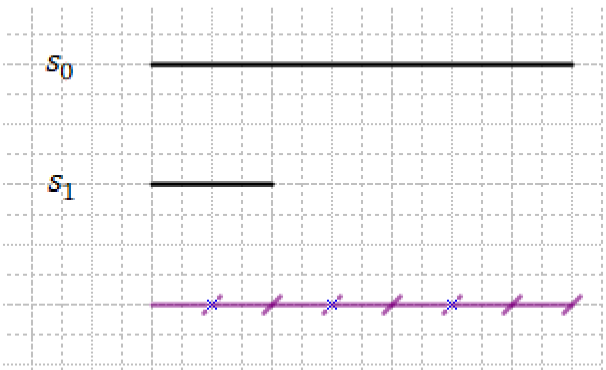
$$s_0 = n_1 s_1 + s_2, \text{ where } s_2 < s_1$$

$$s_1 = n_2 s_2 + s_3, \text{ where } s_3 < s_2$$

$$s_2 = n_3 s_3 + s_4, \text{ where } s_4 < s_3$$

...

where, ultimately, $s_{k-1} = k. s_k$. Then s_k is the common measure for both s_0 and s_1. In the example above we only need to perform this process twice so that $s_1 = 2s_2$ and $s_0 = 7s_2$. Hence, we can say that line segments s_0 and s_1 are commensuarable with s_0 and s_1 being in the ratio 7: 2. This is illustrated in the diagram below.

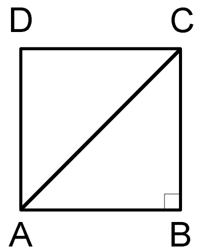


This process of repeated taking away is today called the Euclidean algorithm.

5.2 Incommensurability

Incommensurability is the opposite of commensurability. In other words, there exist two lines whose lengths are such that no reference/unit line will fit into them an integer number of times, *irrespective of how short we make our unit line*. It is the impossibility of finding a unit line of a fixed length, however, small, which is key here.

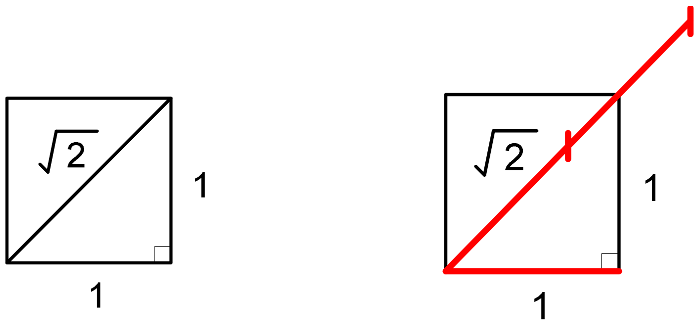
Our two original lines are therefore said to be incommensurable with respect to each other. The classic example of this is illustrated by unit square as shown below.



By Pythagoras' theorem we have

$$AC^2 = AB^2 + BC^2$$
$$= 2AB^2.$$

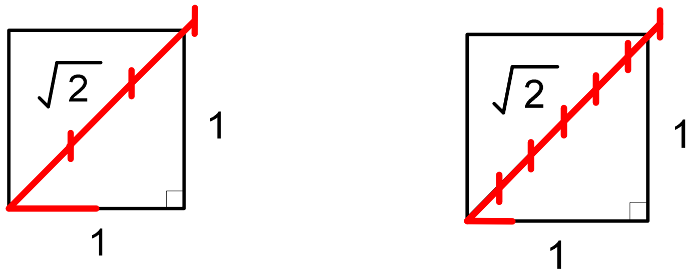
So $AC^2/AB^2 = 2$, implying $AC/AB = \sqrt{2}$. In the language of ratios we have AC : AB :: $\sqrt{2}$: 1. The question now is, Can $\sqrt{2}$ itself be expressed as a ratio $p\!:q$? To find out we try to construct a line which will act as a common measure to the side and diagonal of the square. The first obvious thing to do is to use the side of the square as a common measure. However, in trying to fit this line to the diagonal the side of the square is not commensurable with the diagonal, as shown in diagram (b) below.



(a) (b)

From the perspective of geometry it cannot be denied that the diagonal of the square exists. It also cannot be denied that the number “1” can be used to represent the magnitude of the sides of the square. But we have found in diagram (b) that, geometrically speaking, the side of the square does not fit an integral number of times into the diagonal.

Maybe we can overcome this problem by choosing a smaller unit line, as shown in red in diagrams (c) or (d) below. It turns out that even these unit lines will not fit an integral number of times into the diagonals of the respective squares.



(c) (d)

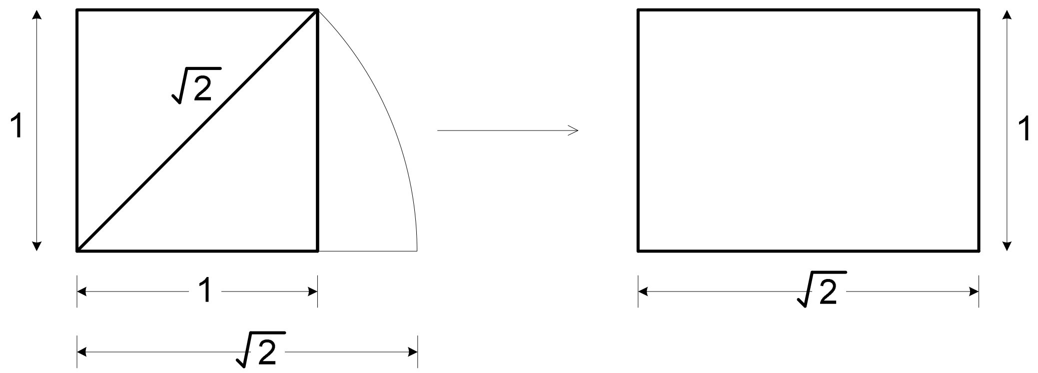
Ultimately, we find that there is no unit line of a finite size, but sufficiently short enough, to fit an integral number of times into the diagonal. As such, the diagonal is incommensurable with the sides of a square.

The modern arithmetic demonstration of this is to assume that $\sqrt{2} = p/q$, where p and q are integers and are co-prime. This means that p/q is in its lowest terms. Then

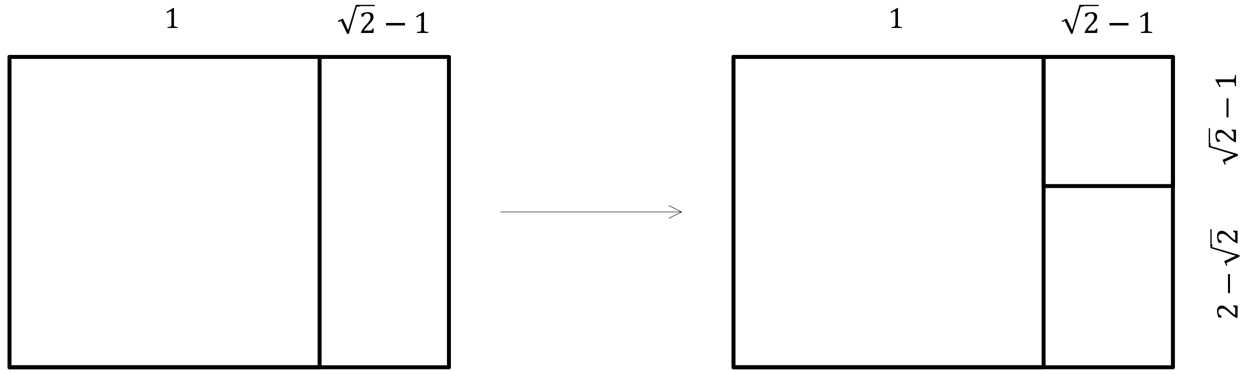
$$2 = \frac{p^2}{q^2} \quad \Longrightarrow \quad p^2 = 2q^2$$

Therefore p^2 is even which implies p is even. But this contradicts our assumption that p/q is in its lowest terms. Another way of looking at this is that p has to be an odd number for p/q to be in lowest terms. But we have just found that p^2, and hence p, is even hence a contradiction. Therefore, $\sqrt{2}$ cannot be written as p/q.

The Pythagoreans did not prove the incommensurability of $\sqrt{2}$ this way. Firstly, $\sqrt{2}$ did not exists as a number so they could not have compared a non-number to numbers p and q. Secondly their analysis of incommensurability would not have been arithmetic. It would, in fact, have been geometric. Here they probably used the division algorithm by repeatedly subtracting shorter line segments from longer line segments in order to find a common measure between line segments. A straightforward approach to geometrically illustrating the incommensurability of $\sqrt{2}$ is as follows: from the unit square, rotate the diagonal onto the horizontal line. From this we can construct a rectangle of sides $\sqrt{2}$ and 1.



We then use the Euclidean algorithm which is to subtract the shorter side from the longer side. We then repeat this process on the next shorter and longer sides:



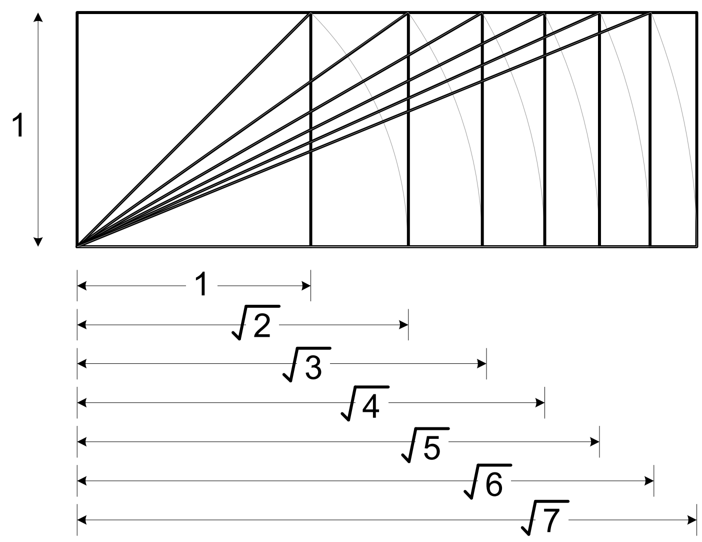
and so on. It so happens that the process of subtraction never ends. It is not that we need to repeat the process a thousand or a million times. It is that we will never get to a point where the shorter side fits into the longer side an integral number of times.

It is interesting to note that the Pythagoreans knew they were able to construct a geometric object where two sides could be represented by natural numbers but where the third side (the diagonal) could not. Given the attitude the Pythagoreans, and other ancient mathematicians, had towards numbers this suggested to them that numbers were not adequate enough to represent certain types of lines. Only geometry could do this. Hence mathematics had to be fundamentally geometric (note that the term "commensurability" is a geometric term for which the arithmetic description is "the ratio of integers". A similar arithmetic description can be given for "incommensurability", namely irrationality. In modern parlance it might then be said that $x^2 = 2$ only had a geometric solution, i.e. the diagonal of a unit square, not an arithmetic solution, i.e. $\sqrt{2}$).

The only reason incommensurability was discovered was because the Pythagoreans had decided that the universe and life could be described in terms of whole numbers only. Without realising it they had made a distinction between whole numbers and other types of as yet unknown numbers (the fractions and the reals) simply by the act of philosophising about the nature and importance of the class of numbers called whole numbers. Without this philosophising the distinction between commensurables and incommensurables would not have been brought into sharp relief. Whole numbers therefore acted as a contrast which allowed incommensurable magnitudes, and ultimately irrational numbers, to be identified.

The concept incommensurability was fairly easily accepted by the ancient Greeks. It was obvious geometrically that the diagonal line of a unit square existed and could easily be constructed. But to achieve commensurability between the diagonal and one side would require endlessly repeating the process of subdividing lines resulting in the commensurable line being infinitely short. But infinite subdivisions and infinitely short lines were anathema to the ancient Greeks and had no part to play in their mathematics. This lead to the Greeks excluding infinite processes and the infinitely small from their mathematics.

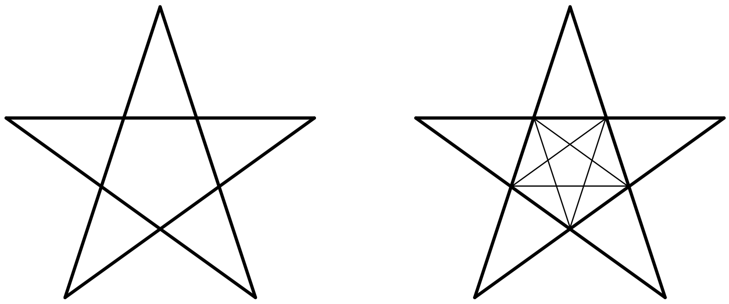
Note, however, that it is easy to geometrically construct the roots of all integer numbers, as illustrated in the diagram below. From our modern perspective the root of a perfect square gives an integer. Since integers were accepted by the Greeks as numbers it begs the question about whether or not the other roots should also be classified as numbers. But the Greeks didn't see it this way. All they saw was geometry, i.e. line segments. So, in the diagram below try looking at the diagonal lines simply as lines. Try not to numericalise them as "so and so metres long", i.e. don't give it a measurement. From this perspective the Greeks had no reason to worry about not being able to deal with what we call irrational numbers since the inherent nature of any line (or curve or plane of volume) was not numeric but geometric.



It is commonly repeated that ancient Greek mathematics underwent a crisis with the discovery by the early Pythagoreans (6^{th} – 4^{th} centuries BC) of the incommensurability of the diagonal of the unit square. But separate studies by Knorr [53], Fowler [30], and von Fritz [31] of the source material from Pythagorean and ancient Greek times suggests otherwise. The Greeks simply got over the surprise of finding incommensurable line segments and the inability to express these as numerical ratios. And by focusing on geometry rather than numbers there was no distinction between 1 and $\sqrt{2}$ since they shared the essential feature of both being line segments. However, if 1 and $\sqrt{2}$ were to be considered as numbers there would be no commonality between these two objects since 1 is what we now call rational and $\sqrt{2}$ is irrational. Furthermore, since the Greeks were subsequently able to enlarge their view of mensuration by developing a geometric theory of incommensurable magnitudes (see section 5.3) only geometry was equipped to deal with things such as surds. If surds did need to be considered numerically they were only ever considered approximately, to whatever accuracy they desired. And when you deal with approximations you have reduced a surd, an irrational number with infinite decimalisation, to a rational number with finite decimalisation, i.e. you are dealing in ratios.

The historical records are contradictory when it comes to finding out how incommensurability was discovered. One story (see [53]) suggests that this was discovered when attempting to double the square. This involves taking a square whose sides and area are integers (such a square of side 2 and area 4), doubling the area to 8 and determining if this new square has integer sides. In this case the answer is no, and in general it is not possible to double the integer area of a square with integer sides in order to get another square of integer area and integer sides.

Another story (see [31]) suggested that it is the Pythagorean Hippasus (approx. 470 B.C.) who discovered the incommensurability via the study of the pentagram (shown as the left diagram below). The central portion is a pentagon, from which another pentagram can be constructed by joining the corners of the pentagon (shown as the left diagram below).



Inside the smaller pentagram lies another regular pentagon implying that a third pentagram can be constructed, and so on for all smaller pentagons/pentagrams therein. Hippasus was then supposed to have wanted to show the commensurability of certain larger line segments to certain smaller line segments of any given pentagram. In doing so he arrived at a contradiction that a line segment would have to be shorter than itself, thus showing that his original assumption of commensurability was incorrect (see [17] for details).

However incommensurability was discovered the surprise caused by this must have been short lived since, in general, the ancient Greeks were not troubled by incommensurability either geometrically or arithmetically. And, they quickly became used to accepting numeric approximations to incommensurable magnitudes.

5.3 Eudoxus' theory of incommensurability – To come

6 How the issue of infinity arises naturally in geometric construction

6.1 The conception of infinity in ancient Greeks times

The objects of study in geometry were always finite in length, divided at finitely many points. The Greeks were aware of the concept of the infinitely small by the fact that a line could be indefinitely divided and they were also aware of the infinitely large by the fact that a line could be extended indefinitely. Euclid knew this and wanted to avoid addressing infinity (in the sense of a line being infinitely long) in his definitions of lines. To do this he defined a *line segment* (something finite) as distinct from a *line* (something infinite). But before he could define a line segment he first had to define a point. Hence (paraphrasing), Euclid defined

- a point to be that which has no part;
- a line segment to be that which is obtained by drawing a line between two points;
- a line to be that in which the line segment is produced(/extended) continuously in a straight line.

There is no hint in the last statement that the process of extending the line segment should carry on forever, simply that the line can be extended in a continuous manner. This hidden implication is that such a line can be extended indefinitely, but this implication is ignored.

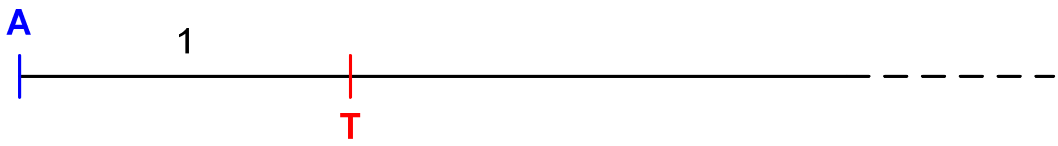
The infinite divisibility of a line segment of a line segment is also ignored/avoided simply by accepting that, for practical purposes, lines are divided only a finite number of times.

Later on, in section 7, we will see how we can construct points, line segments, arcs of circles and circles using straight-edge and compass. We shall also see how we can construct geometric figures such as triangles and other polygons, etc. We shall also see how we can perform geometric arithmetic using straight-edge and compass, in other words how to add, subtract, multiply, divide, and even take square roots of line segments, again using only a straight-edge and compass. But how do we use these instruments to construct an infinite number of line segments, or to construct line segments which are infinitely small? We can't hence, one of the assumptions in constructing geometric objects is that all constructions have to be performed in a finite number of steps, in finite time.

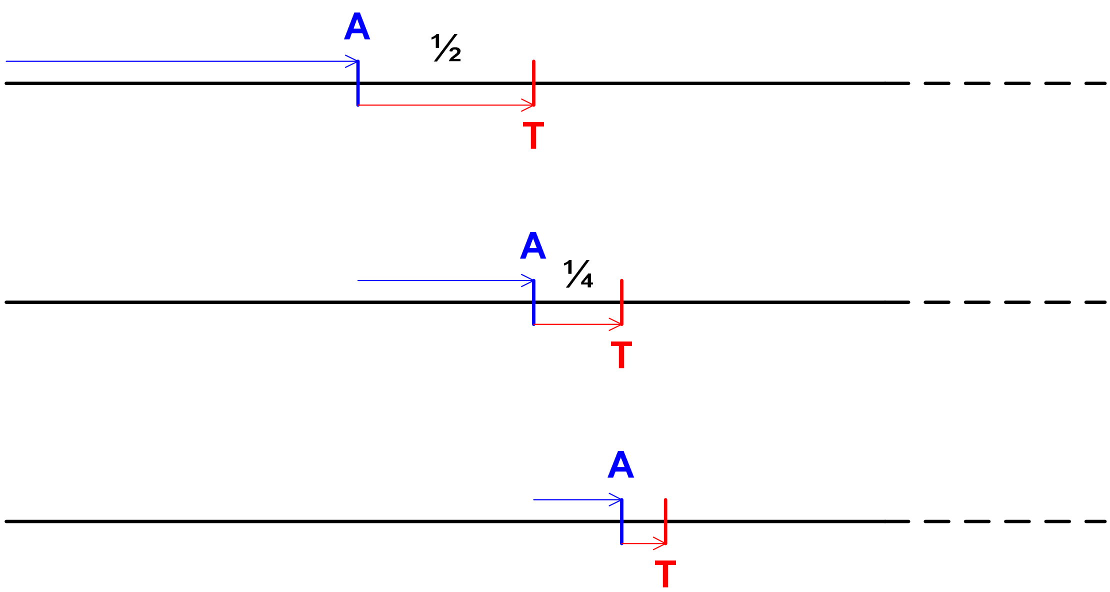
In hindsight we can see incommensurability as a form of infinity. For example, when dealing with a unit square we have seen that its diagonal is incommensurable with respect to its sides. What this means at a practical level is that we have to subdivide a side of the square into an infinite number of infinitely small segments in order to find a unit line which would fit an integral number of times into the diagonal. This change in perspective from the idea of incommensurability to the infinitely large and infinitely small would come to dominate mathematical thinking from the 16th century onwards. But before we come to this we look at the state of play in understanding the nature of infinity in ancient Greek times.

6.2 Achilles and the tortoise

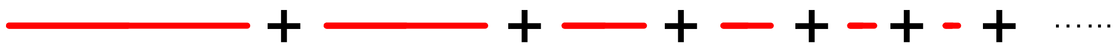
The first difficulties in how to understand infinity were recorded by Zeno (~495 BC – ~430 BC). Zeno, a contemporary of Pythagoras, questioned the possibility of infinite divisibility of space and time. He devised a number of famous paradoxes illustrating the problems when applying infinite processes. One of the more famous paradoxes resulting from the idea of infinitesimals and infinities is that of Zeno's story about the race between Achilles and a tortoise. Let stage 0 of the race be such that the tortoise (T) has a 1 metre head start on Achilles (A), as illustrated below



If Achilles runs twice as fast as the tortoise then stage 1 of the race is where Achilles has covered 1m. But then the tortoise will have moved forward ½m. By stage 2 Achilles will have covered ½m and the tortoise will have moved forward again by ¼m. Since this process can be repeated for ever the tortoise will always be in front of Achilles, and Achilles will never catch the tortoise.



So, in chopping up the distance covered by Achilles into discrete pieces we find that the issue relates to adding up an infinite number of ever decreasing, and ultimately, infinitely small distances. The question now is, How does one add up an infinite number of infinitely small distances? Geometrically we can represent this as an infinite number of ever decreasing lines, viz



Does this sum give a line infinitely long? On the other hand, if the last line is infinitely short does this mean that this last line has zero length? If so, we have an extremely long line of finite length. The problem with this is that it would take an infinite amount of time to get to the infinitely short line segment, so even this line is finite it would take forever to get to the end of it. It is impossible to get an answer to this from a geometric perspective since there is no geometric form of analysis by which will allow us to deal with infinitesmimals.

Neither Zeno nor his contemporaries believed that Achilles would never overtake the tortoise. He simply used this example to illustrate the logical paradox of having an infinite number of distances which ultimately became infinitely small. Clearly Achilles would overtake the tortoise in a finite amount of time and over a finite distance. But the Greek of the day weren't able to offer an explanation as to the contradiction between reality and Zeno's logic. This from Hogben (p17, 1936):

> "If we go on pilling up bigger and bigger quantities, the pile goes on growing more rapidly without any end as long as we go on adding more. If we can go on adding larger and larger quantities indefinitely without coming to a stop, it seemed to Zeno's contemporaries that we ought to be able to go on adding smaller and still smaller quantities indefinitely without reaching a limit. They thought that in one case the pile goes on for ever, growing more rapidly, and in the other it goes on for ever, growing more slowly. There was nothing in their number language to suggest that when the engine slows beyond a certain point, it chocks off."

Note that the situation above can be represented arithmetically by the series

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots \qquad (*)$$

The Greeks knew about this type of series (although they didn't express it numerically). They knew that, despite the fact that the series carries on forever, a final answer could be obtained if the next fraction along was at least half that of the previous fraction. And they also knew that if this was not the case they would not obtain a final answer.

Another type of series which carries on forever is

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots \qquad (**)$$

Today we know that (*) converges and that (**) diverges, and we have the mathematical means to show this.

6.3 The conception of infinity in ancient Greeks times

It would take about 100 years after Zeno for someone, namely Aristotle (384 BC – 322 BC), to give the first viable explanation of what was happening in a situation as above. Aristotle posited two types of infinity: a *potential* infinity and an *actual* infinity. The former type of infinity can be understood as boundlessness, endlessness, or unlimitedness, something for which there is always more to come. A numerical example of this is

$$1+2+3+4+5+\cdots$$

By continually adding 1 to this series we increase the sum without end. For Aristotle it is the fact of a continued, unending addition that defines a potential infinity, an infinity that is forever approached but never reached, and for which there is no known number at the end of the series (and for which "the end of the series" does not make sense).

The actual infinity can be thought of as completeness, wholeness, unity, absoluteness, and is best illustrated by the continued halving of segments of a unit line. Here, one of the line segments which has been cut is again cut in half, this process being repeated indefinitely (this is the case with Achilles and the tortoise). In this case the numerical series is

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots$$

Since the original line has length 1, the sum of these line segments add up to 1 even though there are an infinite number of line segments. Aristotle called this type of infinity an actual infinity. It is the type of infinity which could be "completed" in order to produce an actual result.

However, such a distinction between two types of infinities raised problems about how/when a potential infinity can become an actual infinity. For example,

- how does one add up an infinite number of lines of ever decreasing length? Does such a sum give a line infinitely long? Clearly not since, in the case of the unit line, we know we started with a line of finite length.

- On the other hand, if the "last" line segment is infinitely short does this mean that its length is zero? If so, we have an extremely long line but of a finite length.
- Also, a line of length 1 is a line of finite length whose end we arrive at in a finite amount of time. But the process of adding ever shorter segments of line according to the principle of continual halving requires an infinite number of line segments and an infinite amount of time to complete. So it seems as if the process by which we arrive at a finite line in finite time requires an infinite number of lines and an infinite amount of time to complete.

Returning to the idea of a unit square it is clear that its diagonal exists and is of finite length. So, geometrically the diagonal can be considered a completed or actual infinity. On the other hand, the representation of the diagonal as a number has to be considered as a potential infinity since the process of calculating $\sqrt{2}$ via repeated subtraction (the Euclidean algorithm) does not have an end. It was because of the inability of the Greeks to conceive of numbers other than integers that lead them to this impasse, and ultimately to decide that *numbers were not sufficient to represent mathematics*. There were numbers and number ratios which could represent geometric objects, such as the 3-4-5 triangle, but there were geometric objects which could not be represented numerically such as the diagonal of a unit square. Such were the confusions surrounding infinity and infinitesimals.

As a result, the Greeks decided to banish infinity and infinite processes from their mathematics, and work only in finite terms. This meant that only a numerical approximation could be given to incommensurable magnitudes. And since Euclid's *Elements* did such a good job at demonstrating geometric truths deductively and rigorously it overshadowed his books on number and arithmetic for which there was no equivalent rigour. This led the ancient Greek mathematicians to make a critical decision: *they believed that geometry was more general and fundamental to mathematics than number and arithmetic*.

The attitude that geometry was the foundation of mathematics was to persist until the beginning of the 1800s. It was only then that rigorous mathematical analyses and processes were being developed to answer the problems about how to deal with infinity and infinitesimals. Today we have the mathematical conception and techniques to deal with infinite series where such mathematics is, in fact, an arithmetic analysis not a geometric analysis (or even an algebraic analysis). But because Aristotle's conception of potential/actual infinity dominated mathematics for 2000+ years, such an arithmetic analysis would have to wait until the mid to late 1800s when the foundations of arithmetic were put on a firm footing.

6.4 On points and lines

In the examples above we have seen that points are constructed by the intersection of line or curve segments. But by Euclid's axioms a line segment is the joining of two points. So which comes first? The point or the line? If we are to develop geometry logically from the most primitive elements to the most complicated objects we need to state these primitive elements as being already there given to us. In other words, those primitive elements are objects which are not constructed but somehow already exist.

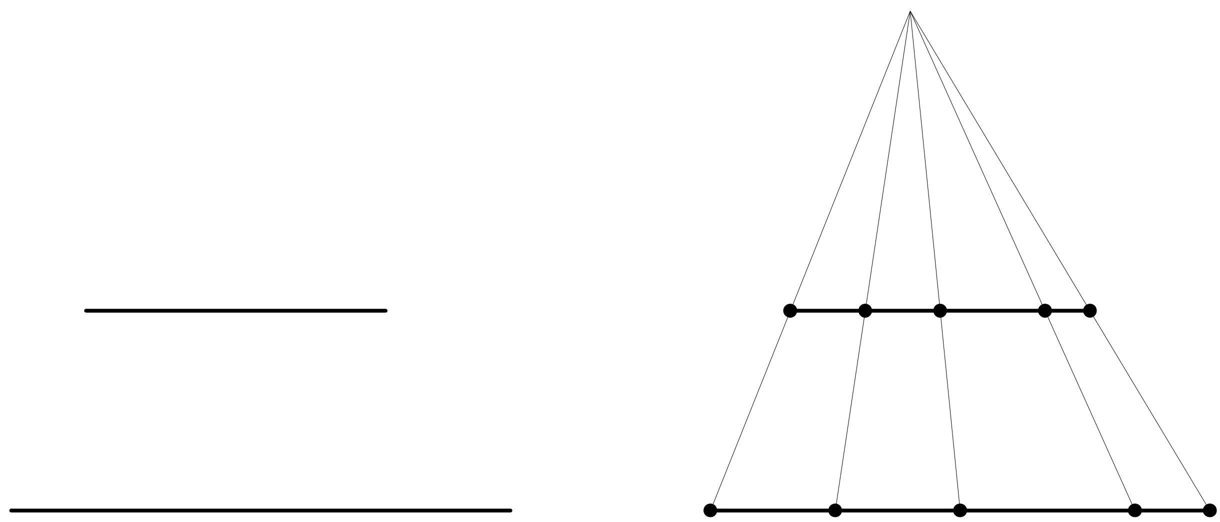
In Book I of his *Elements* Euclid defines a line segment as the joining of two points. So points exist prior to the line segment. This is illustrated in the diagram below. As such he uses his definition of a point as an axiom. We have to start somewhere and it seems logical that there is no more basic an object as a point. So we accept our first point, or possibly our first two points, as already existing and proceed to construct geometry from there (this is akin to Peano's axioms of arithmetic where we start by assuming the number 1 as already given to us, to then construct the rest of the natural numbers by some mathematical/logical process).

If a line segment is something which can be drawn between two points, then the end of the line segment represents points. This means that every time you cut a line segment you create a point where there was none before. Since the attitude of the ancient Greeks was that space was infinitely divisible a line could, in theory, be cut into an infinite number of points, hence the belief that a line was made of an infinite number of points.

But we now have a problem. A line segment contains two end points. If we divide this line segment into two parts we obtain two shorter lines each with two end points. Repeatedly dividing these lines will produce ever shorter lines until we end up with infinitely short lines, say *linelets*. So the original line segment is composed of an infinite number of infinitely short lines. But each linelet, being a line segment, has two end points. So it seems the original line segment has twice the number of points as it does linelets, i.e. twice an infinite number of points as linelets. Furthermore, it would require an infinite number of geometric operations to bring all the points into existence. This is the nightmare of infinity faced by the Greeks (and mathematicians right up to Cantor (1845 – 1918)).

6.5 How a short line equals a long line

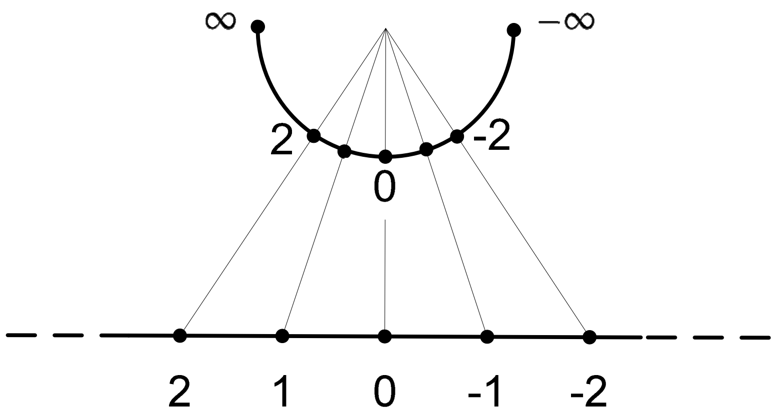
The problem caused by the belief that a line is made up of an infinite number of points can also be seen if we consider the diagrams below. In diagram (i) the two lines are clearly of different lengths. However, every point on the short line can be associated with one corresponding point on the long line. Since both lines have an infinite number of points they both have the same number of points. Hence both lines are of the same length! Clearly this is not the case, and the cause of our problem is to fall into the trap of thinking about infinity in finite terms. It is like comparing apples and oranges: two lines of different finite size being compared with an infinite number of points of infinitesimal size.



(i) (ii)

6.5.1 How a semi-circular arc of finite length equals an infinite line

Another example: below we see a semi-circular line of finite length, and a straight line of infinite length. By equating each point of the curve with each point of the line we see that a line of finite length is as long as a line of infinite length.



Clearly the semi-circle is finite in length, and clearly the line is of infinite length. But also it is clear that there is a one-to-one correspondence between each point on the line and each point on the semi-circle. One could therefore say that the semi-circle represents the completed infinity version of the potential infinity of the line. In other words, we have devised a way of mapping the potential infinity of the line onto the circle so as to construct an actual/completed infinity. We can then assign a terminating value (the end points of the semi-circle) to the infinities of the line.

The problem caused by one-to-one correspondence also applies to infinite sets of numbers. For example, take the set ℕ of natural numbers:

$$\mathbb{N} = \{1, 2, 3, 4, 5, 6, \dots\}.$$

Now take the set 𝕊 of the squares of those numbers

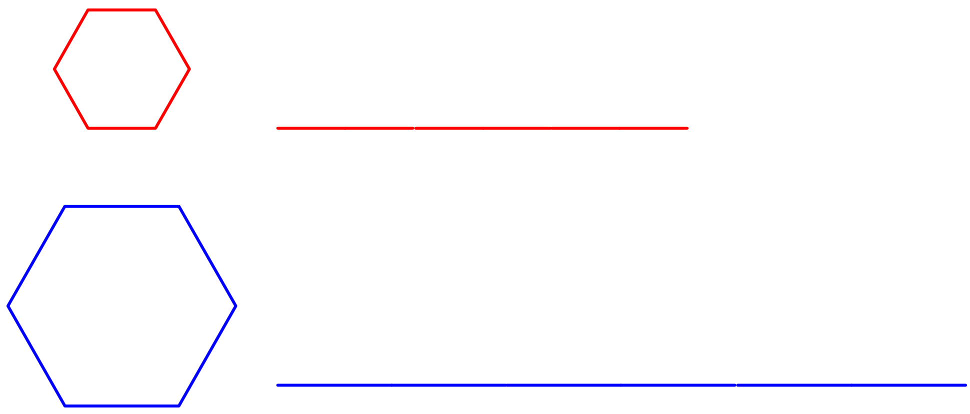
$$\mathbb{S} = \{1, 4, 9, 14, 25, 36, \dots\}.$$

It looks like there are significantly fewer numbers in 𝕊 then in ℕ (particularly since the squares get further apart as we go further along the sequence). But we can place each number in ℕ into a 1-1 correspondence with each number in 𝕊 which means that there are the same number of elements in both sets. Contradiction? The problem here is that we are trying to model the continuous (i.e. lines and curves) by the discrete (i.e. the counting numbers 1, 2, 3, 4…). As Galileo said

When we attempt, with our finite minds to discuss the infinite, assigning to it those properties which we give to the finite and limited; but this I think is wrong, for we cannot speak of infinite quantities as being the one greater or less than or equal to another. One line does not contain more or less or just as many points as another, but each line contains an infinite number. These are the problems which can be created when we try to think intuitively about infinity and the infinitely small.

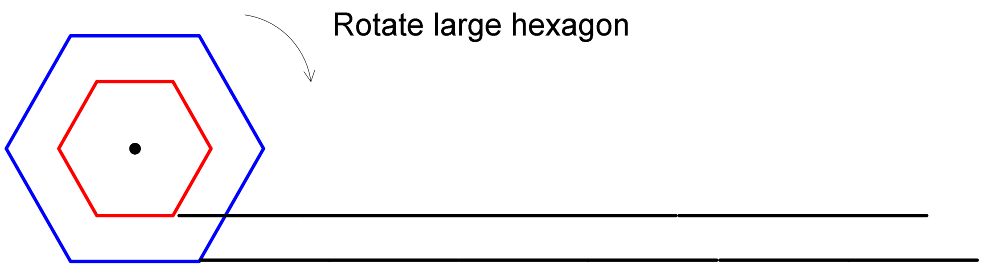
6.6 Galileo's hexagon wheel

Imagine two regular hexagons with one hexagon smaller than the other. Clearly the perimeter of the small hexagon is less than the perimeter of the large hexagon. It stands to reason that if we rotate each hexagon by one revolution the resulting straight-line motion (which equals the length of the perimeter) of the smaller hexagon will be shorter than that of the large hex.



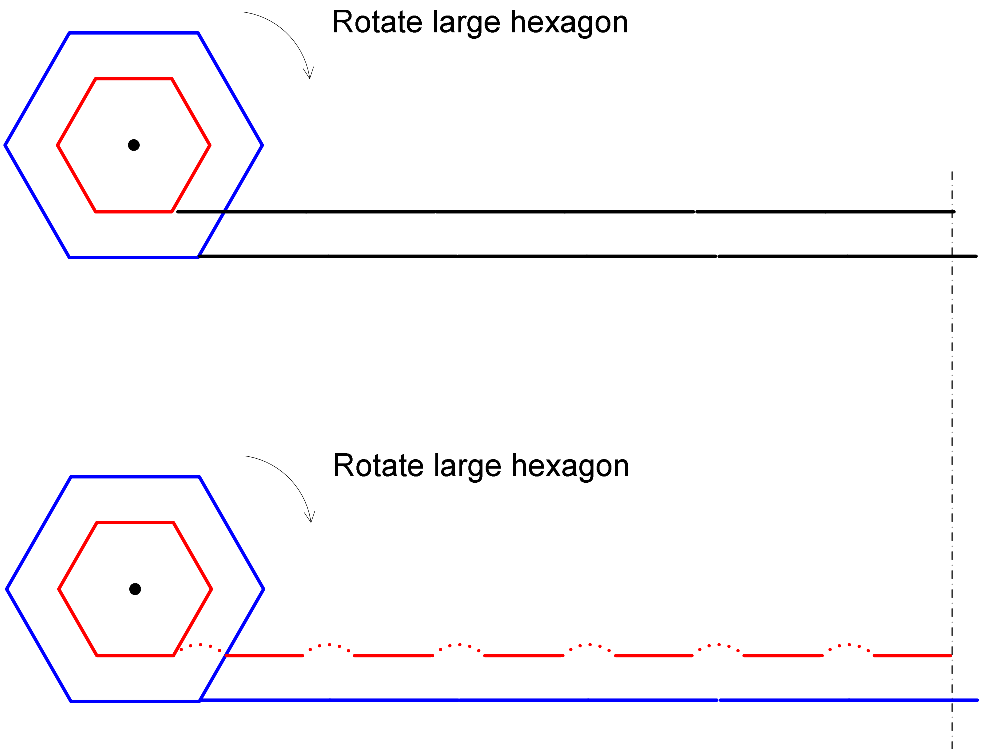
Hexagons *Length of path traced out by one rotation of each hexagon*

Now consider these hexagons attached at their centre. When the large hexagon is made to rotate one complete rotation it will cover the distance of its perimeter. However, since the small hexagon is attached to the large hexagon the former will also cover the same linear distance as the large hexagon. So we have the contradiction that the length of path traced out by the small hexagon is longer than its perimeter.



How do we explain the extra distance covered by the small hexagon? This can be done by performing the physical experimental version of this situation. See youtube's "Aristotle's Wheel Paradox - To Infinity and Beyond" at https://www.youtube.com/watch?v=mrVg9GM5h7Q

What we find is that when we rotate the large hexagon wheel on a level surface we produce a continuous blue line on the level surface. But in performing this rotation the small hexagon, which rests on its own level surface, is lifted off its level surface. The resulting effect is to leave gaps between the red segments representing one side of the small hexagon, as shown in the diagram below.



So the small hexagon does indeed cover a linear distance given by the perimeter of the large hexagon, but the small hexagon is not always in contact with its own the level surface. These gaps are actually the difference in the length of the sides of the large and small hexagons.

We can now ask two questions:

1) What happens to the gaps if we increase the size of the small hexagon whilst still keeping it smaller that the large hexagon?
2) What happens if we use a triangle wheel? A square wheel? A pentagon wheel? A septagon wheel? An octogon shaped wheel? Etc.

In answer to 1) we have that, depending on the relative sizes of the small and large hexagon, the linear distances covered by both hexagons will be approximately equal (but not exactly equal) as illustrated in the diagram above.

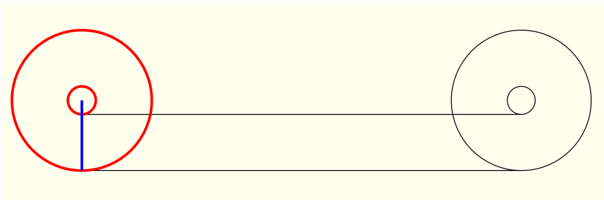
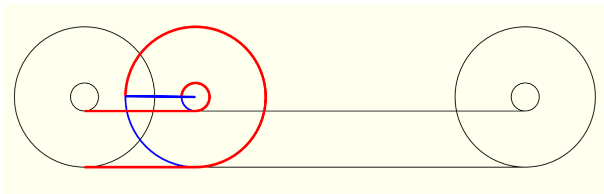
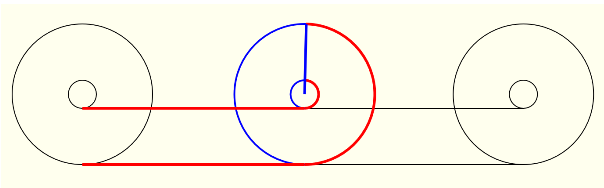
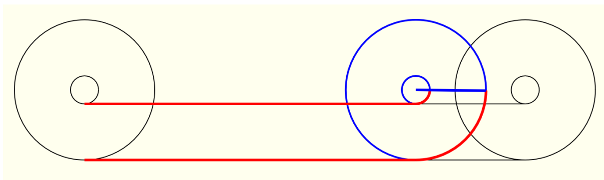
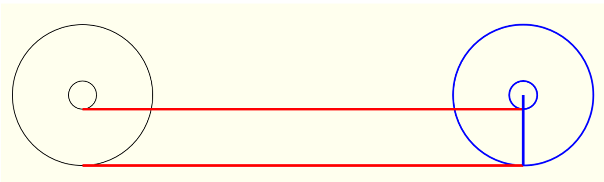
*(*diagram here*)*

And in answer to 2) we have that the smaller polygon will cover the same linear distance as the larger polygon, and for the same reason. But as we decrease the number of sides of our polygon wheel so there will be larger and larger gaps between consecutive solid lines of the smaller polygon; and as we increase the number of sides of the polygon wheel so there will be larger and larger gaps between consecutive solid lines of the smaller polygon, with the distance between each blue strip being shorter and shorter. This is illustrated in the diagram above for a square wheel and an octogon wheel where the hexagon wheel has also been included for comparison

*(*diagram here*)*

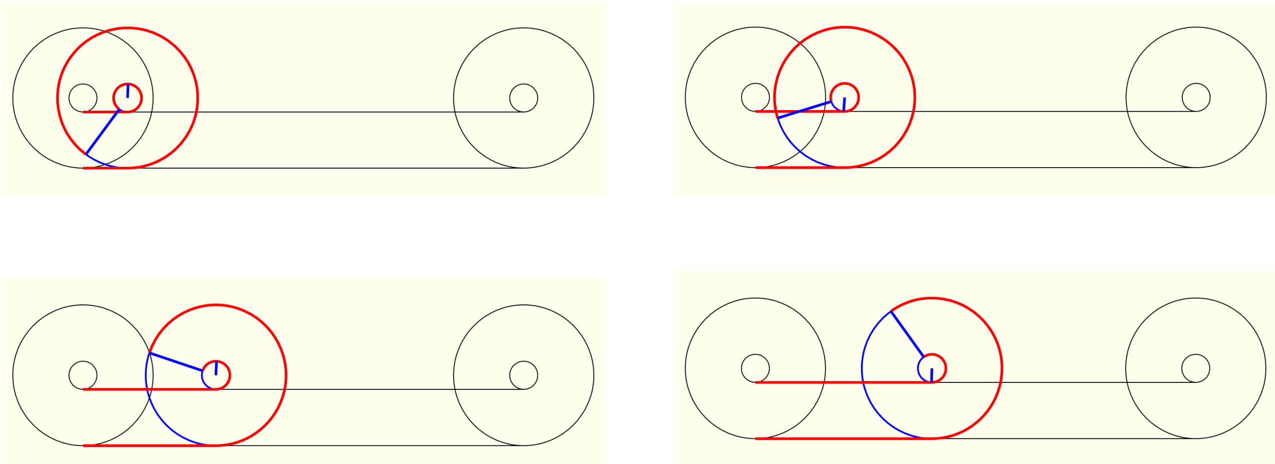
Ultimately we can ask, What if we have a polygons with an infinite number of sides, in other words a circle? What if we rotate two concentric circles, one circle being smaller than the other, the circumference of the inner circle is clearly shorter than that of the outer circle. The smaller circle should therefore cover a shorter linear distance than the larger circle.

But as we see from the sequence of diagrams below that this is not the case (images taken from http://thewessens.net/ClassroomApps/Main/aristotle.html).

So it again seems as if the smaller circle has the same circumference as the larger circle. So how do we now explain this paradox? One way is to appeal to our experience with the hexagon wheel, and other polygon wheels. Then, as we let the number of sides of the polygon approach infinity we obtain a circle. But this is where the problem of infinity raises its head again: in this case logic seems to suggest that we will have an infinite number of blue strips, with an infinite number of infinitely small gaps between each strip.

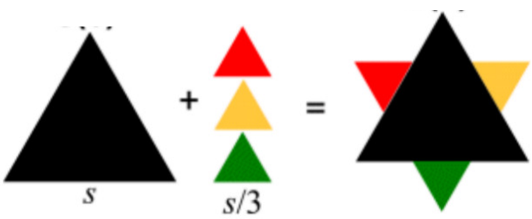
The actual (physical) answer to this is that smaller wheel slips or slides along the level surface as the larger wheel rotates. In other words, the inner wheel is being dragged along for some of its journey. To stop this sliding or dragging along we can decouple the inner wheel from the outer wheel. Then, for the inner wheel to cover the same linear distance as the outer wheel, the inner wheel will have to travel at a greater speed, thus covering several rotations for one rotation of the larger wheel.

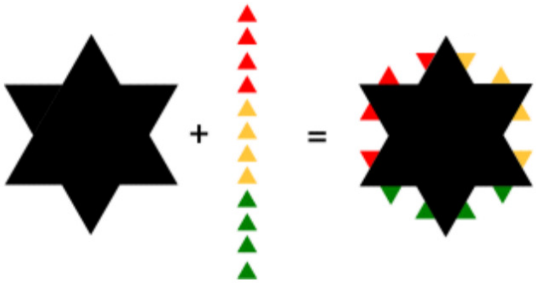


In the case of the hexagon wheel it isn't a situation of slipping/sliding but of jumping a gap that explains the equality of linear distances covered. And as the number of sides of the polygon wheel increases this jumping off the level surface of the inner polygon becomes less and less (i.e. the jumps become smaller and smaller) until ultimately we have an infinitesimal jump, which is translates into a sliding/slipping.

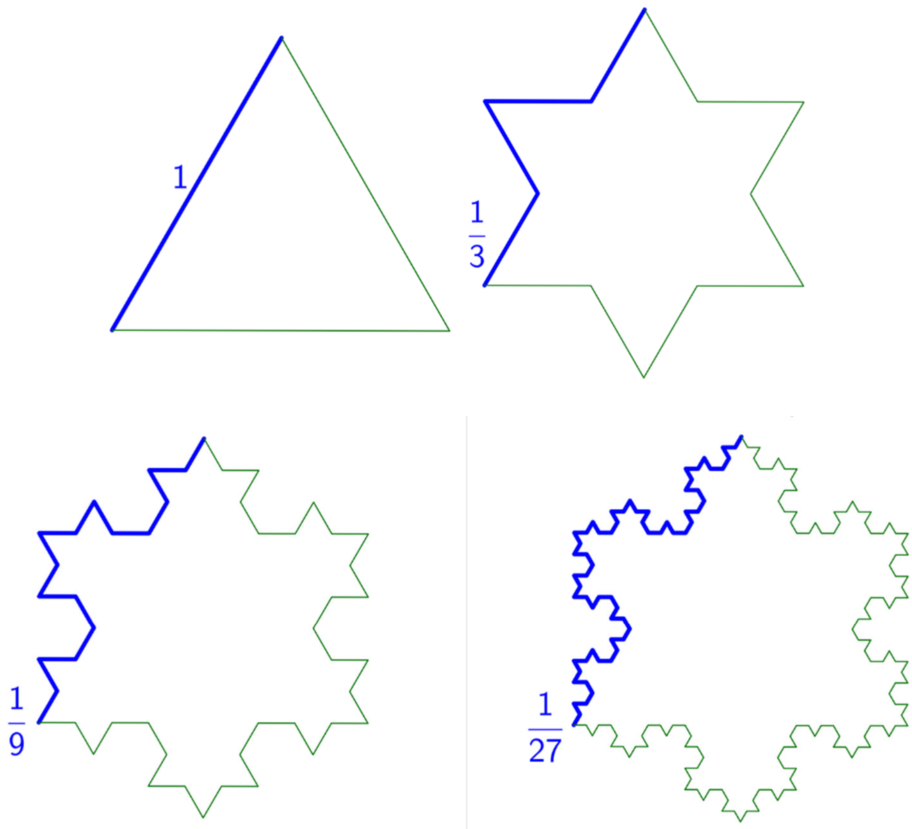
6.6.1 The Koch snowflake

As yet another example we have the Koch curve/snowflake. Here we start with an equilateral triangle of sides 1. Onto the mid-segment of each side we attach another (smaller) equilateral triangle. Onto the mid-segment of each side of these smaller triangles we attach yet another (smaller still) equilateral triangle, and so ad infinitum. The resulting shape is the Koch curve/snowflake below:





From a geometric point of view the process of adding ever more triangles to each side changes the perimeter in the manner illustrated below.



The aim here is to find the total perimeter and the total area of this shape. Intuitively, both the perimeter and the area should be finite since the perimeter is simply one cycle around the shape (ending up at our starting point must mean that the perimeter is finite), and the area is enclosed within a finite perimeter so the area must be finite also. As we shall see later, the area of the Koch snowflake is indeed finite but the perimeter will be seen to be infinite.

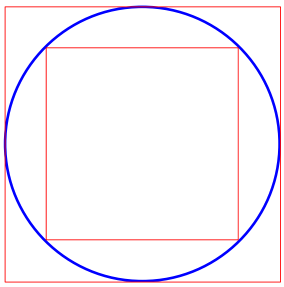
These examples highlight the misconception and difficulties the ancient Greeks had in dealing intuitively with infinity and infinitesimals. Their mindset was to see infinity as if it was a number, i.e. a quantity of a fixed (albeit huge) size. This then led them to misconceive arithmetic performed on geometric objects such as when comparing a finite number of points on two lines segments with the infinite number of points on these lines (see sections 6.4, 6.5, and 6.5.1). For example, a circular arc of finite length equates to a line of infinite length because every point on the arc maps to every point on the line (see sections 6.5.1).

The problem was that there was no geometric way of analysing the behaviour of an infinite number of points on a line. There was no geometric analysis which could distinguish between an infinite number of points on a short line with an infinite number of points on a longer line. So people ended up relying on intuition when solving geometric problems.

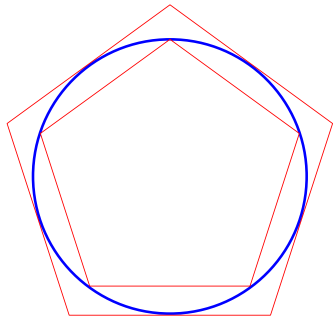
6.7 Archimedes' approach to finding π

To see what it takes to work only with proportional/commensurable line segments and their integer ratios let us look at the way in which Archimedes' calculated an approximation to π (see p194 onwards of [36] for the full account, or see *www.nonagon.org/ExLibris/archimedes-pi* or [5] for modern versions. See also p200 and p254 of [12], or p10 of [38] for alternative modern approaches).

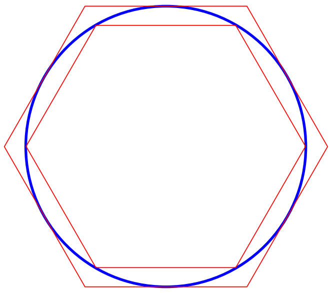
Archimedes knew of the idea of bounding an object as a way of getting ever better approximations to things he was trying to calculate. As such he applied this idea to finding an approximate value for π by inscribing and circumscribing a circle using regular polygons. Archimedes also knew that the greater the number of sides to a regular polygon the more accurate would be the answer to whatever it was he was calculating This increased accuracy can be inferred from the sequence of diagrams below.



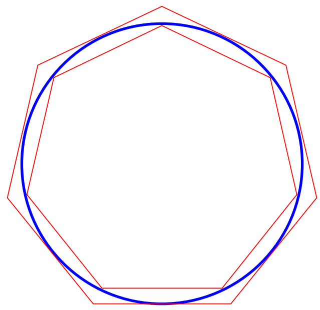
(1): Inscribed and circumscribed squares



(2): Inscribed and circumscribed pentagons



(3): Inscribed and circumscribed hexagons



(4): Inscribed and circumscribed septagons

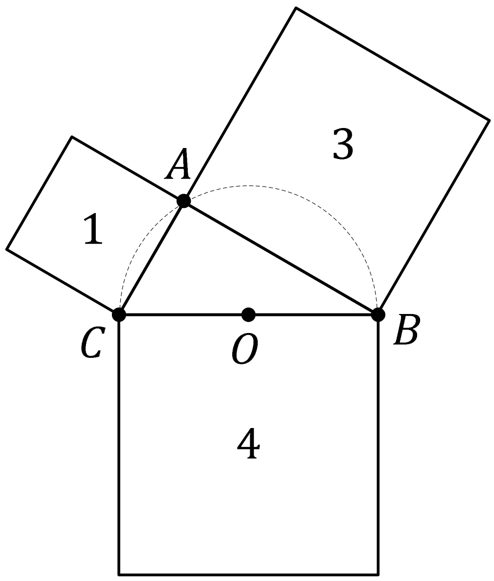
Inscribed and circumscribed polygons give better approximations
As the number of sides of the polygon increases

He used the idea of bounding polygons to find better lower and upper approximations to the value of π, to the area of a circle, and to the area under a parabolic curve. In using this approach Archimedes was able to avoid having to address the difficulties associated with infinities and infinitesimals. As a specific example we will focus on how Archimedes found an approximate value for π. His analysis is not so much difficult as it is elaborate since it involves a significant number of comparisons of line lengths between sides of triangles (all expressed as ratios of integers).

To start with Archimedes knew about what today we call $30° - 60° - 90°$ triangles. He also knew that a hexagon was composed of such triangles in such a way that they formed six equilateral triangles. The approach he used was to start with a $30° - 60° - 90°$ triangles, bisect one of its angles, and then set up relevant ratios of sides of this triangle in order to find an initial approximation to π.

The approach to bisecting one of the angles of the triangle would have the effect of doubling the number of sides of the inscribed/circumscribed polygons, so that starting from a hexagon and repeating his analysis of bisecting angles and setting up the ratio of relevant sides of a triangle, Archimedes ended up with a 96-sided polygon allowing him to obtain π to a certain degree of accuracy. There he stopped, ending up with the lower bound of $3\frac{10}{71}$ and the upper bound of $3\frac{1}{7}$ for π.

We now go through a portion of his analysis. To start with, Archimedes used an approximation to $\sqrt{3}$. No one knows how he obtained his approximation (since most of his works are lost) but he did know about the ratio of areas of squares constructed on the sides of a $30° - 60° - 90°$ triangle (see figure below). To see this ratio consider a circle of radius $OC = 1$. Construct a line segment $CA = 1$ as illustrated in the diagram below. This means that $OA = 1$ (by side-angle-side theory of triangle geometry within a circle) and the triangle OCA is one of the equilateral triangles of a hexagon.



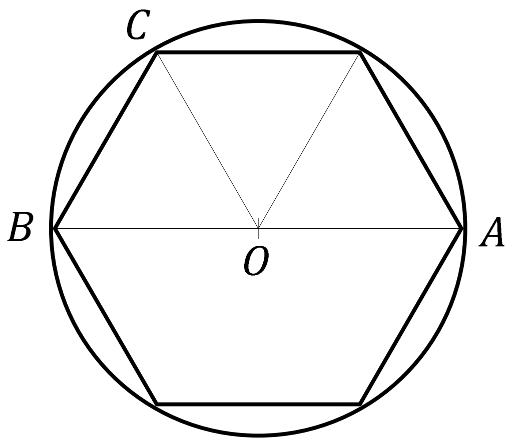
Ratio $BC : CA$ is $2 : 1$. In modern terms we also know the ratio $BA : CA = \sqrt{3} : 1$ (by Pythagoras' theorem). But since irrational numbers did not exist in Archimedes' time he instead used

$$265 : 153 \text{ as a lower bound for } \sqrt{3}$$

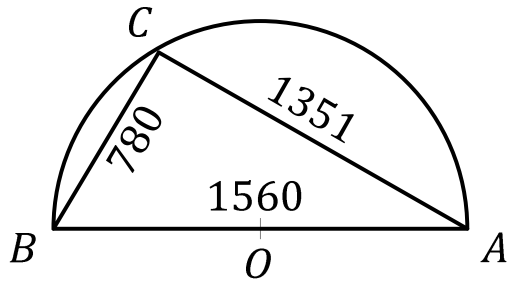
and

$$1351 : 780 \text{ as an upper bound } \sqrt{3}$$

So, starting with a hexagon and letting the diameter of the circle be 1560, which is the value used by Archimedes, we can inscribe six equilateral triangles in a circle, three such of which are illustrated in diagram (i) below.



(i)



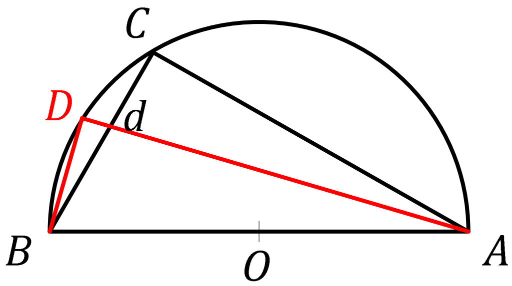
(ii)

Since $AB = 1560$ and all internal triangles are equilateral, we have $OB = BC = 780$. So, an initial approximation to π can be found by dividing the perimeter of the hexagon, now given as $6 \times BC = 6 \times 780$, by AB giving

$$circumference : diameter > 4680 : 1560 = 3 : 1$$

Clearly the perimeter of the hexagon is less than the circumference of the circle, so the value 3 acts as a lower bound for the value of π.

Archimedes does not show this stage of the analysis in his extant work (I include it here for only for completeness). Instead, his original starting set-up was triangle ACB as in diagram (ii) above. To obtain an approximation to π (one better the $\pi = 3$) he bisects angle BAC and extends a line from A, through the bisected angle, to a point D on the circumference of the circle, as illustrated in diagram (iii) below.



(iii)

Bisecting the angle at A bisects the arc BC at D (This was proved by Euclid in his book III proposition 30, invoking the side-angle-side congruence property of triangles within a circle). The effect of this is to gives us a new polygon with double the number of sides of the hexagon, i.e. a 12-sided regular polygon. What we are looking for is the ratio $DB : AB$. Since DB is one side of a 12-sided polygon we will then want to calculate $12 \times DB : AB$ as a better approximation to π.

Based on diagram (iii) we now have the following geometric analysis: AB is the diameter of a circle of radius 1, and BD is one side of a 12-sided polygon. We want to find the ratio $AD : BD$. To do this we proceed as follows:

$AD : BD$	$::$	$BD : Dd$	(ΔACB is similar to ΔADB, and d is the intersection point of AD and BC)
	$::$	$AC : Cd$	(ΔACD is similar to $\Delta ADBd$)
	$::$	$AB : Bd$	Euclid, Book VI, prop. 3
	$::$	$AB + AC : Bd + Cd$	
	$::$	$AB + AC : BC$	

So we have been able to convert the ratio we are looking for into a ratio involving line magnitudes we already know, namely AB, AC, and BC. We now find the value of the ratio $AB + AC : BC$ using the values shown in diagram (ii). Hence we have

$$AD : BD :: 1560 + 1351 : 780,$$

in other words, $AD : BD :: 2911 : 780$.

Now, although Archimedes used the values for the sides of the triangle shown in diagram (ii) above he knew that the true ratio of AC to BC was less than 1351 to 780. Hence, he used $1351 : 780$ as an upper bound for $\sqrt{3}$, and wrote $AC : BC < 1351 : 780$. Therefore

$$\begin{aligned} AD : DB &< (1560 + 1351) : 780 \\ &= 2911 : 780 \end{aligned}$$

We want $AD : AB$, so we now use Pythagoras to find this ratio:

$$AB^2 : BD^2 \quad < \quad (2911^2 + 780^2) : 780^2$$

$$= \quad 9082321 : 608400$$

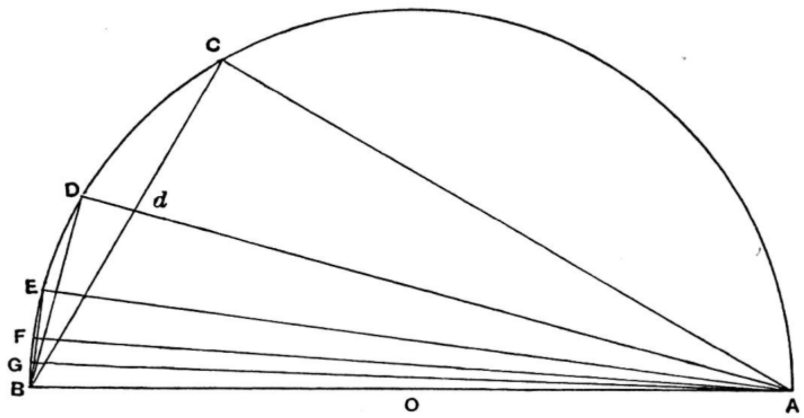
So $AB : BD <$ 3013.688936... : 780. But decimal fractions and irrational numbers did not exist in the days of Archimedes so he "square rooted" the ratio 9082321 : 608400 to the nearest commensurable measure. In this case Archimedes went stated $AB : BD < 3013\frac{1}{4} : 780$ (although it is not known how Archimedes actually simplified his ratios).

Since BD is one side of the 12-sided polygon, a better approximation of the circumference of the circle to its diameter is

$$circumference : diameter > 12 \times 780 : 3013\frac{1}{4}$$

which is equivalent to 3.106280594 (to 9 d.p.) (note that $AB : BD$ represents the ratio *diameter : (part of) circumference*. Since we want the ratio *circumference : diameter* the inequality sign has had to change direction).

Archimedes continued the process of bisecting angle A, constructing lines AE, AF and AG as illustrated below.



He repeated his analysis in order to find the ratios $AE : BE$, $AF : BF$ and $AG : BG$, whereupon he reached a 96-sided polygon, in order to then find yet better approximation to the ratio of the circumference to the diameter (see p12-14, [33], or p199 [36]). The complete list of ratios is shown below.

Polygon	Ratio of line-segments	Value	Bound for π
6-sided	$AC : BC$	1560 : 780	$6 \times BC : AB > 4680 : 1560$
12-sided	$AD : BD$	2911 : 780	$12 \times BD : AB > 9360 : 3013\,\frac{1}{4}$
24-sided	$AE : BE$	$5924\,\frac{3}{4}$: 780	$24 \times BE : AB > 18720 : 5976\,\frac{7}{44}$
48-sided	$AF : BF$	$11900\,\frac{10}{11}$: 780	$48 \times BF : AB > 37440 : 11926\,\frac{51}{99}$
96-sided	$AG : BG$	$23827\,\frac{14}{33}$: 780	$96 \times BG : AB > 74880 : 23840\,\frac{5}{22}$

The final ratio equates to 3.140909654… and acts as Archimedes' lower bound for π (I have used fraction notation to simplify the reading of the ratios). Having obtained a lower bound Archimedes went on to find an upper bound to π using circumscribed polygon. The analysis of this is shown in section 9.4. Ultimately Archimedes obtained the well known bound for π to be

$$3\frac{10}{71} < \pi < 3\frac{1}{7}.$$

In other words, inscribed polygons will never give values greater than $3\frac{1}{7}$, and circumscribed polygons will never give values less than $3\frac{10}{71}$. This is the effort required if you only believe in integers and commensurability.

Now, we know by modern analysis that the sequence

3, 3.1, 3.14, 3.141, 3.1415, 3.14159, 3.141592, 3.1415926, …

is a monotonic increasing sequence which is bounded above. Hence the sequence converges to a number we call π. But Archimedes had no conception of irrational numbers let alone the behaviour of such numbers (via partial sequences, monotonicity, boundedness, etc.). All he had was whole numbers and their ratios. So what a ratio like $74880 : 23840\,{}^{5}/_{22}$ shows is that there is a line segment common to both BG and AB such that we need 74880 of these line segments to fit into BG and $23840\,{}^{5}/_{22}$ to fit into AB. Archimedes (his contemporaries, and future mathematician for centuries to come) could only work with commensurable line segments.

Recall that the ancient Greeks did not accept the idea of performing processes or operation an infinite number of times. Zeno's paradox of Achilles and the tortoise is the classic example of why the infinitely large and the infinitely small were avoided in the mathematics of the ancient

Greeks. Two reasons for ignoring infinity are that Achilles can indeed reach (and overtake) the tortoise, and there can be no such thing as an infinite number of mathematical operations. All answers to problems should be achieved in a finite number of steps.

So, the main take away from the example above is that, given that geometry was the foundation of mathematics, Archimedes was using integers as a convenient way of representing the magnitudes of line segments. He then used integer ratios as a way of comparing such magnitudes. Since the complete form of a number such as $\pi = 3.14159265\ldots$ would require Archimedes to repeat his analysis an infinite number of times, irrational numbers were deemed not to exist. He therefore only dealt with commensurable ratios hence the need to replace AB: $BD <$ 3013.688936… : 780 by AB: $BD <$ 3013¼ : 780. Finally, given that infinity (as in repeating a process an infinite number of times) and the infinitely small (as in the infinitely small line segments which would have resulted from an infinite process) was not considered valid in maths Archimedes applied his analysis only a finite number of times (up to a 96-sided polygon) in order to arrive at a ratio of line segment of finite magnitude.

Furthermore, almost every book or paper which deals with Archimedes' work on π, areas or volumes will tell you that the method of analysis he used is the *method of exhaustion* (a method credited (by Euclid) to Eudoxus). However,

1. the name "method of exhaustion" (credited to Gregory St Vincent, a 16th century mathematician) is highly misleading. Exhaustion, in our case, implies forever bisecting the angle thus forever doubling the sides of a polygon. This leads one to having an infinite number of infinitely short sides in the polygon, ultimately leading to the real value of π.

   But the ancient Greeks never thought in terms of performing a process an infinite number of times. Nor did they consider an infinite number of sides or infinitesimally short sides or infinitesimally small errors. They, simply used the process a finite number of times (as many times as the accuracy they wanted) to find a lower value and upper value to whatever they we seeking;

2. the process used by Archimedes to obtain an approximation to π is better called an *iterative process* because he repeats the same method of analysis a finite number of times only. There is never a hint of Archimedes thinking that this method could be used indefinitely, nor that this method could lead to an exact value for π.

Today we have trigonometry, algebra and an efficient notation so that a more efficient geometric approach to deriving approximate values for π can be presented (see references mentioned earlier). But to the Greeks of the day the method of ratios of line segments was the only way to calculate (approximate) results in a finite number of steps. Although it is easy for us to see such a process as being able to be continued indefinitely, leading us to the idea of limits of the ratio of line segments, the ancient Greeks never conceived of this.

The purpose of the examples above was to show how Archimedes circumvented the issue of infinity (i.e. an infinite number of lines, infinitely short lines, an infinite number of calculations) in his mathematics. It should be noted that situations where such infinities arose was not limited to the "one or two" problems Archimedes (or others) were trying to solve. The problem of infinity is inherent in any geometric problem where curves are involved, one other example of which is where Archimedes calculated the area enclosed by a parabola by using triangles as approximating areas (see section 9.1).

6.8 Why geometry became the foundation of mathematics?

I have mentioned at various points in the previous sections the reason why geometry, over and above numbers and arithmetic, was seen as the foundation of mathematics. It is worth recapping on this. The question is, If arithmetic was the foundation of mathematics and life for the Pythagoreans, and if there was no crisis of faith in integer-mathematics (even after the discovery of magnitudes which cannot be expressed as integer ratios), why was there a near complete shift to geometry? There are four possible perspectives on this.

- <u>Perspective 1: The exactness of geometry; the non-exactness of numbers</u>
  The example of the diagonal of a unit square highlights the issue of mathematical exactness (as opposed to approximation). From the time of Euclid onwards the ancients knew that geometry was the paradigm of rigour and exactness. The incommensurable number $\sqrt{2}$ could be constructed exactly as the diagonal line of a unit square, because every aspect of the construction of a square by straight edge and compass was an exact procedure. Starting with any line of a given length another line of equal length could be constructed perpendicular to it. The construction of a square could then be completed, and the diagonal could be constructed. Every line constructed was exact. Even if there were errors between the length of line of the square these would be attributed to human error not to flaws in the instruments used or the procedure for construction. Thus the diagonal of a unit square was finite and exact.

This is not the case with the numerical representation of the diagonal. Assign a number to the diagonal would require identifying the ratio of length of the diagonal to the length of the side of the square. In simple terms they would have tried ratios such as $14 : 10, 141 : 100, 1414 : 1000$, etc, but they would never have reached a final ratio representing the number of unit line segments which could go into the diagonal a finite number of times. This meant that no number could represent the diagonal exactly.

From this perspective it is easy to see why mathematicians from the ancient times up to the end of the 18th century considered geometry as the foundation of mathematics particularly because, by the 18th century, fantastical or impossible results were being obtained from the use of infinite series (such as $1 - 1 + 1 - 1 + 1 - 1 + \cdots$ being shown to be equal to either 0 or 1 or even ½). With such lack of certainty in numerical results it is no wonder people of the day wanted to geometrically construct every result obtained by algebraic analysis, as a way of confirming such algebraic results.

The attempt to find a ratio of line segments in order to show that the diagonal of a unit square was commensurable to the sides of the square can be interpreted as a first attempt at the construction of an irrational number. Such a construction was geometric: continued subdivision of a reference line (or, seen from another perspective, the choosing of ever smaller unit lines as the reference line) to see how many of these could be made to fit into the diagonal. We would have to wait until the mid to end of the 19th century to have a process of construction of the irrationals which was purely numerical (as well as being based on set theory, where the two most well-known ones are Dedekind cuts and equivalence classes of Cauchy sequences).

- Perspective 2: Numbers as not adequate enough to represent geometry

  Recall the Greeks attitude towards numbers, namely that these were only integers. They had the concept and practice of ratios, i.e. the comparison of two whole numbers. For example, the 3-4-5 triangle, the 5-12-13 triangle etc., are triangles whose sides form integer ratios, and therefore the hypotenuse is commensurable with any of the other two sides.

  But they had no conception of rational fractions (namely as being a comparison of parts to wholes), even less a conception of fractions as non-repeating infinite decimals (i.e. irrationals). So, it must have been initially confusing for ancient Greek mathematicians to meet the situation where one cannot form an integer ratio between the diagonal of a unit square and its sides. … Since geometric figures represent exactness in mathematics, at least in the ideal world of Plato geometric forms, number associated with such figures must also represent exactness. This is exemplified by the whole number, such as 3-4-5 for the lines of a right-angled triangle. Exactness is wholeness. But when they came up against the diagonal of a unit square (an exact geometric figure) there was no equivalent exact ratio of numbers to represent this line. Imagine trying to define an infinite decimal as an exact number when ones only conception of what constitutes a number is that it stops. And 1.4142135… does not stop in the number of whole numbers which form its decimal construction, hence $\sqrt{2}$ is not exact and cannot therefore be a number. Because of this numbers were seen as not being sufficient or adequate to describe geometry. Hence geometry was fundamental to maths in a way numbers and arithmetic could not be.

It must also have been confusing that the method of exhaustion sometimes gave exact results (as for the area under a parabola) and sometimes only allowed for approximate results (as for the calculation of π). For example, in the case of the parabola, it was possible to assign an integer ratio which compared the size of the parabolic area to that of a triangular area (i.e. $P : T_1 :: 4 : 3$). But in the case of calculating a value for π it was not possible to assign an integer ratio to the ratio of the "circumference" of the polygon and the diameter of the circle. This is despite the fact that triangles were used as the principle object in both the calculation of the parabolic area and the calculation of π.

Nowadays we have arithmetic analysis (i.e. real analysis) to better understand the distinction between these two situations. Specifically, it requires a distinction between a series which converges and a series which does not. For example, we now know how to justify (by the use of various tests) the convergence of the series

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots+\frac{1}{n^2}+\cdots$$

Here each term is multiplied by a number whose magnitude is less than 1 (i.e. ½). On the other hand, those same tests allow us to show that

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots+\frac{1}{n}+\cdots$$

diverges even though both terms $1/n^2$ and $1/n$ approach zero as n approaches infinity. We also have mathematical analysis which allows us to show that

3.1415926535 8979323846 2643383279 5028841971 6939937510 5820974944 …

is irrational, hence the reason why no rational fraction (or integer ratio) can ever be found for π.

But such an arithmetic conception was not available in ancient times hence the reason why numbers (i.e. integers) were not adequate enough to represent certain geometric objects (incommensurable lines, areas or volumes).

- <u>Perspective 3: The logical deductiveness of geometry in Euclid's Elements</u>
  Logical deductiveness was highly prized in the mathematics of the ancient Greeks, and this was more clearly demonstrable through geometry. After all, there were definitions and axioms of geometry such as

  - a point is that which has no parts;
  - any straight line may be drawn between two points;
  - any circle may be constructed with a given centre and radius.

  These acted as a frame of reference from which more complex geometrical shapes could be constructed and more complex statements about geometry could be proved (see section 3).

  Furthermore, Eudoxus (395-390BC – 342-337BC) had managed to extend the then known theory of commensurable magnitudes to include incommensurable magnitudes (as presented by Euclid in book X of the *Elements*). This extension allowed for a complete and consistent theory of magnitudes from a geometric perspective. So there was no longer a need to assign numerical values to incommensurable magnitudes.

- <u>Perspective 4: The lack of logical deductiveness of arithmetic in Euclid's elements</u>
  There was no equivalent logical deductiveness in the arithmetic of Euclid's books VII, VIII, and IX to "balance" the rigour he had developed for geometry. No definitions for pure numbers, nor axions of arithmetic, were stated. Hence there was no consistent foundation from which to construct more complicated numbers (such as fractions, negative numbers and roots) and more complicated arithmetic (such associativity, commutativity and other rules which form part of the generalised arithmetic we call algebra).

So, it is all very well saying 1 + 2 = 3 but which number(s) do we take as already given (i.e. as an axiom)? Can we assume the number "2" as already existing and distinct from "1", or is "2" constructed from "1"? If the latter, then we can ask, Is "1" the number we assume as already existing? Nor have we defined how to perform arithmetic. In that case how do we justify that 1 + 2 = 2 + 1? Or that 3×(1 + 2) = 3×1 + 3×2? After all, line segments, lines, circular arcs had been defined, along with conic sections of the circle, ellipse, parabola and hyperbola. Also defined was the process by which these objects could be constructed, namely by straight edge and compass. Beyond this there had been defined the process of geometric arithmetic (see sec 7) by which these various geometric objects could be added, subtracted, multiplied, divided, and even square rooted. But without axioms and definitions for numbers and arithmetic it becomes impossible to explain what $\sqrt{2}$ was or how to add an integer to $\sqrt{2}$ let alone how to show that $\sqrt{2} \times \sqrt{3} = \sqrt{6}$.

Therefore, we arrive at the following conclusion: since the only two areas of mathematical investigation known at the time of the ancients were geometry and arithmetic it was seen that numbers (as integers) were not sufficient to describe mathematics as a whole. Only geometry could do this, hence mathematics was therefore fundamentally geometric.

And with Greek mathematics having entered Europe at about the time of the 12th century (via the Arabian centres of mathematics), and becoming dominant in the 15th century, geometry came to dominate the approach, thinking and methodology of European mathematicians from the 15th century onwards. So deeply ingrained was the geometric mindset of the early modern mathematicians that, despite the emergence and centrality of coordinate geometry, algebra, and calculus in subsequent centuries, it would take until the mid to end 1800s for all geometric consideration to be removed from mathematics. Mathematics would then be cast fundamentally in terms of arithmetic. Today we call this real analysis.

7 On the geometric constructibility: The case of Descartes

7.1 Introduction

It stands to reason that if geometry is considered to be the foundation of mathematics then numbers and arithmetic must be constructed geometrically. Numbers such as natural numbers, ratios, and roots, as well as the arithmetic of addition, subtraction, multiplication, division, and square rooting need to be constructed using only a straight edge and compass. So it was that in the 17th century people like Descartes saw the construction of number and arithmetic through geometry. And, indeed, he showed that the arithmetic operations could be performed geometrically.

> "Just as arithmetic consists of only four or five operations, namely, addition, subtraction, multiplication, division and the extraction of roots, which may be considered a kind of division, so in geometry to find required lines it is merely necessary to add or subtract other lines; or else, taking one line which I shall call unity in order to relate it as closely as possible to numbers, and which can in general be chosen arbitrarily, and having given two other lines, to find a fourth line which shall be to one of the given lines as the other is to unity (which is the same as multiplication); or, again, to find a fourth line which is to one of the given lines as unity is to the other (which is equivalent to division); or, finally, to find one, two, or several mean proportionals between unity and some other line (which is the same as extracting the square root, cube root, and so on, of the given line *(sic)*. And I shall not hesitate to introduce these arithmetical terms into geometry, for the sake of greater clearness." (p237-238, [74])

We are still carrying over the effect of the Pythagoreans discovering that $\sqrt{2}$ could not be expressed as a ratio of two integers, but could be constructed exactly as the diagonal of a unit square. Hence, the idea that all numbers had to be constructed geometrically. Such constructions would be proof that any solution to algebraic equations were valid, because such solutions could be represented as line segments.

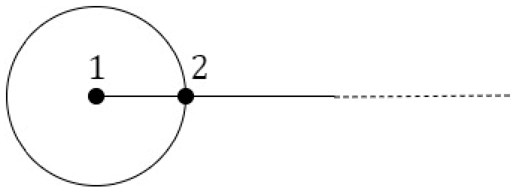
7.2 Constructing natural numbers geometrically

If geometry is the foundation of all mathematics then all numbers have to be constructed geometrically. How do we construct numbers geometrically? Via points. In fact, a number is said to exist because we have constructed a point. So how do we construct points? A view held by Aristotle(384BC – 322BC), and one which was carried forward for about 2000 years, was that points were either created by the intersection of two lines or were the extremities of a finite line. But in order to be able to construct our very first finite line or curve we need two points. Since we have to start somewhere we accept, apriori, the existence of one point if not two points were needed.

Recall that the ancients did not consider "1" to be a number. The "1" was considered as the source of all numbers but not a number itself. But by Descartes' time people had disposed of the idea of "1" not being a number. In that case "1" would need to be constructed like all other numbers. However, for illustrative purpose we will start as if the first point, already given to us, represents the number "1" (this matches the modern conception of Peano's axioms whereby "1" is a number already given to us). This point/number can therefore be said to be an axiom of geometry from which the remaining natural numbers can be constructed.

Then, we use our straight edge and compass to construct a circle whose center is the given point. From the center we can also construct a line. This line will intersect the circle, and it is this intersection which constructs our second point, and hence the number "2", as illustrated below.

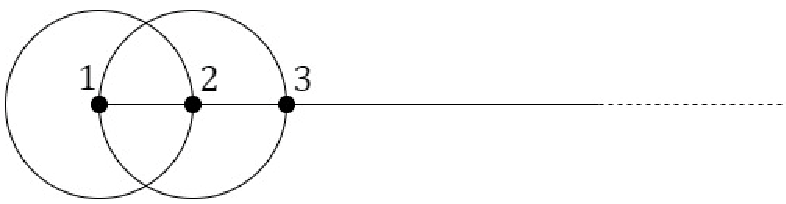
A point already given to us apriori.
This is the number 1.

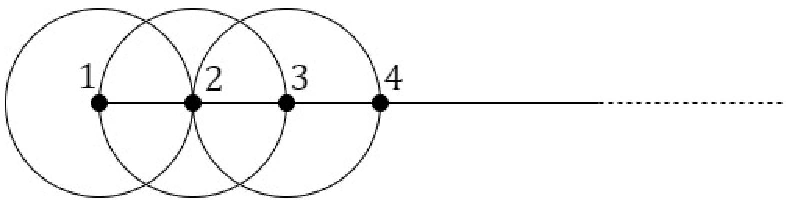


Constructing the number 2
by the intersection of the line and the circle.

Here the dashed part of the line represents the case where the line continues indefinitely and therefore has no end point.

We now use our straight edge and compass to construct more lines and curves as a result of which we construct the remaining positive integers, as illustrated below. To construct the number "3" we place the point of the compass on "2" with radius being the length from "1" to "2". This will intersect the line again, creating a third point, hence the number "3", etc.





What about constructing negative numbers? Well, this was not done. Firstly, negative numbers were not accepted during the renaissance period, and secondly Descartes himself did not believe in negative numbers. However, from a modern perspective it is very easy to construct negative numbers. We simply construct point to the left-hand side of 1. This obviously means that we need to construct 0.

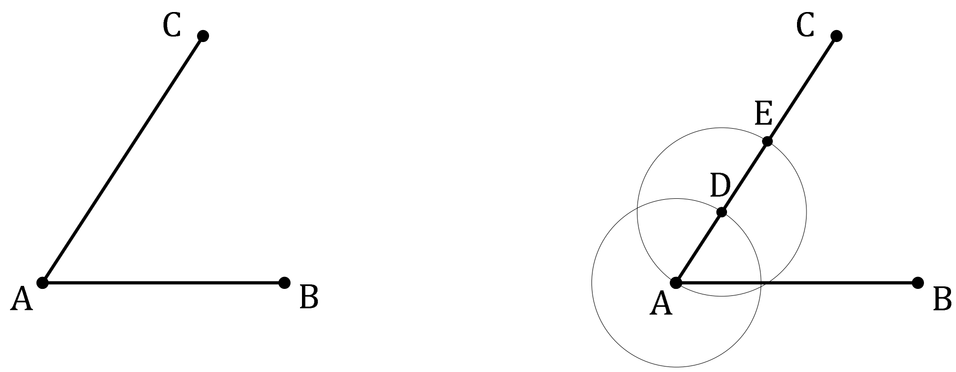
But then if points represent numbers one would need to represent 0 as the absence of a point. One cannot via the intersection of lines and circles since such intersections actually construct points. This is all moot since negative numbers were not constructed geometrically or accepted arithmetically at this stage.

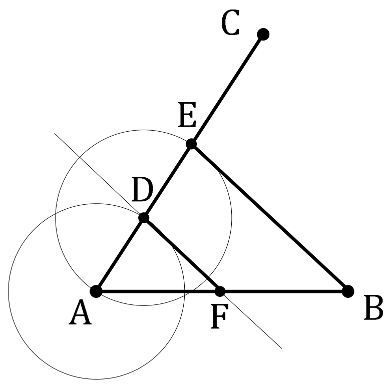
7.3 Constructing fractions geometrically

The ancient Greeks used ratios rather than fractions (what we now call rational fractions have been around since the time of the Babylonians and Egyptians around 4000 years ago but the Greeks used ratios instead). Effectively their construction of ratios was as the comparing of two line segments, i.e. the comparison of two integers.

To construct a rational fraction we do as follows:

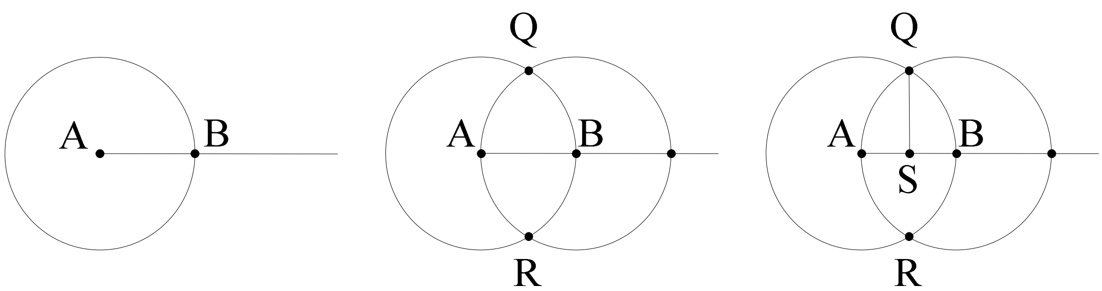
- define two points A and B then construct a line between them. Designate this line segment to be 1 unit long.
- draw a third point C not on AB, and construct a line AC;
- construct a circle, center A, of any given radius less than AC; this circle intersects AC at a given place, thus constructing a new point D;
- construct a circle, center D, of radius AD; this circle intersects AC at a given place, thus constructing a new point E;
- Construct a line from B to E;
- Now construct a line from D which is parallel to BE and intersects AB. Now call the newly constructed point F.





We can then set up the following ratio: $AF : AB :: AD : AE$. We know $AB = 1$ (by definition) and $AE = 2AD$. Hence, in modern notation, we have $AF/1 = AD/2AD$, implying AF = ½, i.e. line segment AF is half the length of line segment AB. The process above can be repeated as often as we want to construct the numbers 1/3, 2/3, 1/4, 3/4 etc.

Another way of geometrically constructing the number ½ is as follow: construct a line AB between two constructed points A and B. This is our unit line segment. Now construct a circle center A, radius AB. Extend line AB by the same amount as radius AB and construct another circle center B.



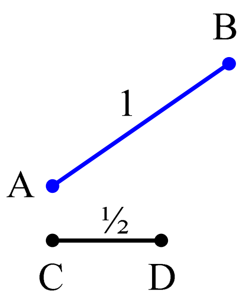
The two circles intersect at Q and R. By drawing a perpendicular between Q and R we cut the radius AB in half at S. The intersection of QS with AB constructs a point, and thus constructs the number ½. The fact that QS cuts AB in half follows from Euclid I.10 and the fact that triangle AQB is equilateral (see p32 onwards for constructing equilateral triangles and perpendicular lines).

Note that there seems to be a contradiction as to what the number Q should be. If numbers are equivalent to points constructed by the intersection of curves then since length AQ = 1 (AQ = AB), point Q is the number 1 as measured from A. On the other hand we have by Pythagoras QS = sqrt(1 – ½) = sqrt(3)/2 long (obtained using AQ=1), implying point Q = sqrt(3)/2. So, which value do we take as the number for Q? Well it depends on the reference frame we use. If we measure everything from line AQ then the number associated with Q is indeed 1. If we measure everything with respect to the line AB then the number associated with Q is sqrt(3)/2. On the other hand, it would indeed be possible to have Q be 1 but only under what is called, in modern maths, a transformation of coordinates.

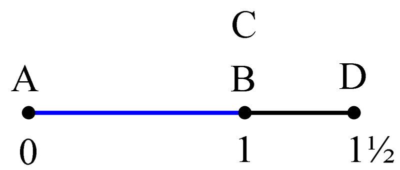
7.4 Constructing arithmetic geometrically – Addition and subtraction

The case of geometric addition and subtraction is straightforward. For example, given two lines AB and CD (diagram (a)) we use a straight-edge and compass construction to append CD onto AB at point B (diagram (b)).

So given Length AB = 1 and length CD = ½ we transfer (by straight edge and compass) line AB to the beginning (or end) of line CD.



(a)

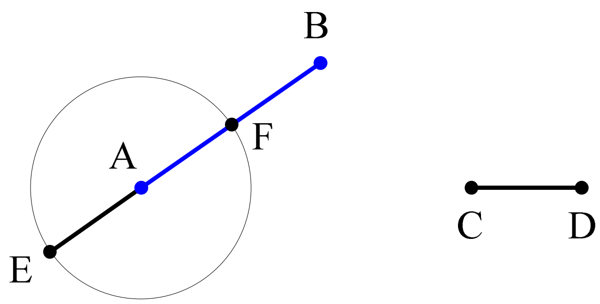


(b)

Hence we can construct the number 1½ from the addition of lines. If CD was constructed to be a square root, say $\sqrt{2}$, then then number constructed to be AD is $1 + \sqrt{2}$, such as when adding AB to the diagonal of a unit square. Subtraction of lines is similarly conceived with the requirement that one subtracts the line of shorter length from the line of greater length. In this case we can transfer CD to lie on top of AB with point D joined to point B:

*(*diag here*)*

Another way of doing arithmetic which allows us to add and subtract using one diagram is to open our compass to length CD and then draw a circle about point A:



Then BE = AB + CD and BF = AB – CD.

Since addition and subtraction have been shown to be geometrically possible one can now perform addition and subtraction arithmetically as a short cut to geometric arithmetic. For example, given that 1 and $\sqrt{3}/2$ are constructable numbers (we will get to how to construct square roots in section 7.7) we can perform $1 + \sqrt{3}/2$ arithmetically as a short cut to performing the addition geometrically.

How far can we go with this? Quite far. For example, we can write

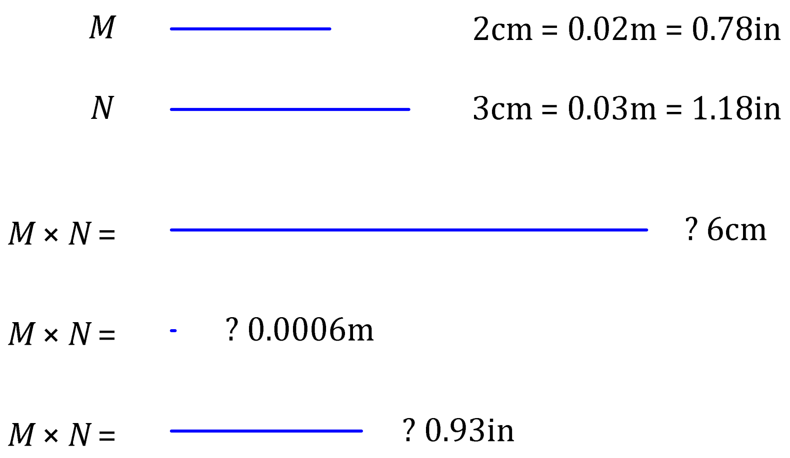
$$4 - \sqrt{\frac{8}{3}} + \frac{15}{34}\sqrt{2 + \frac{\sqrt{3}}{2}} + \frac{3 - \sqrt{5}}{\sqrt{3\sqrt{2}}}.$$

Since each individual number is constructable, and since addition and subtraction can be performed geometrically, the sum above is constructable. Then, however complicated it might be to do so, it is possible to construct lines and circles which intersect in such a way as to ultimately construct a point representing the number above. So we might as well perform the additions and subtractions numerically instead.

7.5 On the unit line segment of arbitrary length and the principle of homogeneity

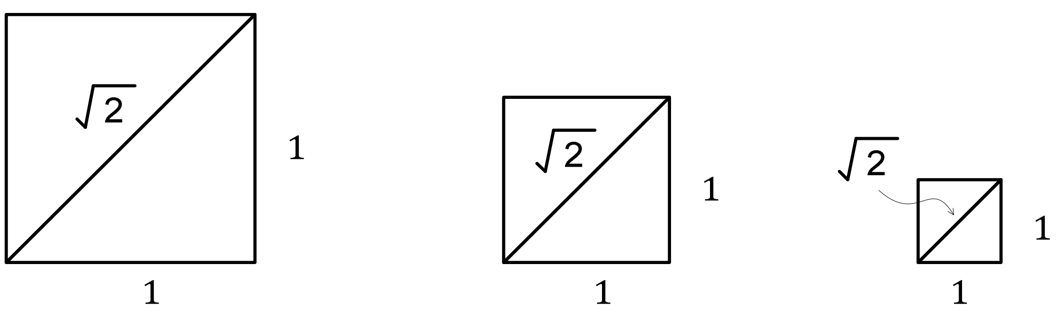
When speaking of magnitudes everyone up until the time of Descartes referred to the size of things, for example “something was 6cm long”. So if a line was 6cm long this was (and still is) in respect of a reference unit we know as 1cm. But suppose 6cm is the result of 2cm×3 cm. What happens if we change to metres or inches?

Then we would have 0.03m×0.02m = 0.0006m = 0.06cm so we obtain a completely different number when working in metres compared to when working in cm. This is the case whatever physical units we use. This highlights a problem when thinking about numbers as magnitudes related to the real world, and not as pure numbers.



On the other hand simply saying that a line was of length 12 made no sense, because the question would then be "12 what?" Without units what does 12 mean? What Descartes did was to replace units related to physical measurements and simply have an *arbitrary unit length* (unrelated to any real world units) as his reference. He therefore moved away from the idea of magnitudes (things measured according to specific units) to the idea of using numbers as pure numbers. As a result, Descartes was then able to say that a line is "12" long with respect to the unit line he had chosen. In other words, the line, which is the product of two other lines, is 12 times longer than the arbitrary unit line.

The same idea applies when trying to justify $\sqrt{2}$ as the diagonal of a unit square. To see this consider the three squares below, each one of which is defined as a unit square.

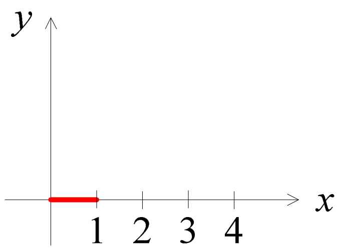


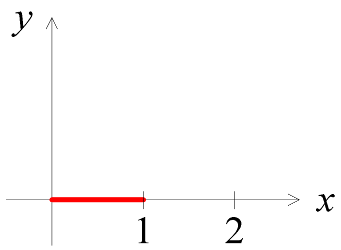
Clearly the sides of each square is of different length compared to the other two squares. But here we are no longer looking at the lines as magnitudes. I.e. we are no longer looking at the actual/real length of the line. Instead we are performing the abstraction of defining the sides of each square to be "1" unit long. It is then easier to see the diagonal of each square as representing the same $\sqrt{2}$ even though the line which represents each diagonal is of a different length. This removes number from geometry and real physical measurement, and allows us to interpret $\sqrt{2}$ purely as number and not as magnitude.

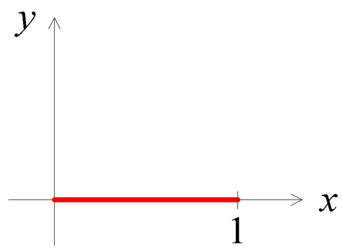
I might say that Descartes idea of using an arbitrary unit was the start of pure mathematics, even though this arbitrary unit was represented by a line. To paraphrase Alfred North Whitehead I might say that the greatest advance in pure mathematics was when 2 was separated from 2 apples (a paraphrase of a general comment made in chap 2 of "Science and the modern world"). This is an example of defining a number as a pure, abstract, entity rather than being inherent part of the physical world. Descartes' conception of an *arbitrary unit line* can be seen as a great generalisation which would help in solving algebraic problems.

What Descartes did was to effect a radical shift in the conception of how numbers were used to represent line segments. In his work on solving geometric problems he conceived lines as being "measured" without any units. In other words, instead of measuring the actual length of a line to be "so many cm or m", something which was specific to a given problem, he simply gave an arbitrary line a length of 1 unit. The length of any other line would then be measured relative to this datum or reference length. In doing this he was effectively normalising a given line, and all other lines would then be expressed as multiples of this reference line. The usefulness of this will be seen in sections 7.3 and 7.6.

The idea of a unit line is seen in the standard Cartesian coordinate system (see diagrams below) where the unit distance from the origin can be of any length we choose. In fact, its length is immaterial. The distance from 0 to 1 is still a unit length.





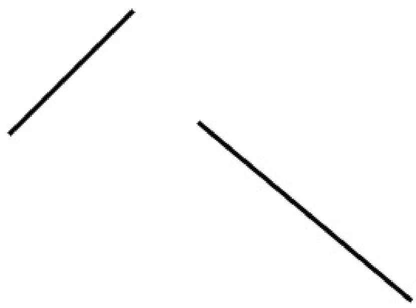


Today we just deal with numbers as pure numbers. And in coordinate geometry we deal with the distance from 0 to 1 as being whatever the distance it is. We are not interested in the actual length of the interval [0, 1], so that however we scale our *x*-axis (as shown in the three diagrams above) we do so without this being an issue.

Returning to the idea of multiplying two line segments, the ancient and early modern mathematicians (up to the 19th century) always saw something like 2cm by 3cm as giving an area, i.e. 6cm^2. This answer does not represent a line segment. So how do we go about performing multiplication of two line segments in order to produce a line segments as the answer? By using the theory of similar triangles. Analysing a geometric configuration using the properties of similar triangles allows us to state that triangles are similar on the basis of comparing line segments and angles (one famous comparison being Side-Angle-Side). In our case we will be interested in comparing sides of similar triangles as a way of setting up the necessary ratios/fraction which will gives us the answer to our multiplication.

7.6 Constructing arithmetic geometrically – Multiplication and division

Taking on board the idea of the arbitrary unit line we are now in a position to multiply two line segments. So, suppose we have two lines, as shown below, which we wish to multiply.



Normally, when given two lines connected at one point, their multiplication is simply done as base × width or base × height. Such an arithmetic was always interpreted as giving an area. But Descartes was able to reconceptualise the multiplication of line segments to produce a line segment as a product. He did this by setting up ratios of line segments based on similar triangles.

So, using straight edge and compass we can transfer these two lines so that they appear as in diagram i) below. What we now want is to find BD × BC.

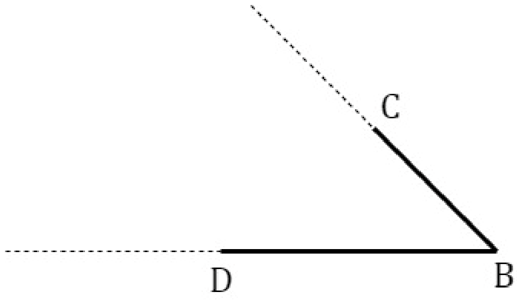


Diagram (i)

Diagram (ii)

From this configuration Descartes sets up two similar triangles with the aid of an arbitrary reference line (diagram (ii)) which he ends up using as a unit line. One can either construct it separately, as in diagram (ii) above, and then transfer it onto the configuration of diagram (i), or one can simply mark of a unit line segment along BD or BC. Either way, we have a unit line segment labelled as AB in the diagram below. Points A and C are then joined together. Finally, a line is drawn from D parallel to AC. This meets the line through BC at E.

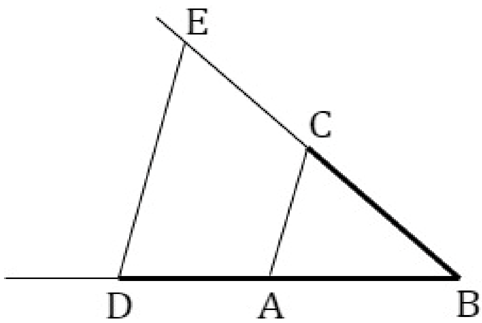


Diagram (iii)

This is Descartes original configuration (reproduced from p1249 [21]). By drawing DE parallel to AC we have triangles BAC and BDE to be similar Descartes is now able to set up the ratio

$$BD : BA = BE : BC.$$

What this says is that length of BD compared to the length of BA is simply going to be considered as the length BD (since BA is a unit/reference line). We can also compare length BE to the length BC. Because we have set up the geometry to contain similar triangles we can then compare these two comparisons, where the comparisons are all just that of line lengths.

We can write this in modern terms as

$$\frac{\text{BD}}{\text{BA}} = \frac{\text{BE}}{\text{BC}},$$

from which $\text{BE} = (\text{BC})(\text{BD})/\text{BA}$. In order to simplify matters Descartes lets $\text{AB} = 1$, from which he obtains $\text{BE} = \text{BC} \times \text{BD}$ (this might then be called the normalised product of the lines BC and BD). Since BE is a line segment this result has to be interpreted as a line segment, not an area. The reason Descartes can do this is because he is using the idea of proportionality of line segments (via the use of similar triangles). The reference line AB acts as a basis for setting up a framework of proportionality via similar triangles. Without AB we could not effect proportional comparison and reconceive the multiplication of two line segments as being a line segment. In this way Descartes has effectively re-defined geometric multiplication.

Another argument which explains why BE can be interpreted as a line segment is based on the idea of dimensionality. Dimensionally speaking, BC×BC is 2-dimensional (an area) and AB is 1-dimensional (a line), so that by dividing the area BC×BD by AB the former has been reduced to a line, and BE can now be interpreted as a line rather than an area. It is by the use of similar triangles and ratios that we can do this.

Note that the configuration of the arbitrary unit lines and similar triangles above is not the only one we could set up. Two other are shown below.

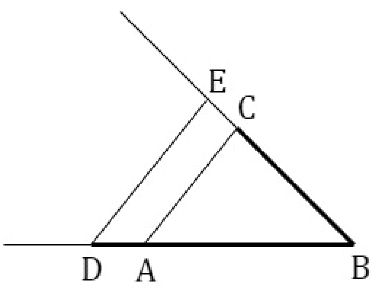


diagram iv)

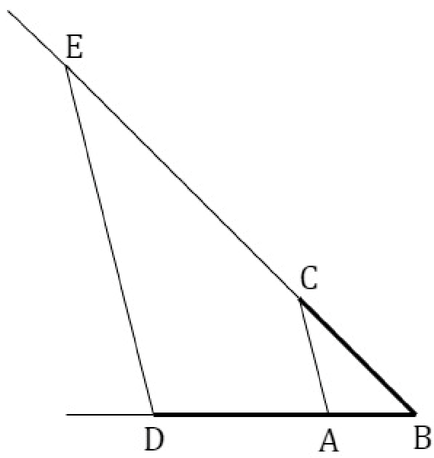


diagram v)

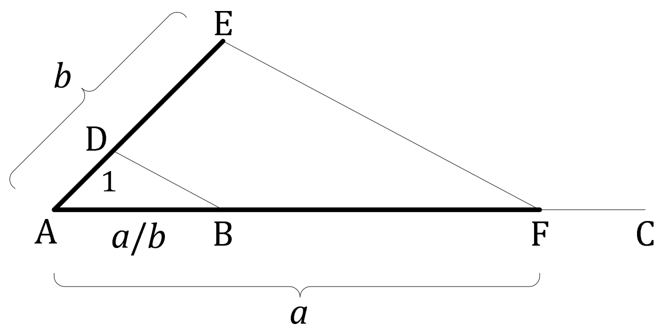
Since the lines BC and BD are of fixed length it looks like different choices of unit line AB produce different lengths for the product BE, and visually speaking this is indeed the case. But surely multiplying two lines should produce the same result as is their product. However, this thinking betrays our modern thinking about arithmetic. In the case of geometric multiplication we are not thinking numerically (such as when 2×3=6 where “2”, “3” and “6” are simply pure numbers without any reference to geometry) but geometrically where ratios of line segments are compared.

At this stage in the history of maths we haven’t yet defined numbers as pure numbers. At this stage, numbers, and algebra, are merely representations of geometry. So geometrically, Descartes’ process of multiplication is relative to AB. This means that, visually speaking, his process is relative to how long AB looks. In diagram v) AB is short (equivalent to being measured, say, in cm) compared to BD, so that BE looks quite long (say, the number 1200cm). But in diagram iv) AB is long (equivalent to being measured, say, in m) compared to BD, so that BE looks shorter (say, 12m). To us, 1200cm = 12m because we have converted between the different units of length. Similarly, BE in diagram iv) equals BE in diagram v) because we have used different units, specifically different arbitrary unit lines for AB.

Such a geometric multiplication can also be done when the line segments are incommensurable, so it is possible to not only associate $\sqrt{2}$ with a line segment (the diagonal of a unit square) but also to perform geometric arithmetic using $\sqrt{2}$.

The ratio concept of multiplication may therefore be seen as a 16th century attempt at defining a new way of performing multiplication, an attempt founded on geometry rather than pure numbers (as in modern day analysis).

In terms of performing division we may use the same general geometric set up with the following changes: let AF = a and AE $= b$. We now want to divide a by b. To do this construct a line from E to F. Again we mark of a line segment, say at D where we now call AD our unit line segment. Then we now construct a line segment from D to AF which is parallel to EF. Again by similar triangles we have AE $:$ AD $=$ AF $:$ AB so that $b : 1 = a :$ AB implying that AB $= a : b$. So the geometric division of line segment a by line segment b is (in modern notation) AB $= a/b$.

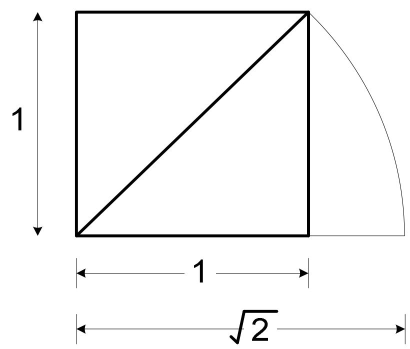


Again, AB equal a/b times the unit line segment, and again, such a result is achieved by setting up similar triangles and comparing line segments, with the aid of an arbitrary unit line as a reference measure.

7.7 Constructing square roots geometrically

In one sense it is actually quite easy to construct square roots. We can, in fact, construct a sequence of square roots as follows:

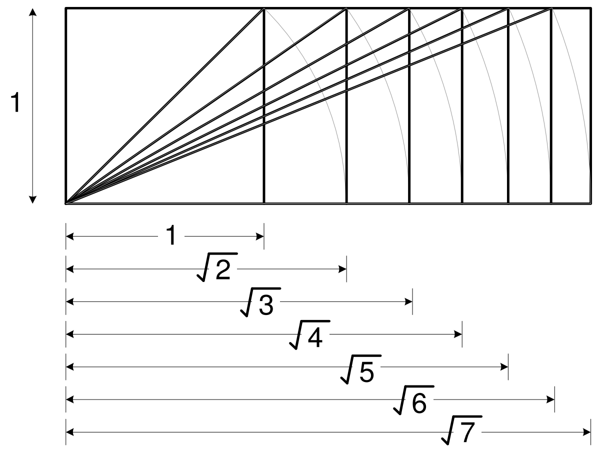
- firstly, construct a unit square (involving the use of a straight edge and compass to construct two line segments which are parallel to each other and perpendicular to two other parallel line segments), then we use a straight edge to draw the diagonal. This gives us $\sqrt{2}$;
- now use a compass to rotate the line $\sqrt{2}$ onto the number line. Since the number line is a geometric representation of numbers, and since we have been able to transfer the diagonal line onto the number line then this diagonal must equate to a number which we call $\sqrt{2}$:



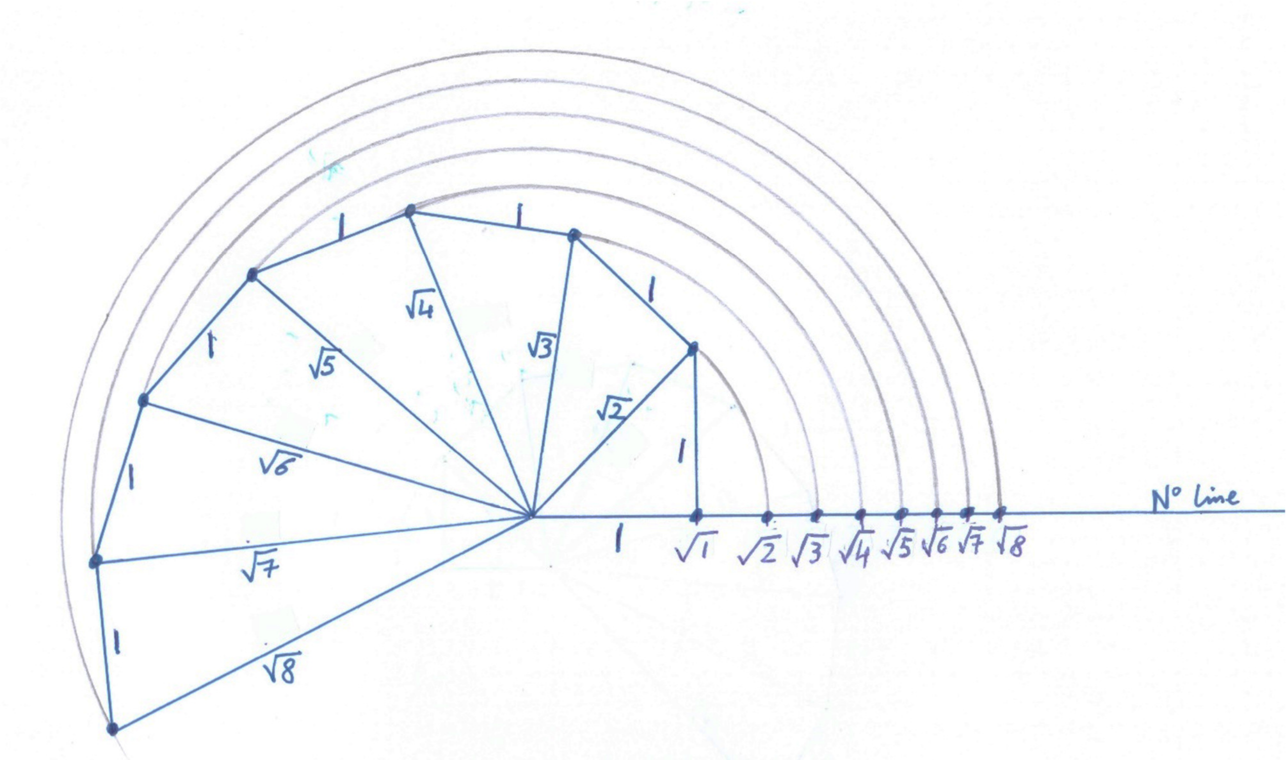
- we now form a rectangle of height 1 and width $\sqrt{2}$ by constructing a line of length 1 perpendicular to the point $\sqrt{2}$, and then joining the top of this line with the top of the previously constructed square;
- with this new rectangle we draw a diagonal and rotate it onto the number line. We have

therefore been able to transfer this diagonal onto the number line, hence $\sqrt{3}$ is a number;

- we can continue the process above indefinitely. This results in the diagram illustrated below.



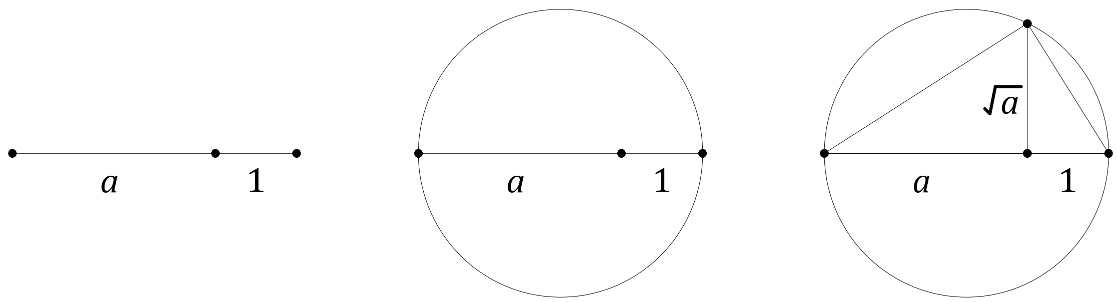
Another simple way of illustrating the process of constructing roots of (integer) numbers is via the root spiral as shown below. Here we start with a triangle of unit base and height. By Pythagoras' theorem the diagonal is $\sqrt{2}$. We then construct the next right-angled triangle by using the hypotenuse of this first triangle as a base. By Pythagoras' theorem this second triangle has hypotenuse $\sqrt{3}$. We can then continue this construction indefinitely.



Recall that numbers are represented by points, and points are constructed by intersecting lines. Each constructed hypotheneuse intersects each unit line, thus constructing the relevant point, and therefore number. Hence all the square roots in the diagram above exist as numbers. Furthermore, the use of a compass allows us to map each point to the number line by the intesection of the circular arc with the number line. New points are therefore constructed which match the ones relating to the hypotheneuses. The number line is now populated by square root numbers as well as whole numbers.

The problem with the diagrams above is that they are not general. We cannot construct any square root we want without firstly constructing all the roots previous to it. This means that we can't construct $\sqrt{n}$ without first constructing $\sqrt{n-1}$. We therefore need a more general way of geometrically constructing square roots. The procedure is as follows:

- we start with two points and then we construct a line segment a of any length joining these two points. We then add onto the end of a a unit line, as shown in diagram (i) below;
- using $a+1$ as diameter we construct a circle of radius $(a+1)/2$ as shown in diagram (ii) below;
- then we construct a perpendicular from the end of line a (i.e. the start of the unit line) to the circumference of the circle. This creates a point on the circumference;
- we then join one end of the diameter to the point on the circle and the other end of the diameter to that same point on the circle, as shown in diagram (iii);



(i) (ii) (iii)

We now have three triangles (the large one and the two internal ones). By the side-angle-side and angle-side-angle aspect of geometry they are all similar to each other. Letting the vertical perpendicular line in diagram (iii) be x we have $a : x :: x : 1$. In modern notation this is

$$\frac{a}{x} = \frac{x}{1},$$

implying $x = \sqrt{a}$. As we did for multiplication and division, so here we have made use of ratios of line segments where these ratios were taken in such a way as to give roots.

The form of the ratio $a : x :: x : 1$, i.e. the order in which the line segments are compared, was known in Descartes' time as a *mean proportional*. In modern terms mean proportionals are geometric means. This can be seen when we replace the unit line segment of diagrams (ii)/(iii) above with the more general line segment b. Then the vertical perpendicular line segment, which we will call x, is said to be the mean proportional of a and b if

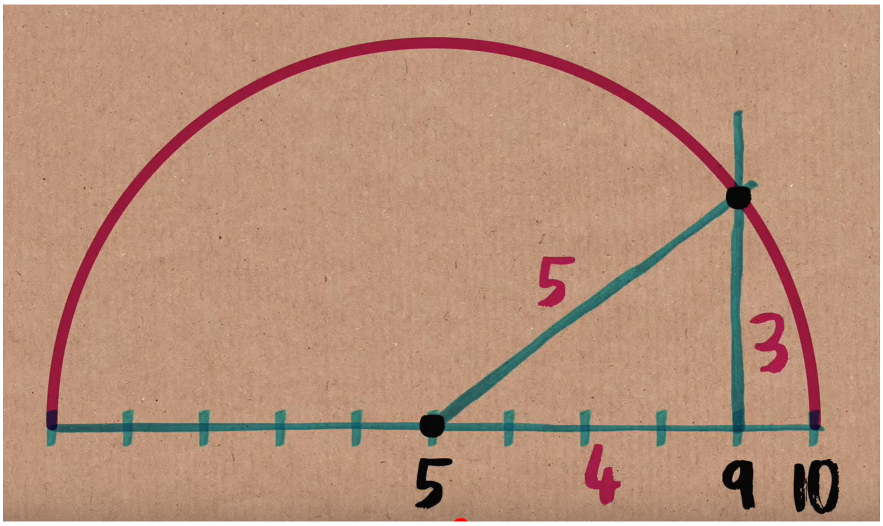
$$a : x :: x : b.$$

In modern terms this is equal to

$$\frac{a}{x} = \frac{x}{b}$$

or $x^2 = ab$. This implies $x = \sqrt{ab}$ which is the geometric mean of a and b. Such a set-up of ratios, i.e. such an order to comparing line segments, is what allows us to perform rooting (with respect to an arbitrarily chosen unit line b).

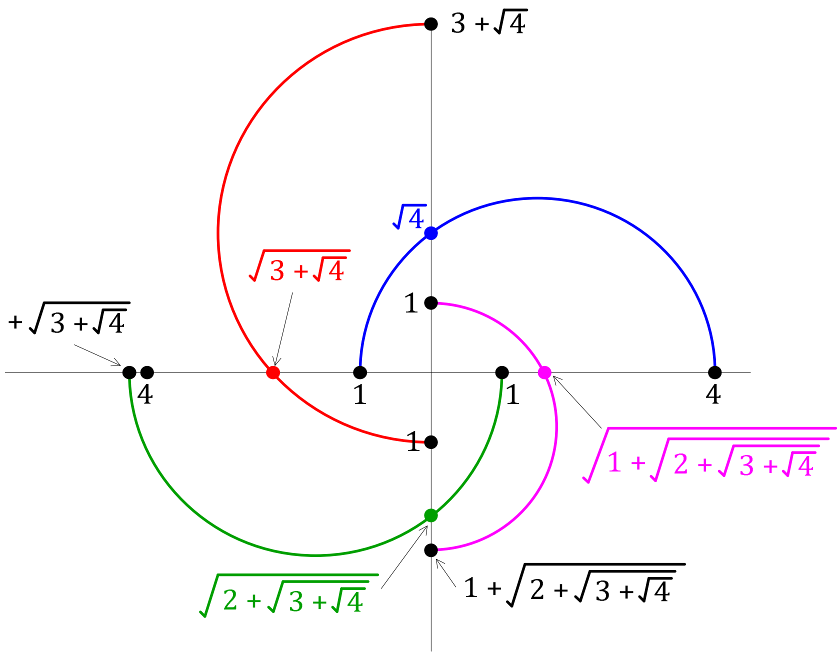
As a concrete example of finding roots geometrically consider constructing the square root of 9. With reference to the diagram below we draw a line of length 9 and then append a unit line segment to obtain a line of length 10. We then draw a semicircle of radius 5, centred at 5, and then draw a perpendicular from 9 to the semicircle. This perpendicular has length 3, the square root of 9, which can be shown via Pythagoras' theorem.



This form of construction allows us to construct more complicated square roots such as nested square roots. For example, we can geometrically construct

$$\sqrt{1+\sqrt{2+\sqrt{3+\sqrt{4}}}}$$

as illustrated in the diagram below. Note here that all values are written in absolute terms, partly because we are only interested in magnitudes (i.e. lengths/distances) and partly because negative numbers were not accepted in Descartes' time.



So, consider drawing a semi-circular arc with diameter from 1 to 4, as shown in blue. By the above analysis the intersection of the vertical with the arc is $\sqrt{4}$ shown as the blue dot. The second root we find is $\sqrt{3+\sqrt{4}}$. We do this by adding a length of 3 to $\sqrt{4}$ and drawing a relevant semi-circular arc of diameter $4+\sqrt{4}$ shown as the red semi-circular arc. This then leads to the number $\sqrt{3+\sqrt{4}}$ shown as the red dot, etc.

Note that the blue and red semicircles can constructed in a straightforward manner since the radius in each case is an integer. For the green semi-circle things are slightly more involved. Here the diameter is $3+\sqrt{3+\sqrt{4}}=2+\sqrt{5}$ which is an incommensurable length. But it is still possible to find out the precise radius we need for the green arc, as follows:

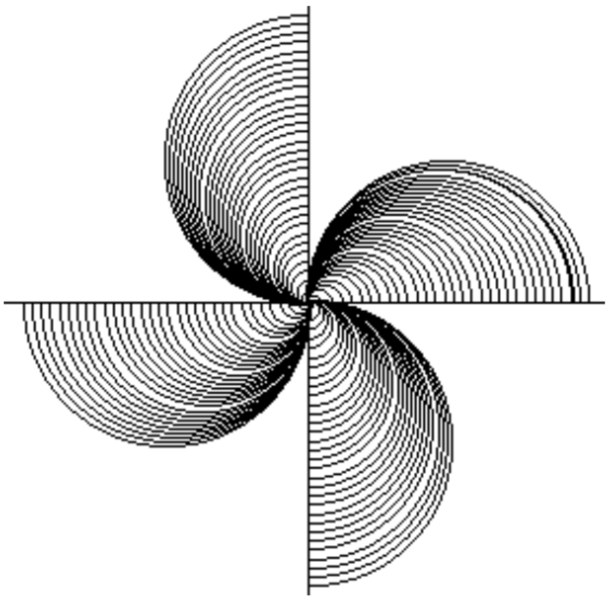
- we know the length of the unit line which is the length from the origin to 1 on the right-hand side;
- we know where $\sqrt{3+\sqrt{4}}$ is located so we can add a line of 2 units to the left of $\sqrt{3+\sqrt{4}}$. We have therefore constructed $2+\sqrt{3+\sqrt{4}}$ on the left-hand side. Note that the addition of line segments is here done in absolute terms. There is no consideration of negative numbers or a negative x-axis. The direction of addition is simply specified as "adding to the left" or "adding to the right".
- we now find the midpoint of the line from $2+\sqrt{3+\sqrt{4}}$ to 1 on the right-hand side in the usual geometric way;
- this midpoint is now the centre of the green arc, hence the green arc can be drawn precisely.

Similarly for the purple curve.

The diagram below shows the construction of roots for

$$\sqrt{1+\sqrt{2+\sqrt{3+\sqrt{\ldots+n}}}}$$

up to $n = 124$.



(from http://www.cut-the-knot.org/arithmetic/constructableExamples.shtml)

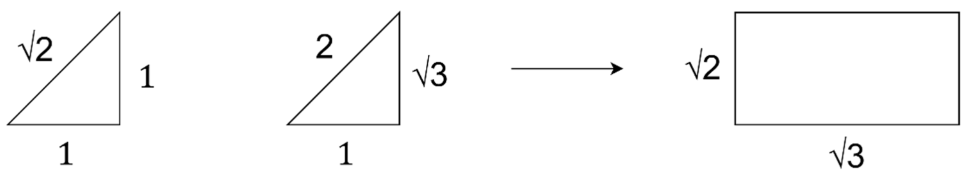
And Vieta (1540 – 1603), a predecessor of Descartes was able to construct the following equation:

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \times \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \times \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}}} \times \dots$$

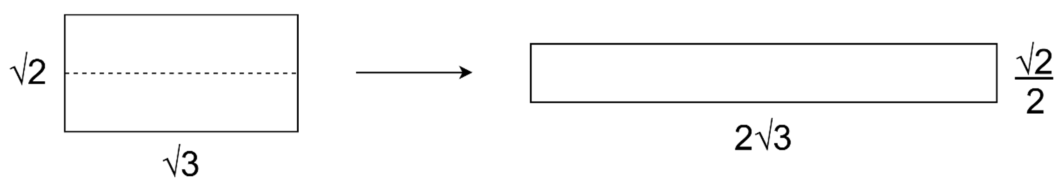
Since π was known to exists as a number in the form of an approximation (if not as an irrational) the case for square roots being numbers was only becoming stronger.

We have seen how to perform geometric arithmetic on whole number and square roots. The interesting thing is that it is also possible to prove the multiplication property of square roots. As illustrated by John Stillwell (p155-157, [87]) it can be shown geometrically that $\sqrt{2} \times \sqrt{3} = \sqrt{6}$. It should be noted that neither the ancients not the early modern mathematicians thought about this.

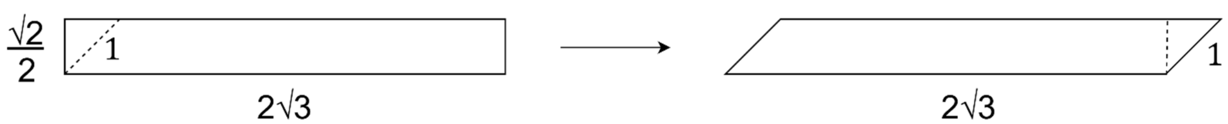
First construct $\sqrt{2}$ and $\sqrt{3}$. We then transfer these two lengths so as to make them perpendicular and then construct a rectangle of height $\sqrt{2}$ and length $\sqrt{3}$.



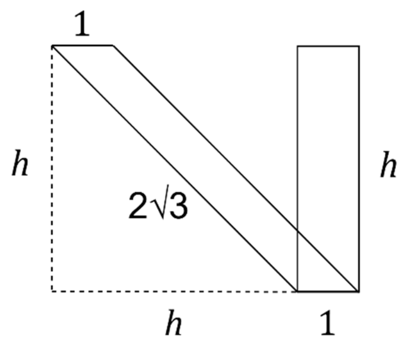
From this we can construct two rectangles of height $\sqrt{2}/2$ and $\sqrt{3}$, and join these two rectangles end to end, as so:



We now construct a right-angle isosceles triangle at one end of the rectangle and transfer it to the other end of the rectangle, viz:



We can now use the slant height of 1 as the base of a rectangle of height h, with $2\sqrt{3}$ as the slant height. We then form an external right-angled isosceles triangle having hypotenuse $2\sqrt{3}$, and height h, as shown below.



From this we obtain

$$2h^2 = \left(2\sqrt{3}\right)^2 \Rightarrow h = \sqrt{6}\,.$$

So a rectangle of sides $\sqrt{2}$ and $\sqrt{3}$, having area $\sqrt{2}\sqrt{3}$, has been transformed into a rectangle of sides 1 and $\sqrt{6}$, having are $\sqrt{6}$. Hence $\sqrt{2}\sqrt{3} = \sqrt{6}$.

Note that we are in a period where we can only perform exact surd arithmetic in geometric form. We can't do exact surd arithmetic numerically. Numeric arithmetic operations on square root numbers allow only for approximate values. Exact numeric arithmetic on surds will have to wait until the development of Dedekind cuts or Cantor's theory of equivalence classes of Cauchy sequences in the late 19th century (see chap 2, [86], for Dedekind cuts and arithmetic, and p846 onwards of [66] for Cauchy sequences and arithmetic). This only acted to further confirm geometry as the foundation of maths.

What the above shows is that it was possible to perform quite complicated geometric arithmetic involving the five basic arithmetic operations. What Descartes did was to subsume arithmetic into geometry since the ops of +. −, ×, /, and $\sqrt{}$ could now be performed geometrically by straight-edge and compass. So we now have a criteria for determining the existence of a number, as described by David Richeson (p242, [74]):

> A point P is a *constructable point* if, starting with our two points, there is a sequence of legal compass-and-straightedge moves [...] such that P is the point of intersection of two constructed curves (lines or circles). A real number a is a constructable number if there exist constructable points P and Q (with $P = Q$ a possibility) such that $|a|$ is the length of segment PQ.

7.8 The geometric constructibility of curves via algebraic equations

During the 16th and 17th century Viete (1541-1603), Descartes and Fermat (1601 – 1665), were among the first to make extensive use of algebra in solving geometric problems. As a result, this led to the acceptance of algebra as a method of analysing and solving geometric problems. One would represent a specific line or curve as an equation, which would then solved the equation, and the resulting solution would then be constructed geometrically so as to justify the validity of the algebraic solution. Algebraic results were not considered to be the actual solution to the problem, hence the requirement that these be constructed geometrically. Two examples will illustrate this, along with an issue arising from requiring everything to be geometrically constructable.

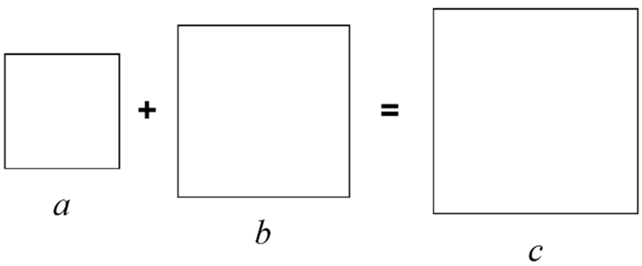
<u>*Example 1*</u>

Consider Pythagoras' theorem. This states a connection between the right-angled sides of a triangle and its hypotenuse. Algebraically this is written as

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

(it should be noted that this is not how the Babylonians and Greeks expressed the relationship. Rather, they described this as a set of written instructions which defined the procedure for establishing such a relationship).

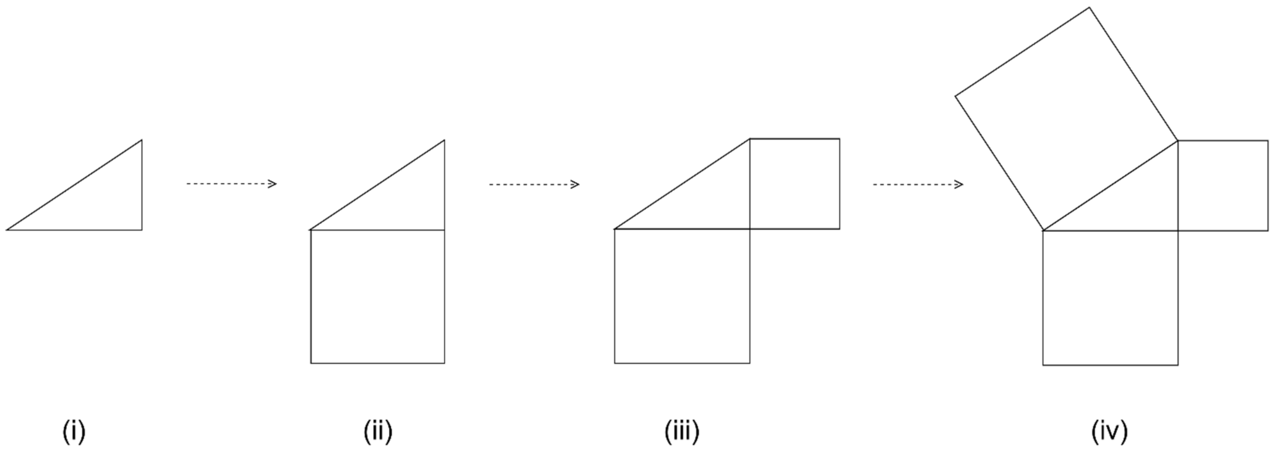
The issue is to show that such a piece of algebra can be constructed geometrically. In geometric terms the equation says is that the sum of two squares equals a third square. This means that we need to be able to construct the following:



To do this we

- construct a right-angled triangle (diagram (i) below);
- then we construct a square on the base of the triangle using the base as the required length (diagram (ii) below);
- then we construct a square on the side of the triangle using the side as the required length (diagram (iii) below);

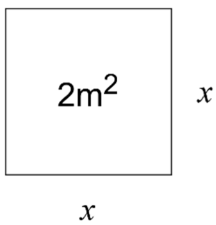
- finally we construct a square on the hypotenuse of the triangle using the hypotenuse as the required length (diagram (iv) below).



Since all the elements of the diagrams above (lines and right-angles) are geometrically constructable we have confirmed geometrically the algebraic relationship $c^2 = a^2 + b^2$.

Example 2

Consider the problem of finding the side of a square of area 2m². The geometric set-up is as below:



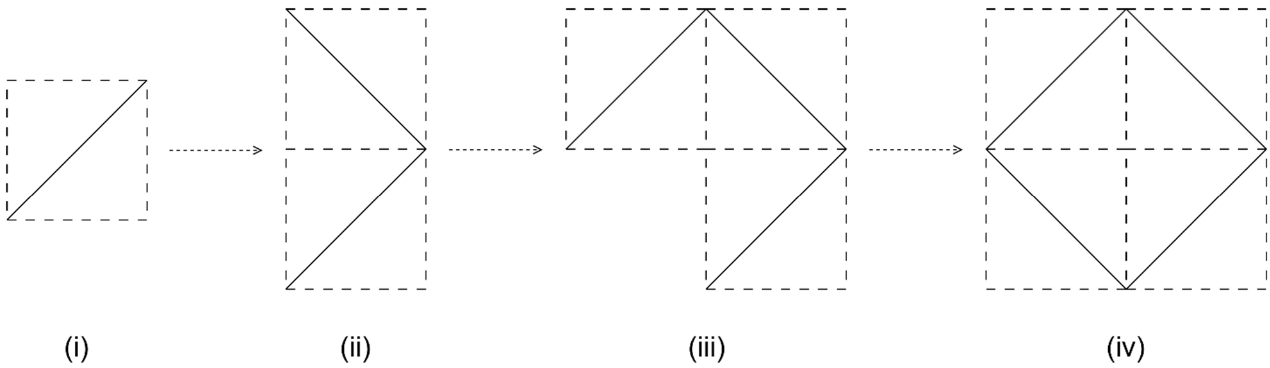
The algebraic equation which represents this situation is

$$x^2 = 2.$$

We perform algebra to obtain $x = \sqrt{2}$ (we ignore the negative root since, in the time of Descartes, negative numbers were not considered to exist). The question now is, Is this algebraic solution constructable? Just because we have moved symbols around does not mean that the result is valid. We have to show that such a result can be constructed by straight edge and compass. This can be shown as follows:

1) Construct a unit square, illustrated in blue in the diagram below;
2) By Pythagoras' theorem (which has been proved constructable in example 1 above) the diagonal of the square can be assigned the number $\sqrt{2}$ (diagram (i) below);

3) On top of the square just constructed, construct another unit square along with its diagonal, such that the diagonal square is perpendicular to that of the first square (diagram (ii) below);
4) Repeat the construction of unit squares so that the newly constructed diagonals are perpendicular to the previously constructed diagonals (diagrams (iii) and (iv) below).



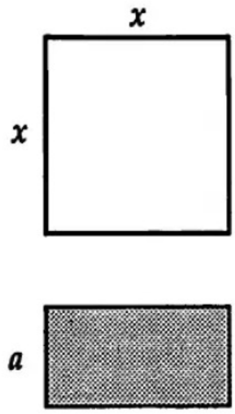
The final construction of diagram (iv) shows a square (solid lines) of side $\sqrt{2}$ whose area is 2m^2.

We now see how much quicker it is to perform the algebra of $x^2 = 2$ to $x = \sqrt{2}$ than it is to find and perform a sequence of geometric transformations. This is exactly what Descartes (as well as Viète and Fermat) by his use of algebra. He was using algebra as a tool to show that a geometric solution existed without the need to go through the intermediate stages geometric manipulations.

Example 3

A major issue with algebra was that, although symbols represented geometric objects such as lines, and although there was now a geometric arithmetic, this did not automatically imply that moving symbols around (i.e. performing algebra) actually represented anything geometrically meaningful.

For example, we can construct a square and a rectangle, one side of which is the same length of the side of the square, as shown below.



We now set up an algebraic expression to represent this configuration. To do this we label the side of the square as x and the height of the rectangle as a. This then gives us an algebraic expression for the areas to be x^2 and ax.

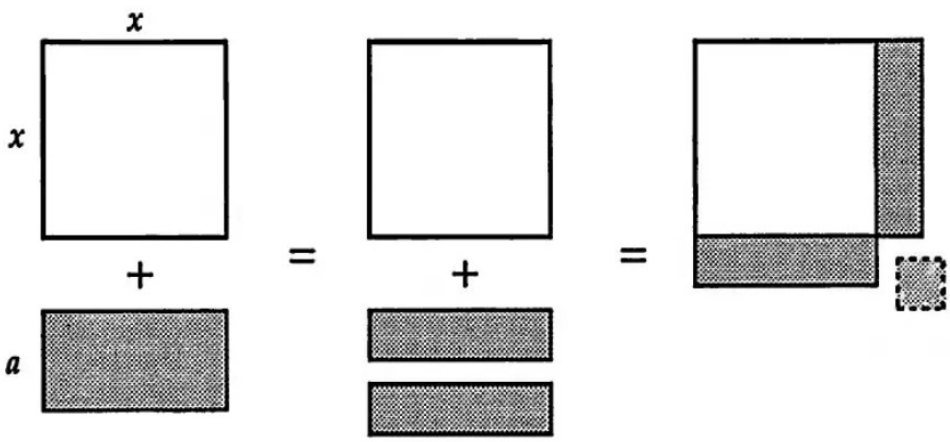
We have the idea that we can add these two expressions and transform the sum into another expressing the difference of squares, i.e.

$$x^2 + ax = x^2 + ax + \left(\frac{a}{2}\right)^2 - \left(\frac{a}{2}\right)^2 = \left(x + \frac{a}{2}\right)^2 - \left(\frac{a}{2}\right)^2.$$

Although we have manipulated symbols (added and subtracted extra terms and factorised) is the final answer on the right-hand side geometrically constructible? If it is our algebraic manipulation of symbols has produced a valid answer, otherwise it hasn't.

Just because the initial algebraic expression $x^2 + ax$ was geometrically meaningful doesn't mean to say that the final result will be able to be represented geometrically. Every manipulation of symbols (by the process of algebra) is supposed to represent the geometric manipulation of lines (or curves) using straight-edge and compass. This means that there should exist not only a geometric representation of the final expression $\left(x + \frac{a}{2}\right)^2 - \left(\frac{a}{2}\right)^2$. But that such an expression should be constructible by a sequence of geometric transformations using the straight-edge and compass (just as there is a sequence of algebraic steps from $x^2 + ax$ to $\left(x + \frac{a}{2}\right)^2 - \left(\frac{a}{2}\right)^2$.

The geometric transformation is shown below. It is a simplified version of what would actually have to be done using straight-edge and compass, but each object below (along with translations and transformations) can indeed be constructed geometrically.



Note that the little square with dashed sides represent the subtraction of ($a/2)^2$.

Example 4

This example comes from p12-15 of [25]. In this section of his book, Descartes gives a number of hypothetical examples of how to geometrically construct the solutions to quadratic and quartic equations. One such equations is

$$x^2 = ax + b^2$$

where a and b were line segments of known lengths, and x is a line segment whose length satisfies the equations above. Today we would construct this by drawing the parabola x^2 and the straight line $ax + b^2$ and locating their point of intersection.

But parabolas are not constructable by straight edge and compass. So, Descartes states the answer to x to be

$$x = \frac{a}{2} + \sqrt{\frac{a^2}{4} + b^2}\,,$$

but does not actually show how he obtains this algebraic answer nor how to construct it geometrically (note the Euclid had shown the geometric version of this formula in Book II, prop. 6). However, Descartes does state that this answer is constructable as can be seen as follows, where only lines, circles and triangles are used. Firstly, construct a right-angled triangle NLM with NL = $a/2$ and LM = b (diagram (i)). Then constructs a circle with centre N and radius NL (diagram (ii)).

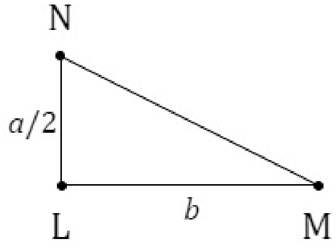


Diagram (i)

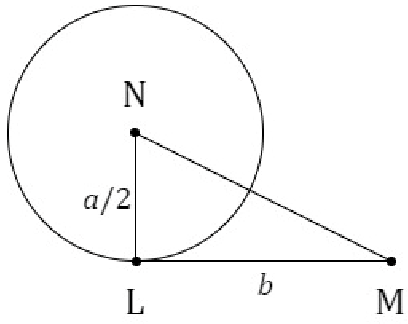


Diagram (ii)

Then extends MN to O on the circumference of the circle (diagram (iii))

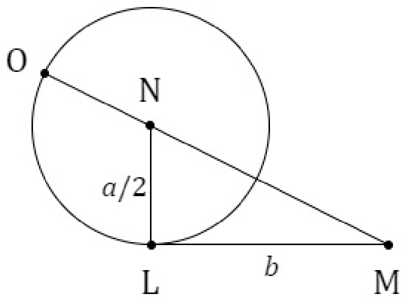


Diagram (iii)

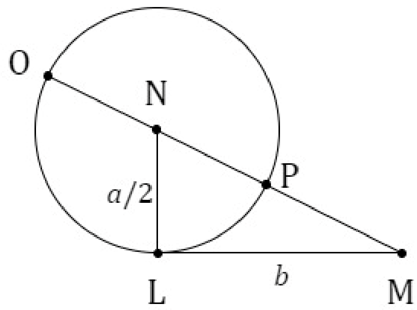


Diagram (iv)

We can then perform the following analysis

- By Pythagoras' theorem we have $\mathrm{MN}^2 = \mathrm{LN}^2 + \mathrm{LM}^2$, i.e. $\mathrm{MN}^2 = \left(\frac{a}{2}\right)^2 + b^2$ so that $MO = \mathrm{ON} + \mathrm{MN} = \frac{a}{2} + \sqrt{\frac{a^2}{4} + b^2}$;
- But $\mathrm{MN} = \mathrm{MO} - \mathrm{ON}$, so that on substituting into our previous equation we have $(\mathrm{MO} - \mathrm{ON})^2 = \left(\frac{a}{2}\right)^2 + b^2$ which simplifies to $\mathrm{MO}^2 = a.\mathrm{MO} + b^2$. Comparing this with $x^2 = ax + b^2$ we see that the required answer x is line MO.

Descartes ignores the other solution (which is line PM in diagram (iv) above) since the other algebraic solution $x = a/2 - \sqrt{}(a^2/4 + b^2)$ is negative. Geometrically this represents NL − NM which impossible to construct since we can't subtract a longer line from a shorter line. But this is where things become contradictory because PM is indeed an actual measurable line segment despite the fact that its construction as NL − NM is not possible directly. Such was the attitude towards numbers in this period.

Many other mathematicians of the age (Viete (1540 – 1603), Clavius (1538 – 1612), and Ghetaldi (1565 – 1626) to name three) also solved quadratics and cubics using combinations of lines and circles. The point here is, whereas today we have one standard equation for a general quadratic and cubic, these specified by the variable of highest degree being 2 and 3 respectively, this was not the case up until the 17th century. From the time of Al-Kwarizmi (c.780 – c.850) and his classification of quadratics, and Khayyam (1048–1131) and his classification of cubics, up to the 17th century, quadratic and cubic equations were classified by their geometric constructability.

It should be understood that quadratics and cubics weren't just equations. They were descriptions of geometric constructions, and thus of sizes of lines, surfaces and volumes. As such only constructions giving rise to positive roots were considered (any other solution being ignored). It is for this reason that the quadratic $ax^2 + bx + c = 0$, instead of being solved algebraically in general terms to obtain the result $x = -\frac{b \pm \sqrt{b^2 - 4ac}}{2a}$, was classified quadratic into three forms:

i) $x^2 + ax = b^2$: the square and the side equals a number;
ii) $x^2 - ax = b^2$: the square less the side equals a number;
iii) $x^2 + b^2 = ax$: the square and the number equals the side.

As for cubic equations, Khayyam had classified these into three categories of two terms, three terms, and four terms, examples of which include $x^3 = cx$, $x^3 = cx + d$ and $x^3 = bx^2 + cx + d$. He had nineteen different forms in all.

All of this demonstrate the length mathematicians of the day (and for some time later) had to go to confirm their algebraic solutions as being geometically valid.

7.9 Extending the idea of what is acceptable geometric construction

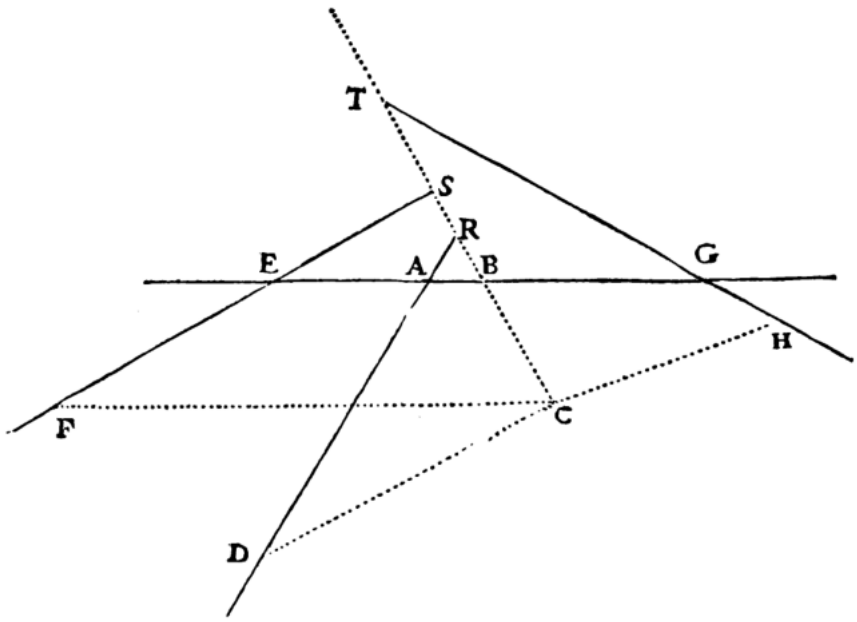
With Descartes' particular application of algebra, not only were algebraic results representing lines and circles obtained, but other types of algebraic results were obtained which represented parabolas and "exotic" curves. As a result of this, it stood to reason that if algebra could produce equations which were geometrically constructable then algebraic equations producing curves considered non-constructable should now be considered as constructable. So it seems as if algebra was a unifying approach to what could be considered constructable, because it (algebra) highlighted a pattern:

Algebraic expression	**Geometric equivalent**	**Geometric representation**	**Instrument**
x^0	A point	▪	(nib of) Pencil

x, as in $y = mx + c$	A line		Straight edge
x^2, as in $x^2 + y^2 = r^2$	A circle or arc of a circle	or	compass
x^2 on its own	A parabola		**?**

It would then simply be a case of manufacturing new instruments to effect the construction of parabolas.

An example of a problem whose geometric set-up consists only of straight lines but whose solution is not constructable is that of Pappus (~290AD - ~350AD) four-line problem, illustrated below (it is called the four-line problem because four solid lines define the initial geometric set-up). Descartes was able to transform Pappus' problem into algebraic form and solve it much more efficiently as a result.



(diagram taken from p27, [25] where I have added the letters x and y)

A detailed solution to this problem can be found in sec 10. Simply put, the aim of the problem is to find the path traced out by point C given that there is a constraint placed on the distance from C to all four solid lines (these distances represented by the dashed lines).

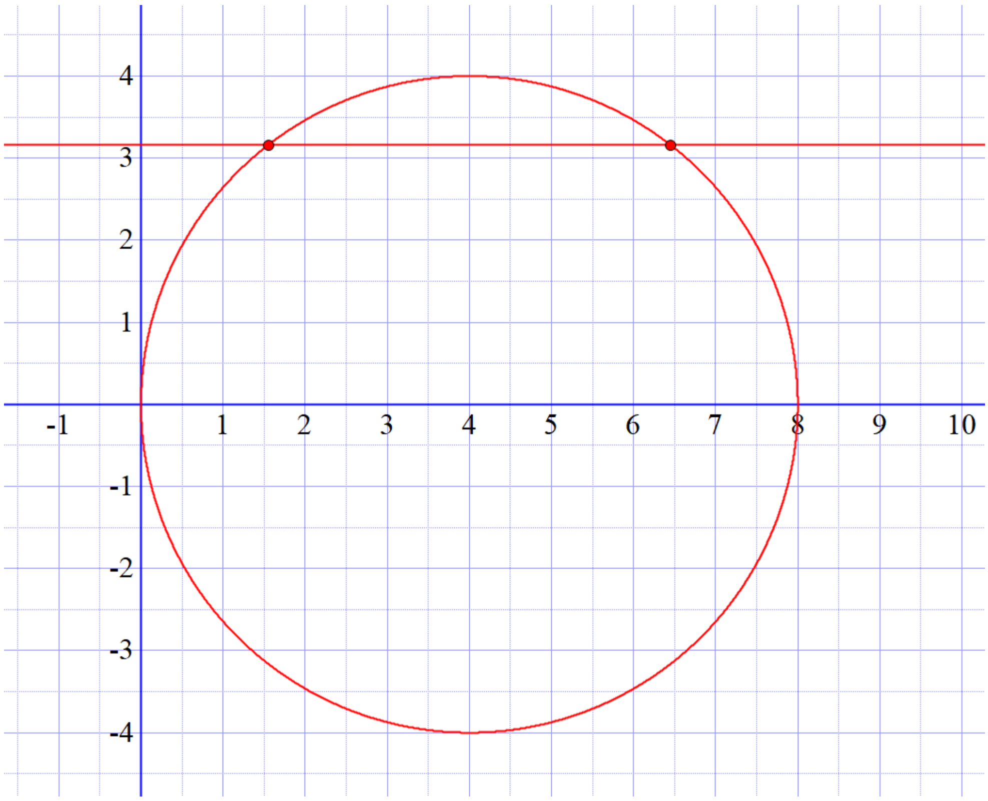
Descartes was able to represent all relevant lines of the geometry into linear equations, perform relevant algebra on these equations, and produce as an equation which was a parabola. The important point to note is that the above set-up is clearly constructable since it consists simply of straight lines. Yet the solution (known to Pappus) is a parabola, and therefore classified as non-constructable. But by the algebraic pattern x^0, x^1, x^2 shown in the table above if a circle, which is constructable, can be represented algebraically using a term in x^2 why can't a parabola, also represented by a term in x^2, be considered constructable in a broader sense? So it is that, although Descartes knew his parabola could not be constructed by straight edge and compass, he did not accept that constructability was limited to the two instruments of straight edge and compass alone.

His reason for thinking this can be illustrated by the fact that he had solved quadratic equations (parabolas) using only lines and circles. This can be verified when we see that $x^2 + bx + c = 0$ can be transformed by first completing the square: $(x - b/2)^2 - b^2/4 + c = 0$. Then we have $(x - b/2)^2 + c = (b/2)^2$. If we let $y = \sqrt{c}$ then we can write this last equation in circle form:

$$\left(x - \frac{b}{2}\right)^2 + y^2 = \left(\frac{b}{2}\right)^2.$$

Solving $x^2 + bx + c = 0$ is the same as finding the intersection of the line $y = \sqrt{c}$ and the circle, both of which are constructable. Descartes had effectively reduced a non-constructable curve to two constructable curves.

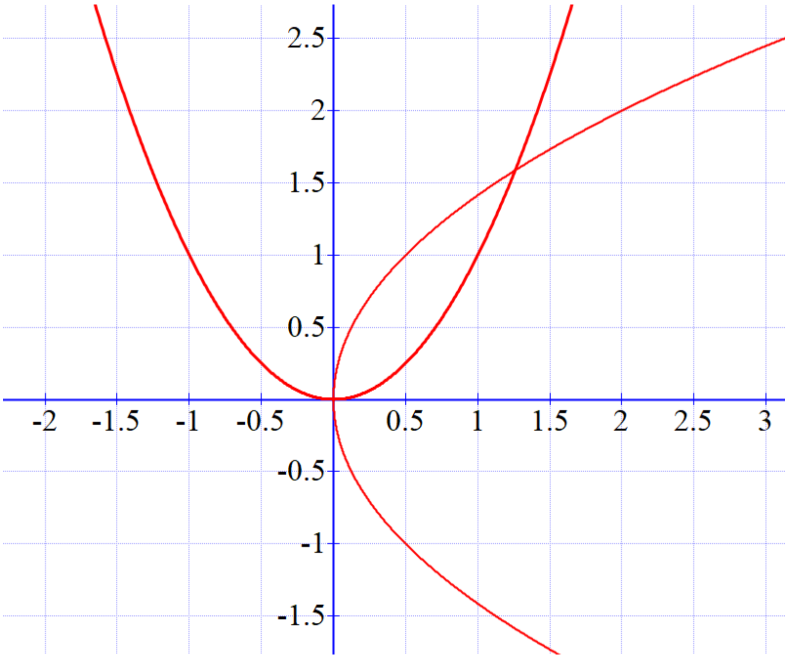
For example, in order to geometrically find the roots of a quadratic equation one would need to plot the quadratic so that we could read off the points of intersection of the quadratic with the x-axis. Since quadratics were not constructable Descartes used the process above as a work-around. So, in wanting to solve an equation such as $x^2 - 8x + 10 = 0$ he would have completed the square to get $(x - 4)^2 - 16 + 10 = 0$, then written this as $(x - 4)^2 + \left(\sqrt{10}\right)^2 = 4^2$. Letting $y = \sqrt{10}$ he would then have $(x - 4)^2 + y^2 = 4^2$ which is the equation of a circle. He could then find the solution to the quadratic as the intersection of the circle with the line $y = \sqrt{10}$, as illustrated below.



The non-constructable quadratic had now been solved by reducing it to two constructable curves, the intersection of which constructs two points which therefore constructed two numbers as the solution to the original quadratic.

In Descartes' mind it therefore stood to reason that any solution derived by a combination of constructable curves must itself have come from a constructable curve. Thus any curve could be classed as constructable if it could be reduced to previously constructable curves. With a kind of bootstrapping mindset, quadratics (and parabolas in general) were constructable since they could be reduced to lines and circles. This means that parabolas could be form part of the family of consrtuctable curves used to solve more complicated problems involving equations of higher degrees.

With this new outlook Descartes claimed $\sqrt[3]{2}$ to be constructable. This number can come as the solution to $x^3 - 2 = 0$. Such a solution cannot be constructed geometrically by straight edge and compass. But Descartes was able to solve such an equation via intersecting parabolas. Simply construct $y = x^2$ and $y^2 = 2x$. The intersection of these two parabolas is $\sqrt[3]{2}$ (verified algebraically by substituting $y = x^2$ into $y^2 = 2x$). This is illustrated in the diagram below. So, all that is required is to build an instrument which allows us to draw parabolas (Descartes did indeed build such an instrument. See [34], p90).



The construction of $\sqrt[3]{2}$ *using previously constructed parabolas*

Descartes was even able to solve cubics of the form $x^3 - ax + b = 0$ using only circle and line constructions, but only to the extent of finding two roots. The intersection of chords with the enclosing circle would then mark the roots (see p212, [25] or section 3, [70]). Nickalls ([70]) superimposes Descartes' construction onto the cubic to show that the latter's construction does indeed identify two of the three roots. Then, following Descartes' logic, cubics of the form $x^3 - ax + b = 0$ should now be considered constructable.

As an example of solving higher degree equation, consider a geometric problem which gives rise to wanting to solve

$$x^4 - 23x^2 - 18x + 40 = 0$$

If we can convert this equation into algebraic forms which represent lines and curves we already know then the aforementioned equation will be geometrically constructable and therefore solvable. So, split the term in x^2 from the form cx^2 to $(c-1)x^2$ and x^2, i.e.

$$x^4 - 24x^2 + x^2 - 18x + 40 = 0\,.$$

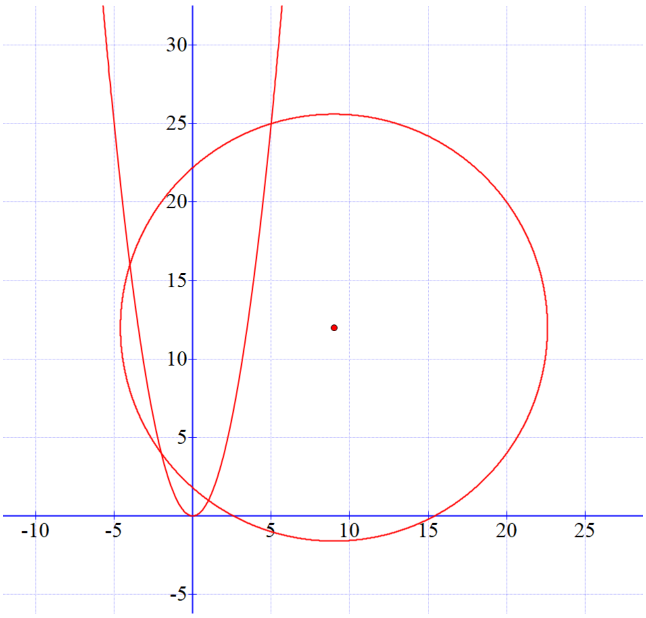
Now substitute $y = x^2$ for the first two terms only to obtain

$$y^2 - 24y + x^2 - 18x + 40 = 0\,.$$

Then

$$
\begin{aligned}
y^2 - 24y + x^2 - 18x &= -40 \\
(y-12)^2 - 144 + (x-9)^2 - 81 &= 36 \\
(y-12)^2 + (x-9)^2 &= 185
\end{aligned}
$$

which is a circle centre $(9, 12)$ and radius $\sqrt{185}$. We know circles are constructable, and from the previous example we know parabolas are constructable. So, the solution to the pair of equation $(y-12)^2 + (x-9)^2 = 185$ and $y = x^2$ are the points of intersection of the circle with the parabola. This means that our quartic equation is constructable. This is illustrated below in the modern context of Cartesian coordinates (note that, contrary to popular belief, Descartes did not invent the Cartesian coordinate system):

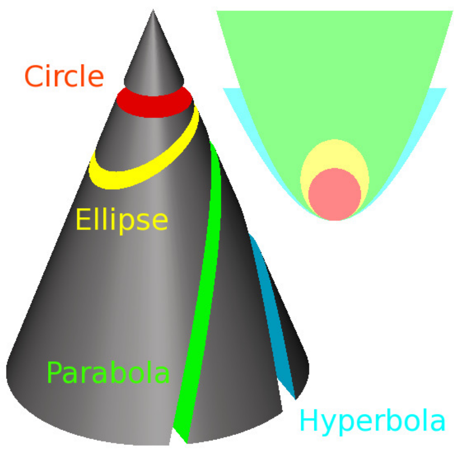


Descartes was also able to solve problems whose solutions were fifth- and sixth-degree equations (see p220 onwards of [25] for Descartes detailed (and very involved) description of how to solve a general 6th degree equation, and p222 for the actual geometric construction of the solution). It only required inventing new devices to actually draw the curve represented by the equation acting as the solution to the problem.

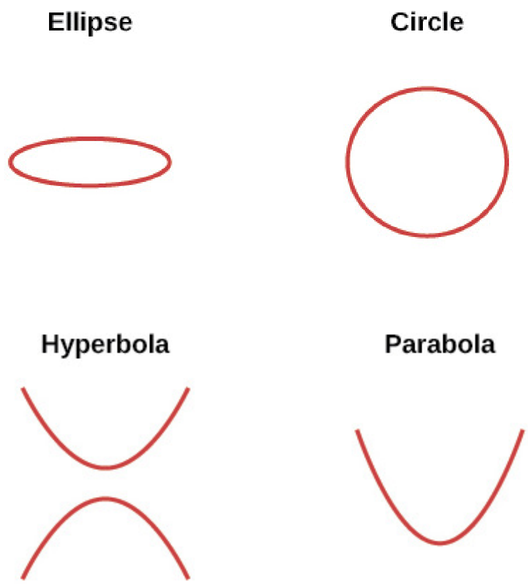
And the building of such instruments confirmed to Descartes the viability of drawing such curves, thus confirming the validity of the algebraic solution. Thus, in Descartes' eyes, this justified his solution as being *geometrically* constructable even though they were not constructable in the ancient classical sense.

It should be noted that Descartes was still a geometer. Even though he worked with algebra to obtain numbers and equations as solutions, these solutions were never the final answer. They only ever represented points, lines or curves. All algebraic solutions had still to be shown as being constructable. As Grabiner [34] says, "[the solution] is this curve, it has this equation, and it can be constructed in this way".

As a side note, despite the parabola not being constructable, it was still accepted by the ancients as a curve. The "construction" of the parabola, as well as the ellipse and hyperbola, was realised by using a plane to slice a cone such that the plane was parallel to the slant side of the cone. This is illustrated by the green part of the diagram below.



(a)



(b)

The circle can also be so "constructed" and should therefore not be considered constructable in the classical sense. Yet it was. The reason for considering the circle as constructable is again down to Plato's philosophy of ideal forms. Since some kind of instrument was required to construct geometric objects, the compass was the instrument which introduces the least amount of corruption when translating an ideal circle, or circular arc, to the real world. Hence the only conic section considered constructable was the circle. All other conic sections would require more elaborate instruments for their construction, corrupting their ideal forms even further due to the accumulation of corrupting influences caused by ever more linkages and straight edge.

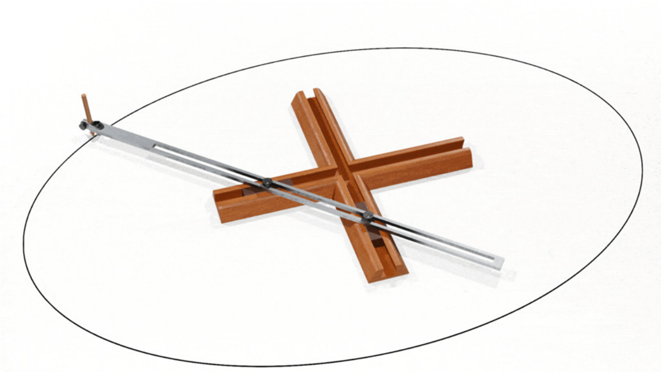
7.10 New instruments used to extend the class of geometric curves

It should be noted that Pappus (c. 290 AD – c. 350 AD) had already solved the 2-line, 3-line and 4-line problem but only by geometric means. Also, he could not go beyond the 4-line problem because this would result in what we now know to be polynomials of degree four or higher. Since the degree of a polynomial represented its dimension in space (x represents a line, x^2 represents a square, and x^3 represents a cube) polynomials of degree four or higher could not be represented geometrically.

The consequences of Descartes' algebraic approach was then not only to confirm Pappus' solutions but also to allow him to expand the class of what could be considered constructable to include curves that could be drawn using any instrument that could be built to draw such a curve (this is not completely true since there were instruments which could draw spirals and conchoids, but Descartes did not consider these to be geometric curves).

Some of these instruments, shown below, consisted of multiple straight edges and compass, and linkages which allowed for coordinate motion. This meant that as one part of a straight edge moved so did another part move. By such motion varied curves could be drawn. It should be noted that even some of the ancients, such as Nicomedes (c.280 BC – c.210 BC), Pappus and Archimedes either built or designed such devices.

- *The trammel of Archimedes.* This is a devices which allows us to construct ellipses.

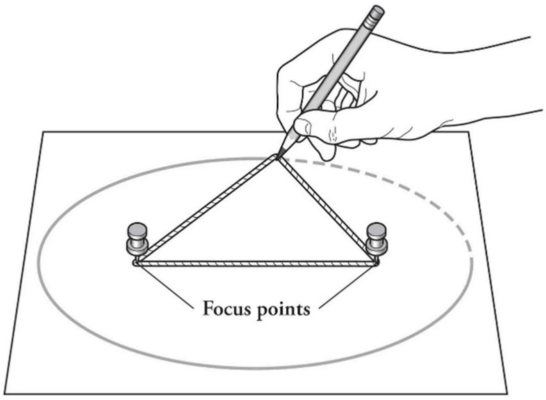


(taken from https://en.wikipedia.org/wiki/Ellipsograph)

It consists of a cross with channels in it, as well as a grooved straight edge connected to sliders inside the channels of the cross, such that the whole mechanism can move freely.

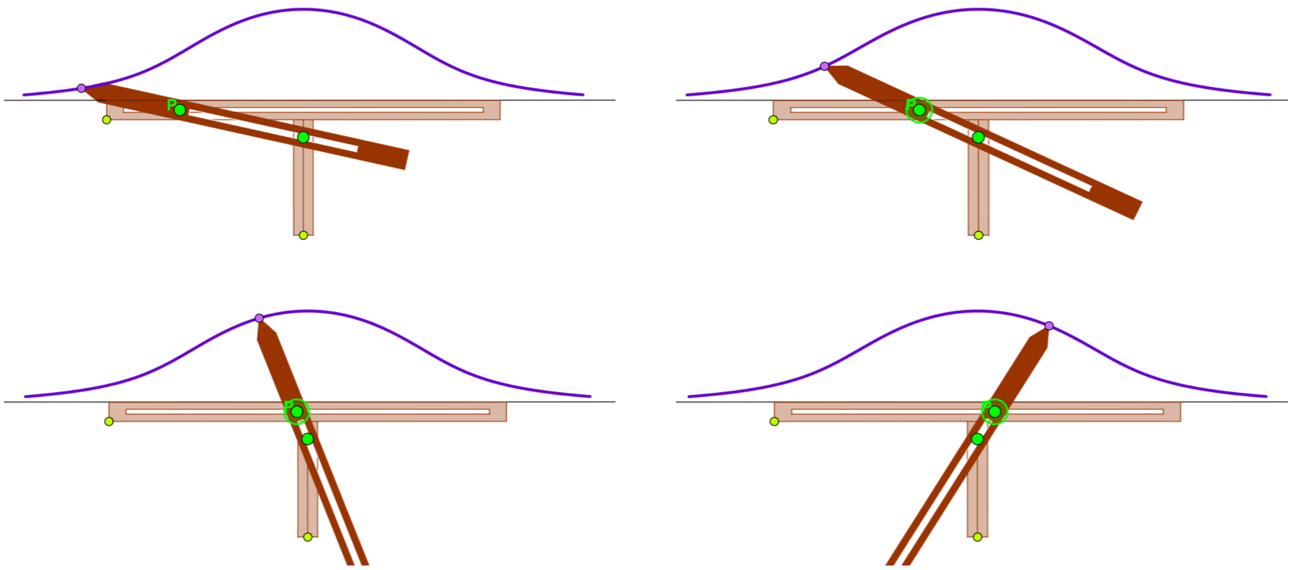
The motion of a pencil placed at the end of the grooved straight edge forces the pencil to trace out an ellipse.

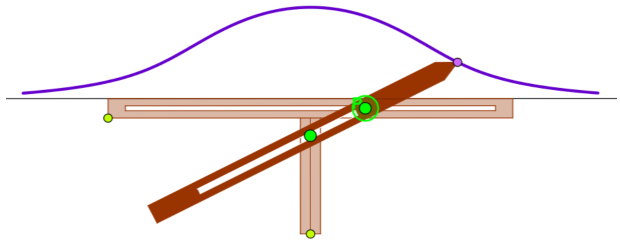
The other classic method for constructing an ellipse, which was also know to the Greeks, is the two-pin and string approach seen below.

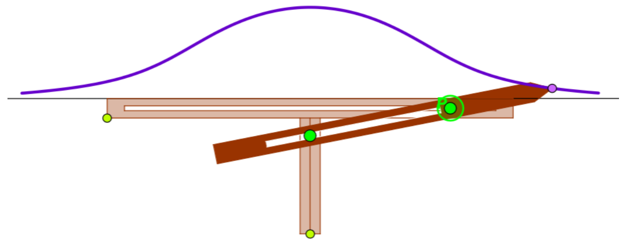


(from http://www.loop-the-game.com/snoop)

- *The trammel of Nicomedes*. This device was built in order to construct the conchoid. It consists of a T-shape with channels in it, and linkages (in green) in each channel. A grooved straight edge (the dark brown straight edge in the diagram below) is then connected to the linkages so that the straight edge can move freely. A pen is placed at the tip of this straight edge so that when the straight edge is made to move it traces out the path shown by the blue curve (images taken from Ramon Nolla at https://www.geogebra.org/m/JZhheAUU)







- Diagram (i) illustrates Descartes' own device for constructing parabolas.

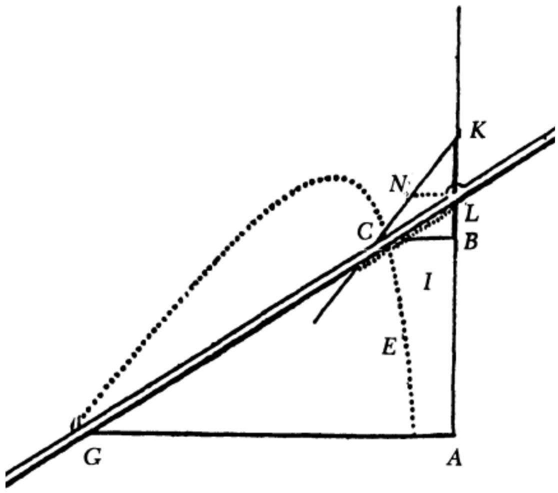


Diagram (i) ([7])

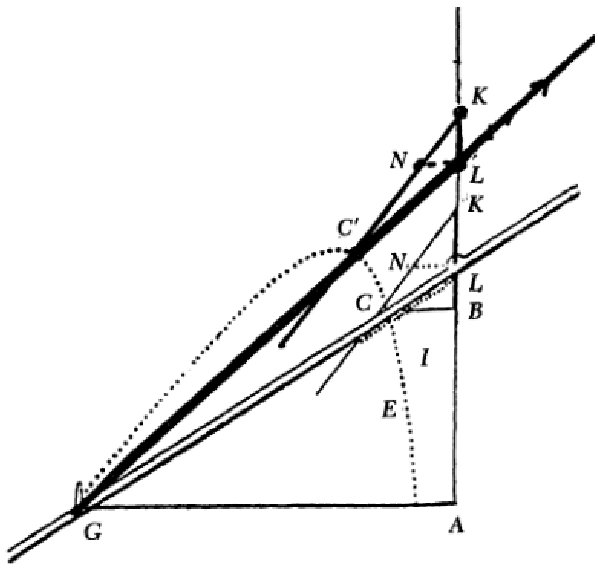
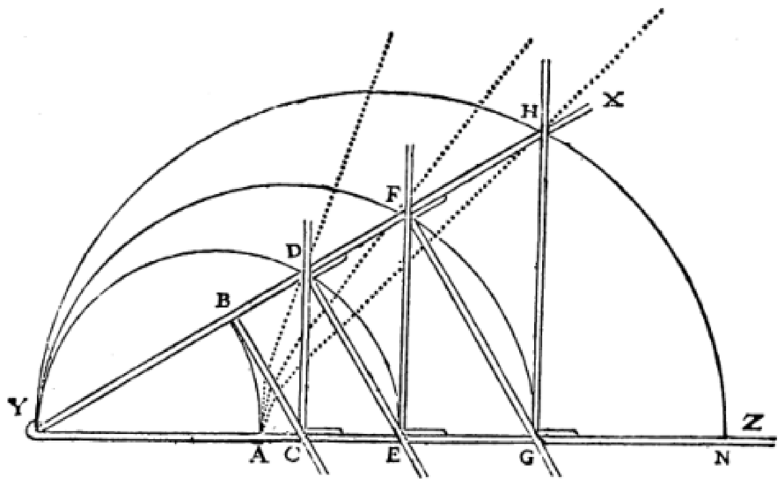


Diagram (ii) ([34])

This device consists of a straight edge GL which is hinged at G, allowing the straight edge to rotate. Furthermore, the point L, on the vertical from A, can slide along GL. Attached to L is a rigid devices NKL which can move up and down the vertical from A. As GL moves up and down the vertical from A, point C (on the extension of KN) will trace the path of a parabola (in order to more easily see this dynamic, Grabiner (p91, [34]) makes an addition to the diagram above to include the straight edge, KL and KN in a second position. This is diagram (ii) above).

- *Descartes' mesolabe* (p46, [25]):



This device allows us to find roots (what was known in Descartes' day as mean proportionals) and thus helps in solving certain algebraic equations. The structure of the device consists of rods YX, YZ and all the diagonal and vertical rods. A rod BC is fixed at B, and rods CD, EF, and GH are made to be perpendicular to YZ but are free to move along YZ. Similarly, for rods DE and FG along YX. Finally, rod YX can rotate about YZ, so that when YX is rotated anticlockwise BC pushes CD along YZ, which pushes DE along YX, etc.

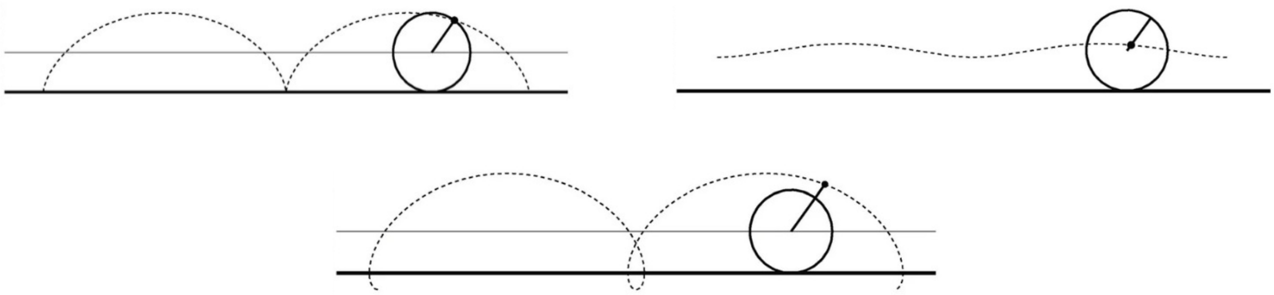
This configuration of rods gives rise to many similar triangles which therefore allow a number of ratios to be set up between line elements. It is these ratios which leads to finding cube roots and setting up cubic equations (see sec 11 for examples of these). The more rods we have the more similar triangles we can set up, and the more ratios can be compared, leading to roots of higher order.

The focus on building so many different and exotic compasses to construct algebraic equations and solutions illustrates the extent and degree to which geometry was still considered to be the essence of mathematics.

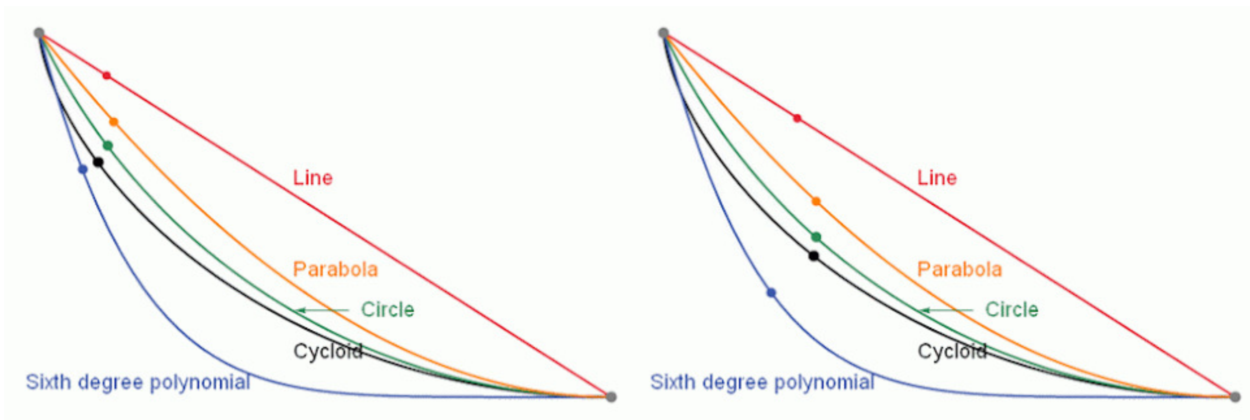
> "By the renaissance, the term "compass" (or "compasses") referred to a variety of tools that could draw circles, ellipses, and other conic sections (drawing compasses), transfer distances (dividers), scale figures by some fixed ratio (reduction compasses), measure spherical or cylindrical objects, transfer distances on maps (three-legged compasses), and perform calculations involving ratios (proportional compasses)." (p228, [74]).

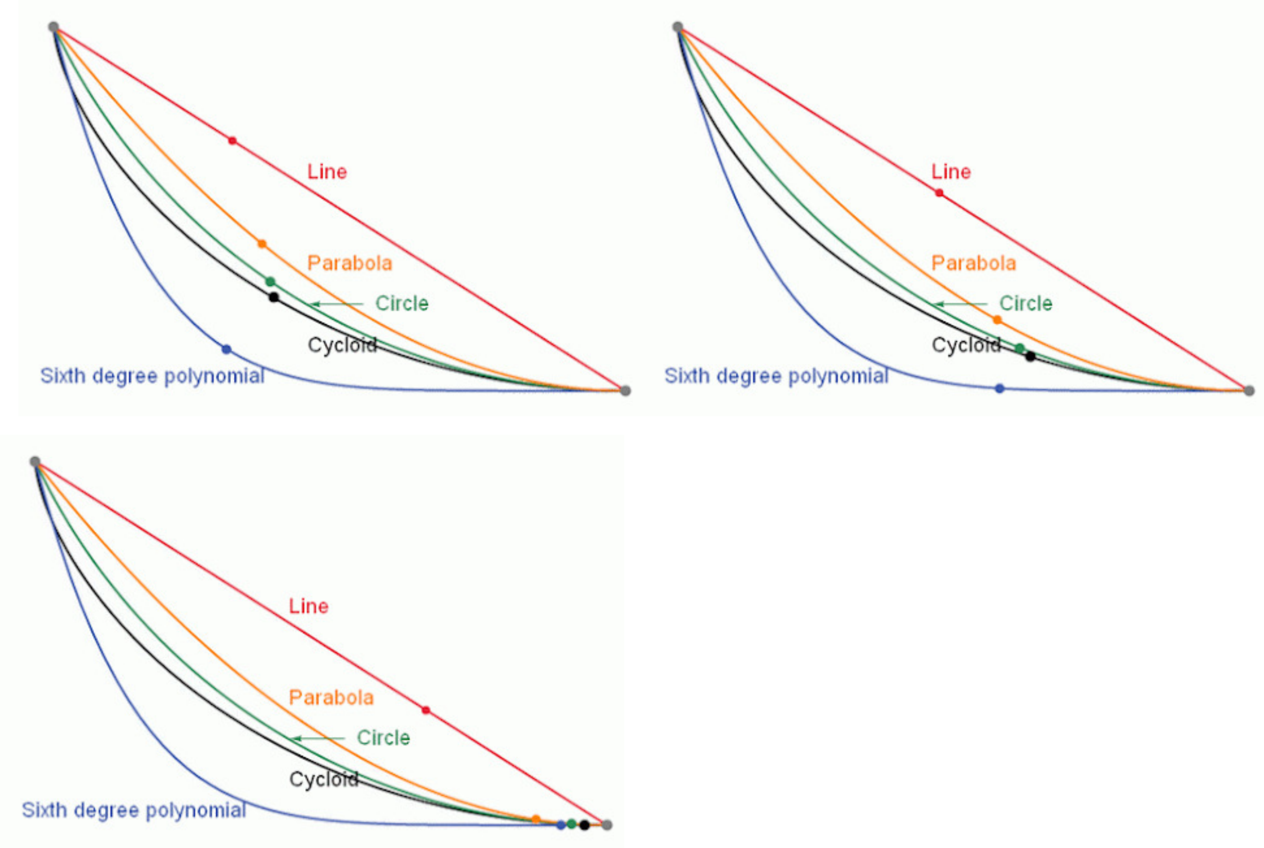
The need to confirm algebraic solutions geometrically continued into the 18^{th} century. Johann Bernoulli (1667–1748) is credited as the person having solved what is known as the brachistochrone problem, or the curve of fastest descent. This problem requires one to find the curve between two points of different heights such that a particle falls under gravity from one point to the other in shortest time (i.e. fastest speed).

Bernoulli used calculus to solve this, obtaining the curve known as the cycloid. This is in fact a family of curves three of which are shown below. The standard shape of the curve is formed as the path traced out by a fixed point located on the circumference of a circle as the circle rolls (without slipping). Depending on where the point is located, either inside, on, or outside the circle, different types of cycloids are formed.



The fact that the cycloid is the curve of fastest descent can be seen in the sequences of images below, taken from a simulation where the black curve is a segment of the cycloid curve and the black point a particle falling along the cure under gravity.

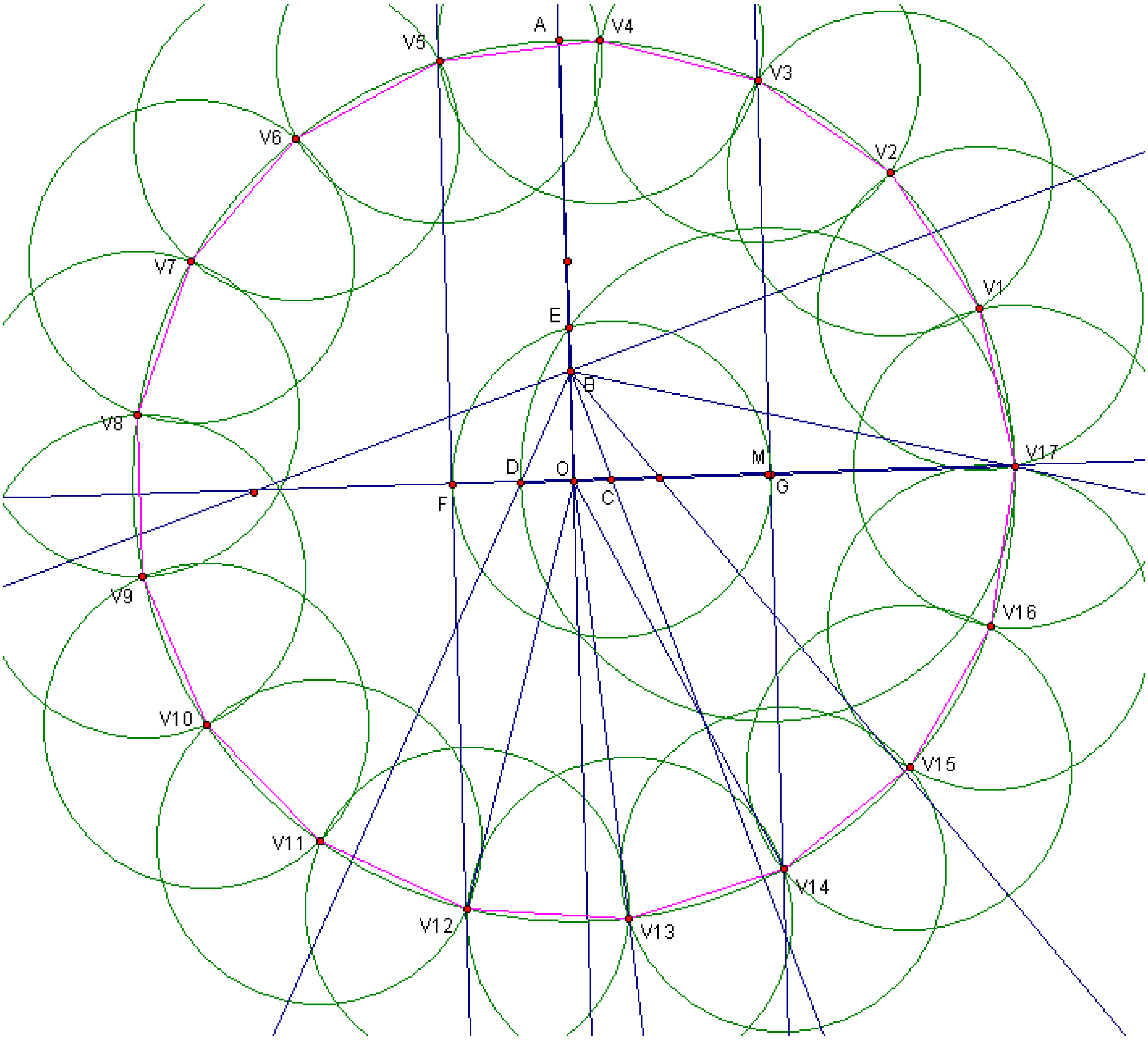




https://medium.com/@priyanshu.pansari/the-brachistochrone-a-mathematical-journey-through-time-and-space-fd8c921e891e

It is worth re-emphasising that although a machine can easily be constructed to trace out the path of a cycloid, such a machine, and therefore the cycloid itself, are inadmissible in classic geometry simply because it generates motion. Curves traced out by moving points do not count as constructable curves. Bernoulli then had to confirm his algebraic solution geometrically in the usual manner of constructing a static geometric configuration.

One of the more famous examples of using algebra to show geometric constructability is that of Gauss's proof that a 17-side polygon is geometrically constructable. Here Gauss used complex numbers to prove this (see chapter 19 of [74] for details) after which it remained to find the actual way of constructing the polygon by straight edge and compass. The procedure for this is quite involved since one needs to construct 17 points on the circumference of a circle, all point being equidistance form each other. At the end of this process one ends up with a complete construction as shown below. The 17-sided polygon is shown in pink where its vertices are the red points labelled V1 to V17 on the large circle centre O.



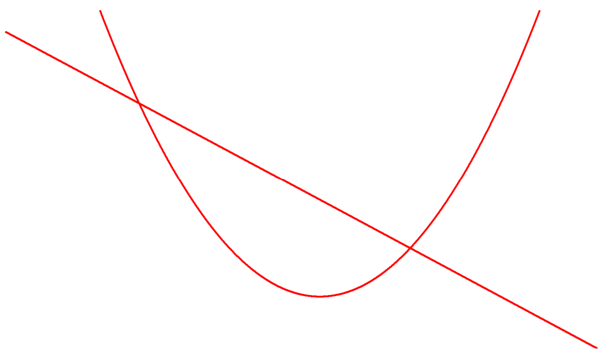
(From https://robertlovespi.net/tag/heptadecagon/)

And even in 1863, Dirichlet, a foremost number theorist, proves $ab = ab$ using the geometric argument of forming a rectangular array of the same number c and then counting the number of rows b versus the number of columns a. A copy of Dirichlet's own text is reproduced in sec 12.

7.11 *Algebra gives absurd results*

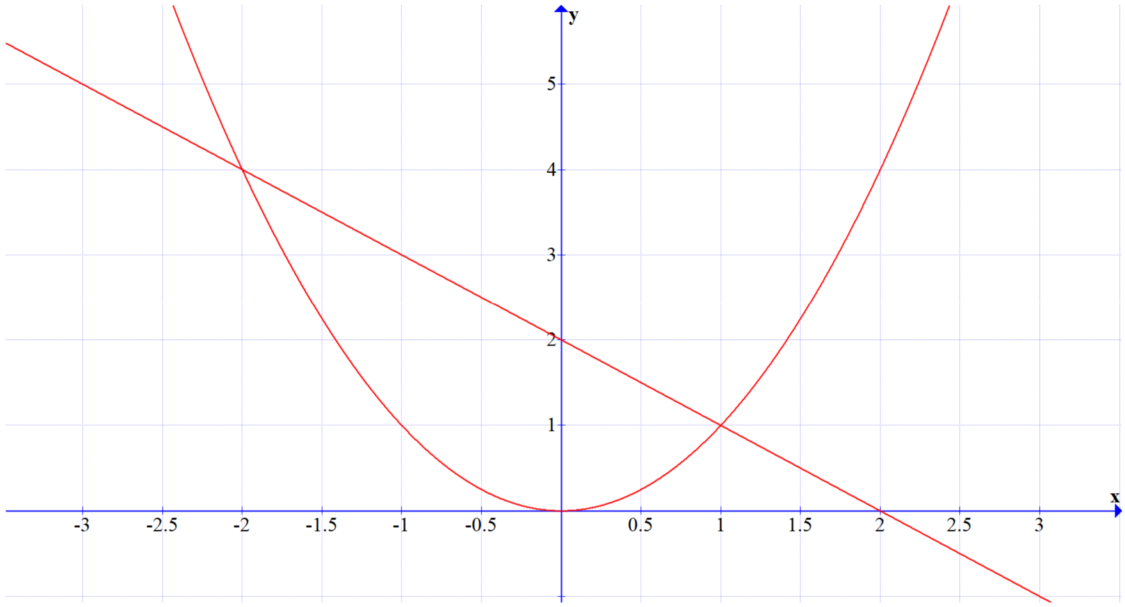
The case for algebra as a method whose results could be considered valid in their own right, and therefore as a method independent of geometry, was not helped by the fact that it could give what was then considered absurd or impossible results. These results were negative and complex numbers. The reason for interpreting such results as absurd or impossible was that numbers, by which is understood natural numbers, were inextricably associated with physical situations. Mathematics as a whole was seen as modelling the physical world. So it was that geometry modelled the physical world, algebraic equations represented specific geometric configurations of lines, the variables of those equations were shorthand for geometric elements of said configurations, and the resulting numerical answers to algebraic manipulation represented magnitudes or sizes. As a result, the numerical solution to such equations could only be interpreted as positive values. If a line was 1m long then, by its nature, the number used to represent the length of that line had to be positive. In fact, the adjective "positive" was irrelevant since magnitudes could not be otherwise. To paraphrase Carl Boyer ([11]) algebra, like geometry, is simply supposed to be a description of magnitude.

For example, if a geometric problem produced the quadratic equation $x^2 + x - 2$, solving this for x would mean solving $x^2 + x - 2 = 0$. This can be solved as $x^2 = 2 - x$, being represented geometrically by the diagram below. Two answers are clearly indicated.



But the solutions to the quadratic are $x = 1$ and $x = -2$. So we now have a problem. The geometry above suggests two (positive) answers, yet the algebra gives on positive answer and one negative answer. Negative answers are considered invalid, so they are rejected.

Today we know that the solution to quadratics does indeed produce two answers, either or both of which can be negative. And such a situation can indeed be represented geometrically in a correct form if we embed $x^2 = 2 - x$ within a Cartesian coordinate system, as illustrated below.



It should be noted that, contrary to popular opinion, Descartes did not invent the modern Cartesian coordinate system. He could therefore not have interpreted negative numbers as simply being numbers relative to a reference line such as the y–axis. As such, he (as well as many others) paid no attention to negative numbers arising from the solution to equations. This left the only possible answer to the problem above as $x = 1$. In fact, "At the time, negative numbers were still not universally accepted; Cardano (1501–1576) called them "numeri ficti" (fictitious numbers) and in 1637, nearly a century later, Descartes called them "faux" (false) numbers." (p289, [74]).

The perceived non-existence of negative numbers is based on the attitude that the variables of equations referred to the world of three dimension:

> "for as *positio* $[x]$ refers to a line, *quadratum* $[x^2]$ refers to a surface, and *cubum* $[x^3]$ to a solid body, it would be very foolish for us to go beyond this point. Nature does not permit it." [Katz and Hunger Parshall, p220].

With this in mind there was no way that any negative numbers arising from the manipulation of symbols could be seen as valid or real. If opening a compass related to measuring a length, and closing a compass implied no length at all, how would you close a compass in on itself to create a "negative" or "non-existent" length? Therefore, something as simple as $x = -1$ made no geometric sense at all.

This highlights the strength of belief in numbers representing magnitudes or sizes rather than just being pure numbers. So here we see the beginnings of the problem of using algebra in the

16th and 17th centuries. In producing negative numbers (and also complex numbers) it brings up what might be called extraneous solutions.

The problem of accepting such extraneous solutions is rooted in the epistemic nature of algebra. Recall that algebra, as used from the 16th to the 19th century, was seen only as a proxy for geometric manipulation. Symbols such as x, y, x^2, x^3 represented dimension such as lines, curves, areas and volumes. This from Descartes:

> "It should also be noted that all parts of a single line should always be expressed by the same number of dimensions [...]. Thus, a^3 contains as many dimensions as ab^2 or b^3, these being the components parts of the line which I have called $\sqrt[3]{a^3 - b^3 + ab^2}$." (p6, [25])

Manipulation of these symbols (according to arithmetic operations of +, –, ×, ÷, and √) represented the manipulation of lines, curves areas and volumes. Again, from Descartes:

> "Often it is not necessary thus to draw the lines on paper, but it is sufficient to designate each by a single letter. Thus, to add the lines BD and GH, I call one $a$and the other b, and write $a + b$. Then $a - b$ will indicate that b is subtracted from a; ab that a is multiplied by b; $\frac{a}{b}$ that a is divided by b; aa or a^2 that a is multiplied by itself; a^3 that this result is multiplied by a, and so on, indefinitely." (p5, [25])

However, manipulation of expressions involving arithmetic operations was never meant for its own sake. There was no such thing as a purely symbolic manipulation in an algebraic equation. Algebraic manipulation always referred to the manipulation of geometric elements so that, if algebra gave negative results (or even complex), these results were simply seen as an absurd solution.

Despite the quote above by Katz and Hunger Parshall there was a natural curiosity to wanting to study equations involving x^4, i.e. quartics. This was motivated in part by the improvements in notation for representing in the use of variables and constants in equations. Such improvements allowed people of the day to see extensions in patterns of said variables. So, if we have an equation in x or x^2 or x^3 why can't we have an equation in x^4? It was, in fact, during the 16th century that much work was done by Italian mathematicians in the study of algebra for its own sake, without reference to any geometry such equations might represent. Specifically, they studied cubic and quartic equation as algebraic equations. This period was dominated mainly by Italians such as del Ferro (1465-1526), Cardano (1501-1576), Fontana (a.k.a

Tartaglia, 1500-1557), and Ferrari (1522-1565). This is where the break from geometry starts to appear and where algebra starts to become more of an independent subject. For example, Cardano in his book *Ars Magna*, poses the following problem:

> "Find a number the fourth power of which plus four times itself plus 8 is equal to ten times its square" [Katz and Hunger Parshall, p223]

i.e. solve $x^4 + 4x + 8 = 10x^2$. This problem had to be solved in purely algebraic terms since it was not possible to construct a curve in four-dimensional space. Except for cubics, quartics, quintics and sextics which could be reduced to combinations of straight lines (linear equations), or circles or parabolas (quadratic equations), there was no possiblity of constructing curves of degree higher than 3 (Descartes himself had managed to geometrically solve an equation of degree 6 using combinations of lines, circles and parabolas (p221-222, [25])).

In solving the aforementioned quartic Cardano was able to obtain the positive roots of $\sqrt{3} + 1$ and $\sqrt{5} - 1$ but ignored the negative roots of $1 - \sqrt{3}$ and $-1 - \sqrt{5}$, saying:

> "if the square of a square is equal to a number and a square, there is always one *true* solution and another *fictitious* solution equal to it. Thus, for $x^4 = 2x^2 + 8$, then x equals 2 or –2." [Katz and Hunger Parshall, p225]

However, it seemed to some, such as Descartes, that the algebra practised by the Italians and by Viète was becoming more and more divorced from geometry, and more and more something which focused only on the manipulation of symbols, producing whatever results obtainable therefrom.

Strangely, it would be the negative numbers which would be the last type of numbers to be fully accepted. Even the complex numbers had been accepted by the early 1800s, and this was only due to the fact that mathematicians such as Wessel and Argand had found a way of representing complex numbers geometrically, along with ways of doing arithmetic on these geometric numbers. Still, at this stage, geometry held sway over what constituted valid and proper mathematics. Only a structured way of defining numbers and arithmetic would allow things like 3, ±3, 3.1, 3.141592654… and $3 + i$ to all be seen as numbers, rather than as irrational, deaf (surd), impossible, imaginary or other.

So, as Martinez (2012) says:

> "So there was, since antiquity, a tension between those who thought that the most fundamental mathematical entities are numbers and those who preferred geometrical figures. Through the centuries, these viewpoints recurred, as some mathematicians sought to explain numerical notions geometrically, while others tried to arithmetise geometry."

Despite such issues, Descartes, as well as Viète and Fermat, had shown algebra to be incredibly useful in solving geometric problems. They were effectively responsible for introducing algebra as a mainstream form of analysis when solving geometric problems.

7.12 Are all numbers constructable?

From what we have seen so far it is clear that during the renaissance period only whole numbers and square roots were constructable and therefore considered valid numbers. A natural question to ask was, Were all numbers involving radicals constructable? Were cube roots constructable? Beyond this, was π constructable?

In general, we now know the answer to be no. For, although $\sqrt[3]{8}$ is constructable because it equals 2, and $\sqrt[3]{7+5\sqrt{2}}$ is constructable because this can be simplified to $1+\sqrt{2}$, something as seemingly simple as $\sqrt[3]{2}$ is not constructable. The reason is that constructable numbers represent constructable points, and these points can only be constructed by the intersections of lines and/or circles. Given that a line is expressed as $y = mx + c$ and a circle is expressed as $x^2 + y^2 = r^2$ such intersections simplify to a general quadratic $ax^2 + bx + c = 0$, where a, b, c are combinations of m, c, r. Therefore, only polynomials of degree at most 2 arise from the use of circles and lines. Algebraically speaking, $\sqrt[3]{2}$ is the solution to $x^3 - 2 = 0$ (or some other cubic which would give $\sqrt[3]{2}$ as an answer), so $\sqrt[3]{2}$ cannot be constructed geometrically. On the other hand $x - \pi = 0$ is a linear an equation whose algebraic solution is $x = \pi$. However, this cannot be constructed due to the fact (not known until the late 1800s) that π is transcendental. This simply means that it cannot be obtained by solving a polynomial with rational coefficients. And all constructable numbers come only from equations with rational coefficients (see sec 2, p127 onwards of [20] for a detailed discussion on this subject). It wasn't until the 1800s that a sufficient enough shift in attitude towards algebra had occurred so as to trust algebraic solutions independently of their geometric constructability. It was then in 1827 that Pierre Wantzel showed, by algebraic theory, that $\sqrt[3]{2}$ and π were not constructable.

What about fourth-roots? These are always constructable since these are simply square roots of square roots, e.g. $\sqrt[4]{1+\sqrt{2}} = \sqrt{\sqrt{1+\sqrt{2}}}$. In fact, all n^{th} roots, where n is a power of 2, are constructable. This is one of the key criteria in the modern proof of the constructability of roots. With this in mind we then see that fifth-roots, sixth-roots and seventh-roots are not generally constructable, although there are exceptions, for example $\sqrt[5]{41+29\sqrt{2}} = 1+\sqrt{2}$.

The non-constructability of $\sqrt[3]{2}$ and π might not have been such an issue if it weren't for the fact that they appeared as part of problems associated with such basic geometric objects as cubes and circles. Two of these problems were squaring the circle and doubling the cube:

- Given a circle, is it possible to construct a square of the same area? If we have a circle of radius 1 unit, therefore having area π units, can we use a straight edge and compass to construct a square of sides $\sqrt{\pi}$?
- Given a cube of a certain volume can we construct a cube of double the volume? In other words if we have a unit cube, whose volume is $x^3 = 1$, is it possible to construct a cube such that $y^3 = 2$? Effectively, this would mean constructing a cube of sides $\sqrt[3]{2}$.

These problems are ancient and classic problems in constructability. They were important problems in mathematics right up to the 19^{th} century precisely because they involved the most simple of plane and solid objects, were easily stated, but for which no solution had up to 1827.

Given that doubling the cube and squaring the circle had indeed been solved by the ancients using, for example, conic sections, or carpenter's squares, lunes (crescent shaped curves), the quadratrix or the Archimedean spiral, why did these problems persist as problems for so long? Because of the attitude the early ancient mathematicians had towards the nature of geometrical objects. They believed in the so-called Platonic ideal of geometrical objects, where (as mentioned in sec 3.2) there existed a world of perfect lines, perfect circles, perfect triangles, etc. In constructing these objects on paper these shapes were being transferred from the perfect world of ideal forms to the imperfect outside world. In the case of a line, the ideal form of a perfect line is one which has absolutely no kink, no curvature, and absolutely no change in direction. This compares with the imperfect line drawn on paper which is a corrupted version of the ideal line because it is subject to all sorts of unevenness due to the physicality of the paper, the nib of the pencil or pen, or other factors. The same can be said of the perfect circle in the world of ideal forms, and the imperfect circle drawn by a compass.

So in order to minimise imperfections, and minimise the move away from exactness, one could only use instruments having a minimum number of mechanisms. These were/are the straight edge and the compass. Even a ruler was not allowed since graduation marks on the straight edge would add an element of imperfection and approximation to construction and measurement. However, the devices used to draw conic sections, lunes, the quadratrix or the Archimedean spiral were made of multiple straight edges and hinges, of were made of moving pieces, thus corrupting the ideal forms further (having multiple hinges and straight edges, or introducing moving elements to a device, leads to accumulation of error caused by individual hinges, straight edges, or motion).

Such a philosophy of mathematics carried over from ancient times to the early modern times so that by the time of Descartes a problem could only be considered solved if it could be constructed using a straight edge and compass. This was the only rigorous way to prove a solution valid. We had to wait until the 19th century for a proof of the non-constructability of these problems. By that time there had been a sufficient enough change towards algebra in the minds of mathematicians as to trust algebraic solutions independently of their geometric constructability.

But at least the irrational number $\sqrt[3]{2}$ can be written as a polynomial with rational coefficients, i.e. $x^3 - 2 = 0$, and therefore has an exact algebraic solution even though this solution is not constructable. However, the irrational number π cannot even be found as the solution to a polynomial with rational coefficients, hence the appellation *transcendental number*, since π transcends the capabilities of finite algebraic operation (addition, subtraction, multiplication, division, and root extraction) on integers to produce such a result. If polynomials are to be used to express the number π then such polynomials are in the form of infinite series. Notice that, for the requirement of constructability, we again return to numbers needing to be whole numbers.

Other examples of transcendental numbers include $\ln 2$, $\sin 1$, e^{π}, i^{i}, and $2^{\sqrt{2}}$. And still today work is done to show whether or not a number is transcendental. From example it is still not known whether the Euler constant 0.5772156649… which is the limiting difference between the harmonic series and the natural log, i.e.

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right),$$

Is transcendental. Furthermore, even though it is known that π and e are transcendental it is not known if $e+\pi$ or πe are transcendental. It might be thought that they should be but one has to be careful about the effects of arithmetic on individual numbers. For example, we know that $\sqrt{2}$ is irrational, but is $(3+\sqrt{2})(3-\sqrt{2})$ irrational? No, since $(3+\sqrt{2})(3-\sqrt{2}) = 9 + 3\sqrt{2} - 3\sqrt{2} - 4 = 5$ is not irrational. And, just as the ancient Greeks placed restrictions on the type of instruments that could be used in geometry in order to construct certain types of numbers, so we today we place restrictions on the type of algebraic expressions which construct certain types of numbers.

7.13 A first step towards pure number?

Given the type of numbers discussed above as being constructable, we arrive at a logical conclusion: since numbers and numerical arithmetic are simply proxies for points and geometric arithmetic, why not simply perform algebra on an algebraic equation designed to model a geometry and automatically accept the numerical results without having to confirm this via the relevant geometric construction? For example, since 1 and $\sqrt{3}/2$ can be constructed geometrically, $1+\sqrt{3}/2$ can also be constructed. In this case we can simply leave the answer as $1+\sqrt{3}/2$ without effecting the construction. Similarly, the sum, difference, product or quotient of other constructable numbers will also be constructable so that we might as well just accept the numerical form of these sums, differences, products and quotients as they are.

Newly constructed numbers such as $1+\sqrt{3}/2$ can then be used to construct even more complicated numbers such as

$$1-\sqrt{1+\frac{\sqrt{3}}{2}}.$$

A multitude of constructable numbers can then be formed, so that there seems to be no end of these. For example,

$$3+\sqrt{\frac{11}{3}-\sqrt{2+\sqrt{3}}}+\sqrt{\frac{15}{4}}+\frac{8-\sqrt{7}}{5+\sqrt{1+\sqrt{3}}}.$$

If necessary, such a number could be constructed. But given that it would probably be quite complicated to do so, why bother?

The ease with which we can construct new numbers from old (already constructed) numbers suggests that we can uses numbers as objects in their own right in the construction of other new numbers without having to constantly return to first principles (i.e. geometry) to justify the existence of such numbers.

Although Descartes did not think this way we can say that, in constructing the arithmetic of line segments, he paved the way for arithmetic on numbers. Although this latter arithmetic arose from geometric construction, such an arithmetic can now be performed *without reference to geometry* and can therefore be made independent of geometry. We are now in the realm of pure number.

8 On numeric constructibility (to come)

Despite the dominance and ubiquity of geometric solutions in the 16th and 17th centuries there were a few mathematicians who studied and worked principally with pure numbers. Some of these include Diophantus (200AD – 284AD), Stifle (1487 – 1567), and Stevin (1548 – 1620). An exemplar of Diophantus' work has already been discussed (sec 4.4). Exemplars of the work of Stifle and Stevin will be discussed below, and exemplars of the work of Wallis (1616 – 1703) and Leibnix (1646 – 1716) will be discussed in part II of this essay. And, pprobably the most famous example of the study of pure number is that of Fermat and his last theorem: Given whole numbers x, y, z, the equation $x^n + y^n = z^n$ has no non-trivial solution for $n > 2$.

8.1 The case of Stifle

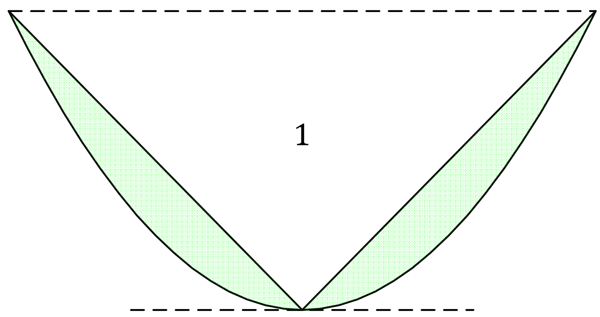
8.2 The case of Stevin

9 Appendix 1: Two examples of Archimedes using the method of exhaustion

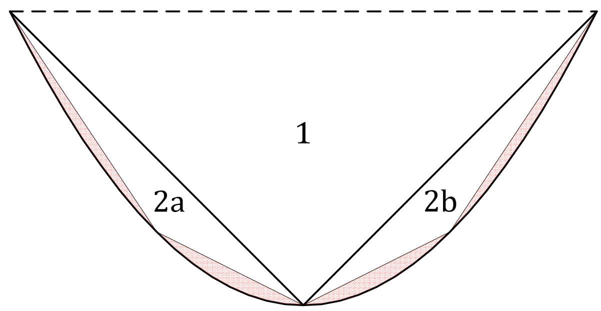
9.1 Archimedes' approach to finding the area contained by a parabola

In this section we will go through another example of how infinities were avoided in ancient Greek times by looking at how Archimedes found the area contained between a parabolic curve and a chord. Here Archimedes used triangles as his principle object by which to analyse the geometric configuration and obtain a result. In this case he actually obtains an exact answer for the area under a parabola (not an approximation as he did for π).

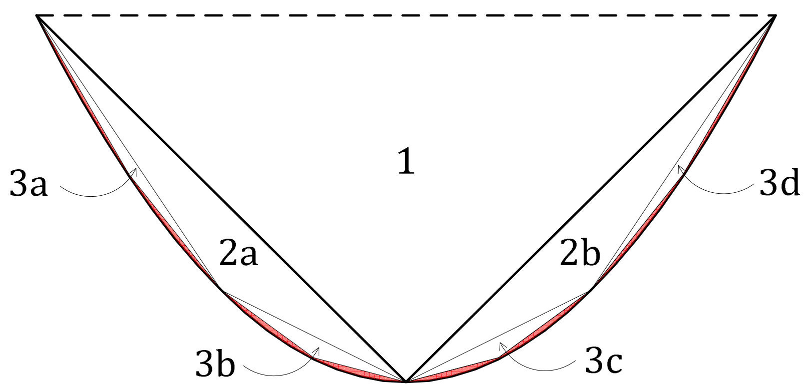
As such, let a parabola be cut by a chord parallel to the tangent at the minimum point of the parabola (diagram (i) below). We now inscribes a triangle, say T_1, whose base is this chord, and whose vertex lies on the axis of symmetry of the parabola. The remaining area shaded is the error between the triangular area and the area contained by the parabola and the chord. Our process will be to construct triangles within the successive remaining areas, firstly shown as 2a and 2b in diagram (ii), then as 3a, 3b, 3c, and 3d in diagram (iii), etc.



(i)



(ii)



(iii)

Clearly one can continue indefinitely inscribing ever smaller triangles. In terms of modern mathematics we can find the exact answer to the area under the parabola by setting up an infinite series which converges to a limit (the more usual name for this process is integration). But given Archimedes' (and the Greeks') aversion to infinities he didn't do this. instead he proceeded as follows: let T_2 be the combined areas of triangles 2a and 2b then Archimedes' first analysis showed that

$$T_2 = \frac{1}{4} T_1 .$$

Archimedes then repeated the process of triangle addition a finite number of times. Then if we let T_3 be the combined areas of triangles 3a, 3b, 3c, and 3d, Archimedes found that

$$T_3 = \frac{1}{16} T_1 = \frac{1}{4^2} T_1 .$$

Repeating this process Archimedes found $T_4 = \frac{1}{4^3} T_1$, $T_5 = \frac{1}{4^4} T_1$, etc. (see section 9.2 for the full development of this). Letting P be the area under the parabola we have, in modern form,

$$P = T_1 + \frac{1}{4} T_1 + \frac{1}{4^2} T_1 + \frac{1}{4^3} T_1 + \frac{1}{4^4} T_1 + \cdots . \qquad (*)$$

Archimedes gets around the issue of the infinite sum above as follows: he rewrites the above as the finite sequence

$$P = T_1 + \frac{1}{4} T_1 + \frac{1}{4^2} T_1 + \frac{1}{4^3} T_1 + \frac{1}{4^4} T_1 + \cdots + \frac{1}{4^n} T_1 + \frac{1}{3} . \frac{1}{4^n} T_1 .$$

He then takes only the first few terms of the series but keeps the last two terms. So, let us have

$$P = T_1 + A + B + C + \cdots + D + \frac{1}{3} D .$$

He then does the following (there is no record of how or why he knew to do this):

$$A + \frac{1}{3} A = \frac{4}{3} A = \frac{1}{3} T_1 \quad , \quad B + \frac{1}{3} B = \frac{4}{3} B = \frac{1}{3} A ,$$

$$C + \frac{1}{3} C = \frac{4}{3} C = \frac{1}{3} B , \quad , \quad D + \frac{1}{3} D = \frac{4}{3} D = \frac{1}{3} C ,$$

Then by back-substitution we have

$$\begin{aligned} T_1 + A + B + C + D + \frac{1}{3} D &= T_1 + A + B + C + \frac{1}{3} C \\ &= T_1 + A + B + \frac{1}{3} B \end{aligned}$$

$$= \quad T_1 + A + \frac{1}{3}A$$

$$= \quad T_1 + \frac{1}{3}T_1 = \frac{4}{3}T_1 \,.$$

So it seems that we have the area under the parabola to be

$$T_1 + \frac{1}{4}T_1 + \frac{1}{4^2}T_1 + \frac{1}{4^3}T_1 + \cdots + \frac{1}{4^n}T_1 + \frac{1}{3}.\frac{1}{4^n}T_1 = \frac{4}{3}T_1 \,.$$

The argument which gave us this result is certainly convincing. The effect of the back-substitution arithmetic above is to collapse the last two terms to form a version of the term before these. This arithmetic effect is repeated all the way back to $\frac{1}{3}T_1$.

But now Archimedes has to prove this result in general since the result above applies only for a limited number of terms in the series, not the infinite number of additions in the series (*) above. His proof method is what we now call *proof by contradiction*. This involves assuming the answer is not true, doing some analysis, and then arriving at a contradiction based on this analysis, implying that the answer must, in fact, be true. This is how it is done:

Assume $P < C$

Here we assume that P (our supposed exact answer) is less than C (another supposed exact answer). In that case the "true" error E is given by

$$E = C - P \,.$$

But we know that $I_n < P$. In other words, the approximate area (which is the sum of all inscribed triangular area I_n) is less than our assumed exact answer P. This means that the real error e is given by

$$e_n = P - I_n \,.$$

Notice the distinction between E and e_n: E is the "smallest" error possible whereas e_n isn't. another way of saying this is that E is the theoretical true error, e_n is the approximate error. Then, at the beginning of our process, say when we have only 1, 3, or 7 triangles inscribed, e_n will be large compared to E.

Now, since $I_n < P < C$ we have $C - P < C - I_n$, in other words

$$E < e_n \,.$$

But this is where a contradiction will appear: the error E remains fixed since it is the difference between two "final" results. But e_n is not fixed. It reduces every time we double the number of triangles. Another way of saying this is that E is the smallest possible error and e_n is a range of decreasing errors, but all larger than E.

In actuality, given an error e_n at a stage n, the error e_{2n} at the next stage of doubling the number of triangle is always less than half of e_n (see section 9.3):

$$e_{2n} < \frac{1}{2} e_n \, .$$

Since there is no stopping the process of adding more and more inscribed triangles, leading to ever better aproximations to the inscribed area, there is no stopping these errors getting smaller and smaller, i.e

$$\ldots < \frac{1}{2} e_{8n} < \frac{1}{2} e_{4n} < \frac{1}{2} e_{2n} < \frac{1}{2} e_n \, .$$

This means that sooner or later a value of e_n will catch up with, and undertake, E:

$$e_n < E \, .$$

This is the contradiction. We can't have both errors being less than each other. Hence $P \nless C$.

<u>*Assume $P > C$*</u>

Now assume that P (our supposed exact answer) is greater than C (another supposed exact answer). In that case the "true" error E is given by

$$E = P - C \, .$$

The approximate error is again given by

$$e_n = P - I_n \, .$$

But since we have $I_n < C < P$ we now have

$$P - C < P - I_n \, .$$

So again we have

$$E < e_n \, .$$

Now, the argument relating the error e_n at a stage n to the error e_{2n} at the next stage (of doubling the number of triangle) still applies, namely

$$e_{2n} < \frac{1}{2} e_n \, .$$

Hence the value of e_n will sooner or later catch up with, and undertake, E:

$$e_n < E \,,$$

resulting again in a contradiction. Hence $P \not> C$. Then, given that $P \not< C$ the only option remaining is for the area P under the polygon to be $P = C = \frac{4}{3}T_1$.

There are now two important comments to make:

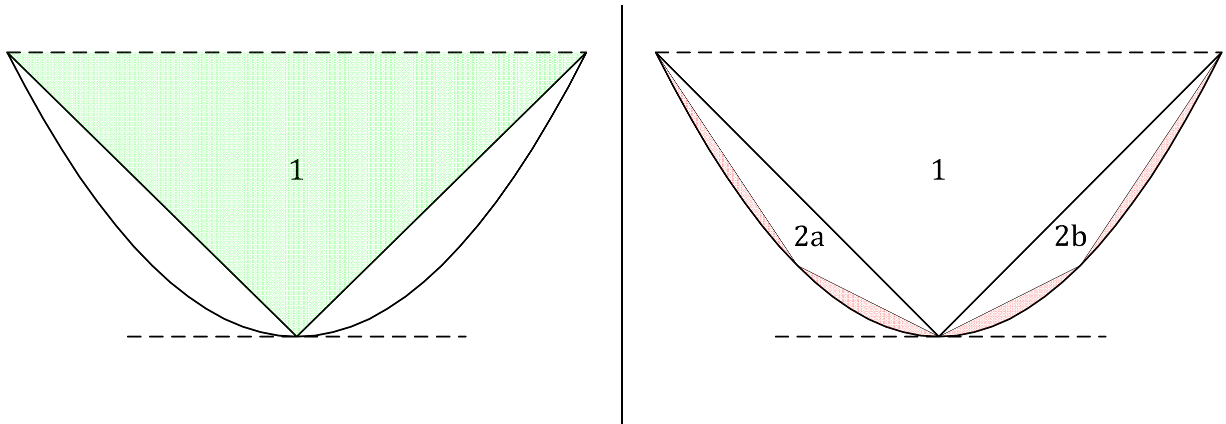
1) Looking more closely at the logic of Archimedes' proof we see that infinity has not been eliminated It has only been sidestepped. Archimedes' argument about the contradiction between the errors, and therefore between the areas P and C, are based on squeezing out the cases $P < C$ and $P > C$, leaving only $P = C$. But squeezing the error out to zero, or squeezing areas out to equality, inherently contains the idea of infinity and the infinitely small (how else would the errors reduce to zero?)

2) In both the previous section and this section we have seen that Archimedes used the so-called method of exhaustion by utilizing an ever-increasing number of triangles as his way of solving the problem of ? and the area under a parabola. But in one case he obtains an approximate answer, and the other case he obtains an exact result. And herein lies the problem: the same method (which tries to avoid infinities) does not lead to the same degree of accuracy. In the case of finding ?, avoiding infinitesimals, namely the ever-decreasing tangent line-segments, does not make these infinitesimals go away. Their effect is still directly present by the fact that we cannot obtain an exact answer in a finite number of steps. But in the case of the area under the parabola it is possible to avoid infinitesimals so as to obtain an exact answer. This is done by ignoring the infinitesimal aspect of the sums of areas of ever smaller triangle and instead using the different approach of squeezing out the only two possible alternative answers to the area.

9.2 Why $T_2 = ¼ T_1$, etc

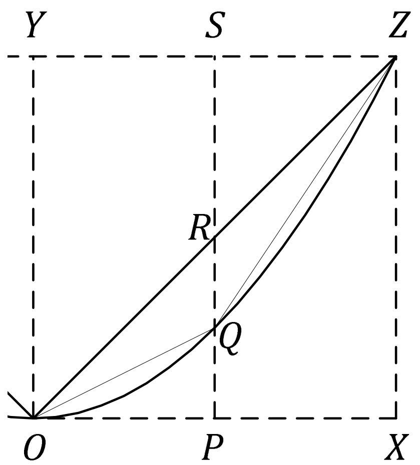
Here we see how Archimedes deduced the answer that the area under the parabola is four-thirds the area of the largest inscribed triangle. As such, consider triangle 1 shown shaded in diagram (i) below. The area of this triangle approximates the area under the parabola. Let us call this area T_1.

Now form two new triangles 2a and 2b as in diagram (ii) below.



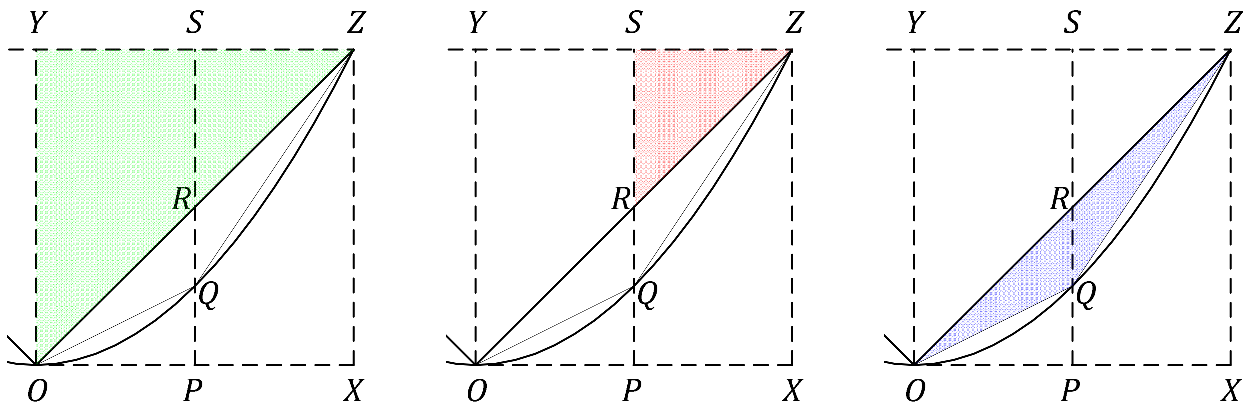
(i) (ii)

Let T_2 be the combined areas of triangles 2a and 2b. Thus $T_1 + T_2$ is a better approximation to the area under the parabola. Our aim is now to express T_2 in terms of T_1. To help us in this endeavour we focus on one half of diagram (ii). Using modern notation throughout (see [84] for Archimedes' actual approach) we draw a square $OYZX$, then bisect OX at P (diagram (iii) below). The line PS intersects the triangle and the parabola at R and Q respectively.



(iii)

The important regions to look at are triangles OYZ (which is a half of triangle 1) shown in green below, triangle RSZ shown in red below, and triangle OQZ shown in blue below. Our aim will be to find a way of linking the blue area of triangle OQZ with the green area of OYZ.



Now, notice that triangle OQZ is composed of two small triangles OQR and QRZ, both having the same base RQ and the same "height" $OP = PX$. So,

$$\text{area } \Delta OQR = \frac{1}{2}(RQ)(OP) \text{ and area } \Delta QRZ = \frac{1}{2}(RQ)(PX).$$

Since $OP = PX$ we have area $\Delta OQR =$ area $\Delta QRZ = \frac{1}{2}(RQ)(PX)$, hence we have the area of the blue shaded triangle to be

$$\text{area } \Delta OQZ = (RQ)(PX) .$$

Now we consider the red area of triangle SRZ respectively:

$$\text{area } \Delta RSZ = \frac{1}{2}(SR)(SZ).$$

Since $SZ = PX$ we have

$$\text{area } \Delta RSZ = \frac{1}{2}(SR)(PX).$$

But $SR = 2RQ$, hence

$$\text{area } \Delta RSZ = (RQ)(PX). \qquad (*)$$

Now we consider the green area of triangle OYZ respectively:

$$\text{area } \Delta OYZ = \frac{1}{2}(YZ)(OY) ,$$

and

$$\text{area } \Delta RSZ \quad = \quad \frac{1}{2}(SZ)(SR)$$

$$= \frac{1}{2}\left(\frac{1}{2}YZ\right)\left(\frac{1}{2}OY\right).$$

$$= \frac{1}{4}\Delta OYZ.$$

Since $\Delta OYZ = \frac{1}{2}T_1$ we have $\Delta RSZ = \frac{1}{8}T_1$ by (*). By symmetry across the vertical axis OY there is another triangle similar to ΔRSZ, hence we have a total area $T_2 = 2 \times \Delta RSZ$, i.e.

$$T_2 = \frac{1}{4}T_1 .$$

We now repeat the above analysis by placing a triangle in the segment bounded by arc OQ and the line OQ, as well as within the segment bounded by arc QZ and the line QZ. By symmetry across the vertical axis OY there will be two other triangles similar to these. Total area T_3 of these four smaller triangles is then be found to be

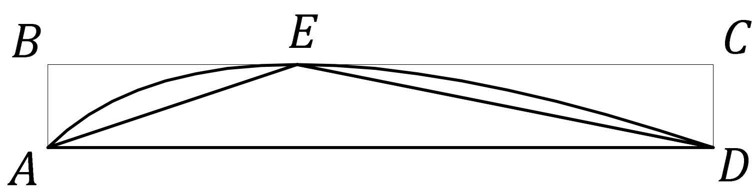
$$T_3 = \frac{1}{4}T_2 = \frac{1}{4^2}T_1 .$$

Continuing the process of constructing ever smaller triangles with the segments bounded by the relevant arcs and lines we ultimately obtain

$$T = T_1 + \frac{1}{4}T_1 + \frac{1}{4^2}T_1 + \frac{1}{4^3}T_1 + \cdots$$

*9.3 The error between the area under the triangle and the area under the parabola (*to finish*)*

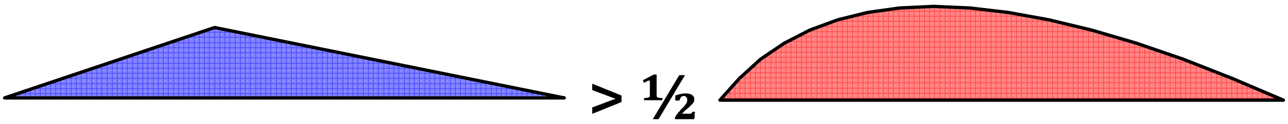
Archimedes knew that the area of any triangle ABC inscribed to a parabola ABC was greater than half the area under the parabolic segment. This is shown by considering the parallelogram $ABCD$ as illustrated below.



Here the area of the triangle is half the area of the parallelogram. Since the area under the parabola AEC is less than the area of the parallelogram $ABCD$ it follows that

$$A_{triangle} > \frac{1}{2}A_{parabolic\ segment}$$

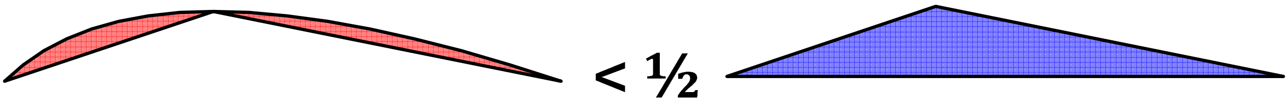
i.e.



Now, the area of the segments AE and EF represents the error between the true area under the parabola to the approximate area given by the triangle. Hence these areas in total are less than half the area of the parabolic segment.

$$A_{segment\ AE} + A_{segment\ EF} < \frac{1}{2}\, A_{parabolic\ segment}$$

i.e.



This means that the error will continue decreasing every time new triangles are added such that *(*link to the inequality $e_{2n} < \frac{1}{2} e_n$ shown in sec 9.1*)*

This is what must have suggested to Archimedes that the real answer to the area P under the parabola was in fact

$$P = \frac{4}{3} T_1 \, .$$

To include

$$\begin{aligned}
1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots + \frac{1}{4^n} + \frac{1}{3} . \frac{1}{4^n} \quad &= \quad 1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots + \frac{1}{4^{n-1}} + \left(\frac{1}{4^n} + \frac{1}{3} . \frac{1}{4^n}\right) \\
&= \quad 1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots + \left(\frac{1}{4^{n-1}} + \frac{1}{3} . \frac{1}{4^{n-1}}\right) \\
&= \quad \ldots \\
&= \quad 1 + \frac{1}{4} + \frac{1}{4^2} + \left(\frac{1}{4^3} + \frac{1}{3} . \frac{1}{4^3}\right) \\
&= \quad 1 + \frac{1}{4} + \left(\frac{1}{4^2} + \frac{1}{3} . \frac{1}{4^2}\right)
\end{aligned}$$

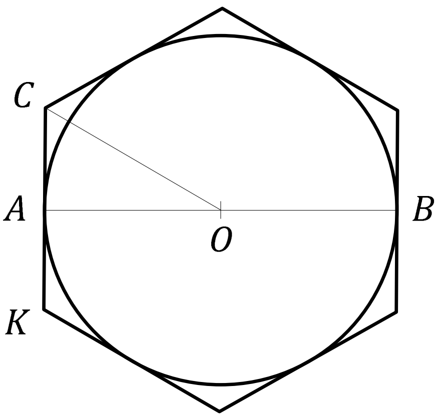
$$= \quad 1 + \left(\frac{1}{4} + \frac{1}{3}.\frac{1}{4}\right)$$

$$= \quad 1 + \frac{1}{3} = \frac{4}{3}.$$

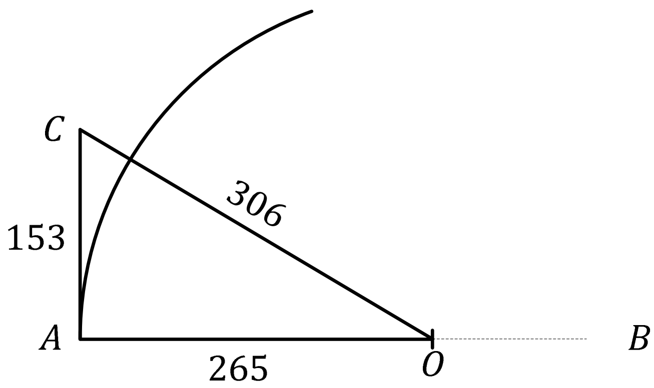
$$1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots + \frac{1}{4^{n-1}} + \frac{1}{4^n} + \frac{1}{3}.\frac{1}{4^n} = \frac{4}{3}.$$

9.4 Archimdes' upper bound for π

Having obtained a lower bound for π Archimedes repeated his polygon analysis using circumscribed polygon in order to obtain an upper bound to π. He initially started with a hexagon as illustrated in diagram (i) below.



(i)



(ii)

Archimedes repeatedly performed bisection of angle AOC as well as analysis similar to that he performed on inscribed polygons in order to develop better upper bound approximations to π.

He started by working with triangle OAC. In modern terms we would take the radius $OA = 1$ from which standard trigonometry gives the exact value $AC = 1/\sqrt{3}$. Hence, we would say that $OA : AC = 1 : \sqrt{3}$. However, not accepting roots as numbers Archimedes used 265 to 153 as his lower bound approximation to $\sqrt{3}$. Although these values are as illustrated in diagram (ii) above Archimedes knew the ratio $OA : AC$ to be greater than the ratio of these values. Hence, we have

$$OA : AC > 265 : 153$$

Since OC is a side of a hexagon and AC is half a side of a hexagon we know that

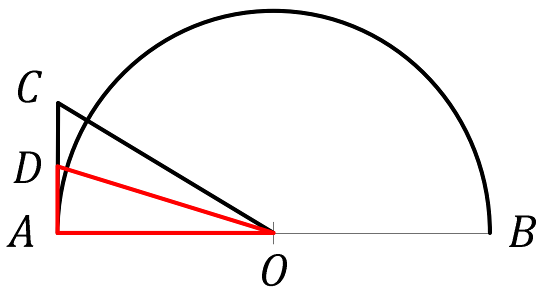
$$OC : AC :: 306 : 153 .$$

Since there are 6 sides to the hexagon, and since CA is half of a side of the hexagon we have

$$circumference : diameter < 6 \times \frac{2CA}{2OA} = 6 \times \frac{153}{265},$$

which in decimal form is 3.4641509 (to 7 d.p.). This is our initial upper bound to π.

As before, we now bisect the angle at O and produce a line from O to D as shown below. Line DA is half the side of a 12-gon, and line OD acts as an approximation to OA (half the diameter of the circle). Our aim will then be to obtain the ratio $OD : DA$, after which we can approximate π as $12 \times 2DA \div 2OD$, i.e. $12 \times DA \div OD$.



So, we start by using Euclid Book VI, proposition 3 (which gives the ratio in which D cuts AC) to obtain

$$OC : OA :: CD : DA. \qquad (*)$$

Then

$$(OC + OA) : OA :: (CD + DA) : DA \qquad (**)$$

Algebraically this is the same as adding 1 to both ratios in (*). So, in modern terms we would have $OC/OA + 1 = CD/DA + 1$ leading to $(OC + OA)/OA = (CD + DA)/DA$. Ratio (**) simplifies to

$$(OC + OA) : OA :: CA : DA,$$

from which

$$(OC + OA) : CA :: OA : DA. \qquad (***)$$

(in modern terms this is equivalent to cross-multiplying by CA and cross-dividing by OA).

So we have been able to convert the ratio we are looking for ($OA : DA$) into a ratio involving line magnitudes we already know (OA, OC, and CA). We now find the value of ratio (***), so

$$OA : DA > (306 + 265) : 153 .$$

We want $OD : DA$, so we now use Pythagoras to find this ratio:

$$OD^2 : DA^2 = (OA^2 + DA^2) : DA^2$$

$$> (571^2 + 153^2) : 153^2$$

$$> 349450 : 23409 ,$$

from which Archimedes found

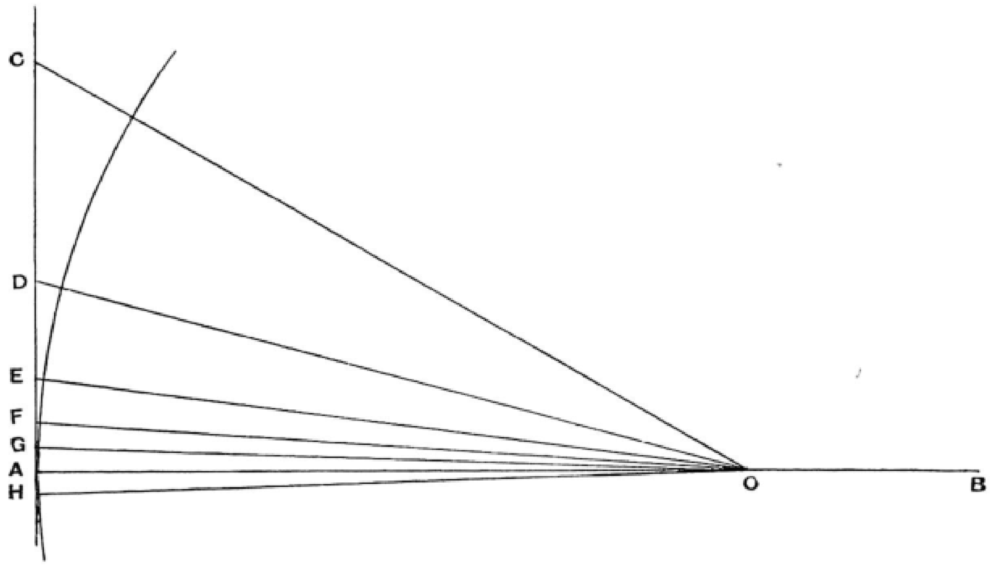
$$OD : DA > 591\tfrac{1}{8} : 153.$$

But DA is half a side of a 12-sided polygon and OD is approximately OA (half the diameter of the circle), so a better approximation to π is

$$circumference : diameter > 12 \times 2DA : 2OA :: 12 \times 153 : 571.$$

which, in our number systems, is equivalent to 3.215412 (to 6 d.p.). This is a better upper bound approximation to π.

We then proceed to continually bisect the angle at O to produce line segments OE, OF and OG, each of these being half of the line segments of 24-, 48, and 96-gons.



From this analysis we can then find new ratios $OC : CA$, $OD : DA$, $OE : EA$, $OF : FA$ and $OG : GA$. Archimedes did not calculate the ratio for $OG : GA$. He used a different method in order to obtain the perimeter-to-diameter ratio $GA : OA$. But to be consistent with what we are doing here we could have a suitable value for OG to be $OG = 4676$, with Archimedes' $OA = 4673\tfrac{1}{2}$. The table below summarises the results.

Polygon	Ratio of line-segments	Value	Approximation for π
6-sided	$OC : CA$	$306 : 153$	$6 \times CA : OA < 918 : 265$
12-sided	$OD : DA$	$591\frac{1}{8} : 153$	$12 \times DA : OA < 1836 : 571$
24-sided	$OE : EA$	$1172\frac{1}{8} : 153$	$24 \times EA : OA < 3672 : 1162\frac{1}{8}$
48-sided	$OF : FA$	$2339\frac{1}{4} : 153$	$48 \times FA : OA < 7344 : 2334\frac{1}{4}$
96-sided	$OG : GA$	$4676 : 153$	$96 \times GA : OA < 14688 : 4673\frac{1}{2}$

In decimal terms the last column of the table above is

$$3.464150943\,,\quad 3.215411559\,,\quad 3.159728945\,,\quad 3.146192567\,,\quad 3.142826575$$

These ratios represent upper bounds to the ratio of the circumference-to-diameter of a circle.

So it is that we have the two ratios ------ and $14688 : 4673\frac{1}{2}$ both derived from a 96-gon. Archimedes is able to simplify these as follows (expressed in modern terms)

……………..

and

$$\frac{14688}{4673\frac{1}{2}} = 3 + \frac{667\frac{1}{2}}{4673\frac{1}{2}} < 3 + \frac{667\frac{1}{2}}{4672\frac{1}{2}} = 3\frac{1}{7}$$

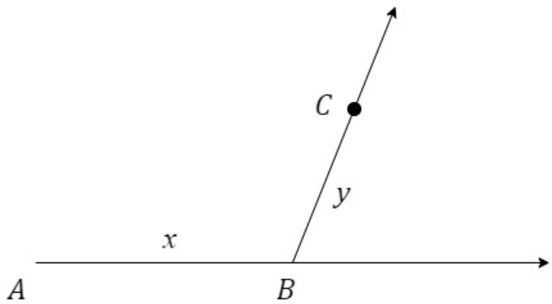
Hence

$$3\frac{10}{71} < \pi < 3\frac{1}{7}.$$

In other words, inscribed polygons will never give values greater than $3\frac{1}{7}$, and circumscribed polygons will never give values less than $3\frac{10}{71}$. This is the effort required if you only believe in integers and commensurability.

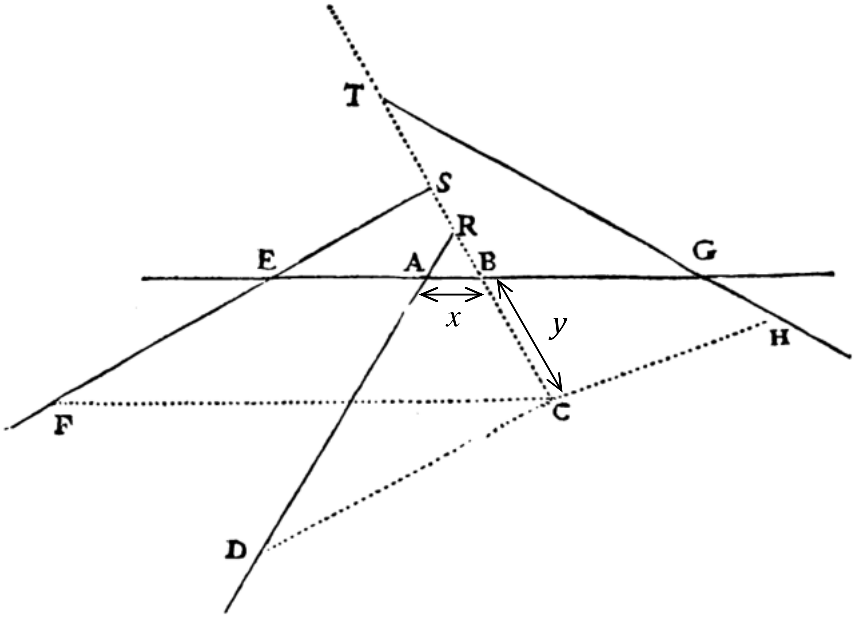
10 Appendix 2: Descartes' solution to Pappus' 4-line problem

Descartes illustrates his approach to the algebraic solution to geometry problems by solving Pappus's 4-line locus problem. A locus is essentially a path traced out by a point as this moves under a certain constraint. For example, if we have the geometric configuration below, where A is the starting point of our measurements (i.e. our origin) and the line through AB is our independent reference axis (i.e. the *x*-axis), then we want to find the path traced out by point C under a specific constraint.



If the constraint on C is that it always remains the same distance from B then the path traced by C is a circle. If the condition on C is that the line drawn from it to B always remains at the same angle, then the path traced by C is a straight line.

In solving Pappus' 4-line locus problem Descartes sets up the diagram below. Consider any four lines EG, FS, DR and TH (represented as solid lines), not all of them colinear. Now place a point C anywhere in the plane. The aim is to find the path traced out by C if the ratio of two line segments, say CD : CF, always equals the ratio of the other two line segments, in this case CB : CH (where these line segments are shown as dotted) (note that any pair of line segments could be chosen to form the ratios).



(diagram taken from p27, [25] where I have added the x and y letters)

We therefore want

$$CD : CF = CB : CH. \qquad (*)$$

Another way of saying this is that we want the ratio of these two ratios to be constant, i.e. $(CD/CF)/(CB/CH) = 1$. Descartes then decides to measure everything with respect to one fixed line segment, in this case line EG (hence we might call line EG the x-axis), and he starts measuring from point A (hence we might call point A the origin). He then sets $AB = x$ and $CB = y$ which is the distance we are looking for. This approach helps simplify a problem down to two common variables, one known and one unknown. By the use of algebra he develops equations for CD, CF, CB, and CH in terms of the common variables x and y, and finally obtains an equation which represents the path traced out by C.

Note that it looks like Descartes is setting up his problem in a Cartesian coordinate system, but this is not the case. Firstly, he does not refer to A, B, C, D, … as numbers, just as points. Secondly, he is not measuring distance y on a fixed vertical axis. The fact that y here is at an angle to the vertical doesn't matter. This would just be an example of an oblique coordinate system. But here y does not represent a reference axis but simply a length. Since point C moves around the plane, so y changes length (which is not what the modern-day y-axis does). Furthermore, from ancient time to Descartes' time (and some time beyond) numbers were used to represent points. But in the modern-day Cartesian coordinate system points are used to represent numbers, specifically the numerical pair (x, y).

So it is that Descartes did not invent the Cartesian coordinate system as we know it today, i.e. as being an ordered pair (x, y) located in an orthogonal reference frame. But he did develop the approach of setting up equations involving algebraic symbols x and y referenced to some line (line EG in the diagram above). The Cartesian system then developed over a period of time on the basis of Descartes' work.

To return to our problem, Descartes sets up equations relating to CD, CF, CB and CH combining these (by substitution) as he goes along developing the next equation until he obtains one single equation in x and y. This equation then represents the curve generated as C moves so as to maintain the equality of the two ratios *(*repetition with 2 //s above*)*. As an example, the equation representing line CB is derived as follows:

- Triangle ARB is determined: All angle here are known because $\measuredangle RAB$ is known by construction. Also, for any given point C (determined so as to keep the equality of the ratios in (*)) the line from C through B forms a determined angle RBA. Hence $\measuredangle ARB$ is also known;
- This means that every side of triangle ARB is a fixed multiple of every other sides of triangle ARB. So even though the triangle may be smaller or larger than shown above its internal angles will remain the same. As such the ratio of sides will remain the same;
- So the ratio of AB to BR is known (because of the sine rule). In this case Descartes writes $\mathrm{AB} : \mathrm{BR} = z : b$ where z and b are unknown but determined values;
- Hence $\mathrm{AB} = \frac{b}{z}.x$ (note that throughout his analysis Descartes mixes the use of ratio notation, ":", and equality notation, "=")
- Then $\mathrm{CB} = y + \frac{b}{z}.x$. This applies only when point C lies outside of BR. Descartes separately deals with the case when C lies within BR and when C lies on the other side of BR (see p29 of [25]).
- We now have an equation for the unknown length of one of the line segments of (*). We simply need to find equations for the other line segments. For example, starting with $CR : CD = z : c$ Descartes finds that $CD = \frac{cy}{z} + \frac{cb}{z^2}x$ (although there seems to be a typo in his equation of the last line of p29 of [25]). Equations for CF and CH are similarly found.
- With appropriate algebra (namely, substitution) Descartes obtains the final equation of y in terms of x to be

$$y = m - \frac{n}{z}x + \sqrt{m^2 + ox + \frac{p}{m}x^2} \qquad (**)$$

with m, n, o and p as constants representing combinations of constants b, c, z, and other constants used in Descartes' analysis. Equation (**) can be simplified to

$$y^2 + ax^2 + by + cx + dyx + e = 0$$

where constant a, b, c, d, e are relevant combinations of the constants in (**). This equation is the standard form of a general parabola so that, given the constraint that CD：CF to CB：CH remains constant, (**) the path of C is a parabola.

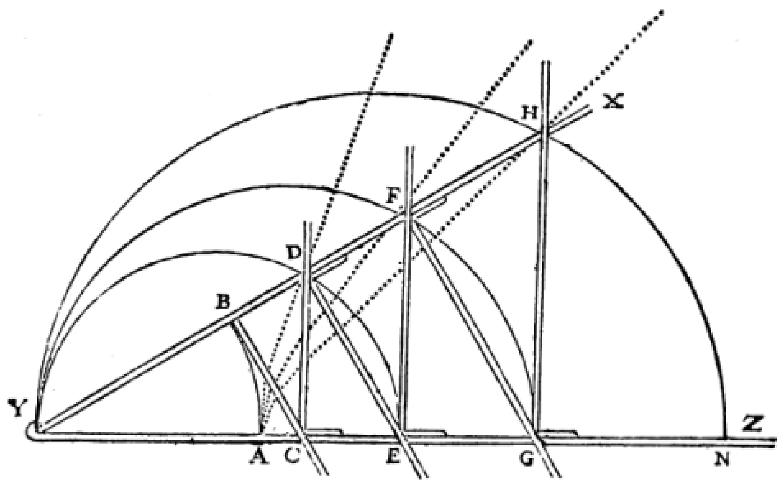
Imagine trying to perform all the steps above geometrically! So, arithmetic on line segments is dealt with much more efficiently and powerfully by Descartes' use of symbolism and algebra. This allows for a more powerful way of solving geometric problems.

However, all this algebraic manipulation, which acts in lieu of the geometric manipulations of lines and curves via straight-edge and compass, represents a deconstruction of the original geometric figure. How is one to know that the final equation resulting from all this algebraic deconstruction represents anything constructable? Such an equation would only be constructable if each step in the algebra were constructable? However, just because algebra allows us to arrive at certain equations doesn't mean to say that each step in the algebra is constructable, and therefore that the final equation is constructable.

And we have seen that a set of linear equations, representing a constructable geometry (the four-line geometric configuration of the diagram above) produces an equation, a parabola, which is not constructable in the classical sense. But Descartes did not accept such a perspective. Given that he had been able to recover the solutions to Pappus' original 3-line and 4-line problems using algebra he had no problems accepting that his algebraic solutions to 5-line, or many-more-line, problems could also be considered to represent geometric curves and therefore be constructable.

11 Appendix 3: Use of Descartes' mesolab for constructing $\sqrt[3]{2}$ and a cubic equation

Consider Descartes' mesolab, reproduced below. The configuration of rods sets up similar triangles.



So, if we take triangles *YBC*, *YDC* and *YDE* then we can set up the following ratio of lines:

$$BC : CY :: CY : DY :: DY : EY \quad \text{(a)}$$

(think of *BC* as the "short perpendicular" and *CY* as the hypotenuse. Then we effectively have a number of short perpendiculars proportional to hypotenuses). In modern notation we write

$$\frac{BC}{CY} = \frac{CY}{DY} = \frac{DY}{EY}. \quad \text{(b)}$$

This gives i) $BC.DY = (CY)^2$ and ii) $CY.EY = (DY)^2$. Substituting i) into ii) for *DY*, and simplifying, we obtain

$$CY = \sqrt[3]{(EY)(BC)^2}.$$

Since $EY = 2AY$ we end up with

$$CY = \sqrt[3]{(2AY)(BC)^2}. \quad \text{(c)}$$

Since Descartes took YB as his unit line, we have $YB = YA = 1$, hence

$$CY = \sqrt[3]{2}.\sqrt[3]{(BC)^2}. \quad \text{(d)}$$

On the other hand, if we rotate *YX* so that *BC* = 1 we can find that

$$CY = \sqrt[3]{2}.AY \quad \text{(e)}$$

This is obtained by setting any ratio in (b) equal to k. Since all ratios are equal we have

$$\frac{BC}{CY} = \frac{CY}{DY} = \frac{DY}{EY} = k\,.$$

Multiply all three fractions to get $BC/EY = k^3$. Using $k = BC/CY$ this simplifies to $BC/EY = (BC/CY)^3$, leading to

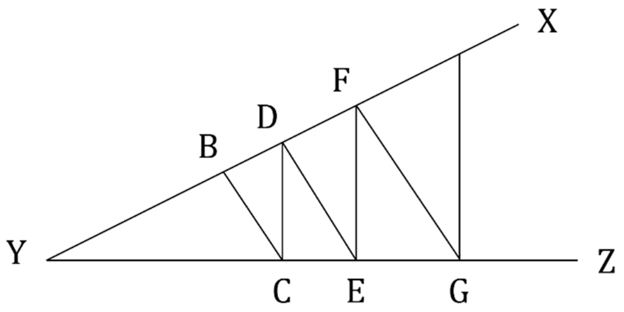
$$(CY)^3 = (EY)^3(BC)^2.$$

Since $EY = 2AY$ we have $(CY)^3 = (2AY)^3(BC)^2$. Given $BC = 1$ we obtain the desired result (e). If we then set AY as our unit length then

$$CY = \sqrt[3]{2}\,.$$

In any case, it is the comparison of the three ratios in (a) which allows for the construction of the cube root.

We can also use the mesolab to set up, and solve, a cubic equation. To see this let us look at a simplified diagram of the mesolab above:



Choose BY as our unit line, and set the mesolab so that $CE = 2$. Then by similar triangles we have

$$BY : CY :: CY : DY :: DY : EY\,.$$

Using modern notation we now consider two separate equality between the ratios:

(a) $\dfrac{BY}{CY} = \dfrac{CY}{DY}$ and (b) $\dfrac{CY}{DY} = \dfrac{DY}{EY}$

From (a) we have $(BY)(DY) = (CY)^2$ and from (b) we have $(CY)(EY) = (DY)^2$. Substitute the former equation into the latter equation for DY to obtain

$$(CY)(EY) = \frac{(CY)^4}{(BY)^2}\,.$$

But $BY = 1$, so the above simplifies to $EY = (CY)^3$. Let $CY = x$. Then $EY = CY + EC = x + 2$. Hence we have

$$x^3 = x + 2.$$

In fact we can set CE to any value a, so that $x^3 = x + a$. Note that, despite the fact that we have constructed an equation for the given geometry, and that we can solve this equation to find a numerical answer, the solution (the line segment CY) and its construction are still geometric.

12 Appendix 4: Dirichlet's proof that *ab* = *ba*

The following comes from p1 of John Stillwell's (1999) translation of P. Dirichlet *Lectures on Number Theory*.

CHAPTER 1

On the divisibility of numbers

§1. The product of two or more factors

In this section we treat a few arithmetic theorems, which indeed may be found in most text books, but which are of such fundamental importance for our science that a rigorous proof is necessary. First among them is the theorem that the product of a series of positive integers is independent of the order in which they are multiplied. Confining ourselves initially to the case of *three* numbers a, b, c, we construct the following schema

$$\begin{array}{cccccc} c, & c, & c, & c, & \ldots & c \\ c, & c, & c, & c, & \ldots & c \\ c, & c, & c, & c, & \ldots & c \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ c, & c, & c, & c, & \ldots & c \end{array}$$

consisting of b horizontal rows, each containing the number c equally often, namely a times. We ask ourselves what is the sum of all these numbers. First we can say: since the number c appears a times in each horizontal row, it follows from the basic principle of multiplication that the sum of each row is ca, where c is the multiplicand and a the multiplier. Also, since there are b such horizontal rows, the sum of all the numbers is $(ca)b$, where ca is now the multiplicand and b the multiplier. But now we can determine the same sum in another way, by viewing the above schema as a vertical rows, in each of which the number c appears b times. The sum of the numbers in a vertical row is cb, and hence the total sum is $(cb)a$. From this we obtain the first result, that

$$(ca)b = (cb)a,$$

and by setting the previously arbitrary number c equal to 1 we conclude that

$$ab = ba,$$

that is, *in a product of any two positive integers the multiplier and multiplicand may be interchanged.* Thus one may drop the distinction between multiplier and multiplicand in the terminology, and refer to both by the name *factors*.

13 Bibliography

The list of titles below forms the complete bibliography for all three parts of this essay. Not all titles listed are relevant to this part of the essay.

1. Alexander, A., (2014), *Infinitesimals: How a dangerous mathematical theory shaped the modern world*, Scientific American / Farrar, Staus and Giroux.
2. Baron, M., (1969), *The Origins of Infinitesimal Calculus*, Pergamon Press
3. Berggren, L., Borwein, J., Borwein, P., (1997) *Pi: A source book*, Springer-Verlag.
4. Bingham, T. R., (1971), "Newton and the development of the calculus", *Pi Mu Epsilon Journal*, Spring 1971, Vol. 5, No. 4 (SPRING 1971), pp. 171-181.
5. Bonsangue, M., (2016), "In search of Archimedes: Measurement of a circle", *Mathematics Teacher*, Vol 110, No 1, (August 2016)
6. Bos, H. J. M., (1974), "Differentials, Higher-Order Differentials and the Derivative in the Leibnizian Calculus", *Archive for History of Exact Sciences*, 26.XI.1974, Vol. 14, No. 1 (26.XI.1974),
7. Bos, H. J. M., (1981), "On the Representation of Curves in Descartes' Géométrie", Archive for History of Exact Sciences, 1981, Vol. 24, No. 4 (1981), pp. 295-338.
8. Bos, H. J. M., (2001), *Redefining geometrical exactness: Descartes' transformation of the early modern concept of construction*, Springer.
9. Boyer, C. (1956), *The history of analytic geometry*, number 6 and 7 of the Scripta Mathematica studies, Published by Scripta Mathematica.
10. Boyer, C. (1959), *The history of calculus and its conceptual development*, Dover, New York.
11. Boyer, C. (1959), "Descartes and the geometrization of algebra", *The American Mathematical Monthly*, Vol. 66, No. 5, pp. 390-393.
12. Burgin, M. (2022), *Trilogy of Numbers and Arithmetic: Book 1: History of numbers and arithmetic*, Vol 12, World Scientific Studies in Informations Studies.
13. Burton, D. (2011), *The history of mathematics: An introduction*, (Seventh Edition), McGraw Hill.
14. Byrne, O. (1847), *The First Six Books of The Elements of Euclid in which Coloured Diagrams and Symbols Are Used Instead of Letters for the Greater Ease of Learners*, William Pickering, London.

15. Carroll, M. T. et al, (2013), “Indivisibles, Infinitesimals and a Tale of Seventeenth-Century Mathematics”, *Mathematics Magazine*, Vol 86, Issue 4, p239-254.

16. Child, J. M. (ed.), *The Early Mathematical Manuscripts of Leibniz*, translated from the Latin texts published by Carl Immanuel Gerhardt, with critical and historical notes by J. M. Child, The Open Court Publishing Co., Chicago and London, 1920, p148.

17. Choike, J. R., “The Pentagram and the Discovery of an Irrational Number”, *The Two-Year College Mathematics Journal*, Nov., 1980, Vol. 11, No. 5 (Nov., 1980), pp. 312-316

18. Cooke, R. (2005), *The History of Mathematics. A Brief Course (Second Edition)*, Wiley.

19. Coolidge, J. L., (1963), *A history of geometric methods*, Dover.

20. Courant, R., Robbins, H. (1996), *What is mathematics? An elementary approach to ideas and methods*, Oxford University Press.

21. Crippa, D. (2017), “Descartes on the Unification of Arithmetic, Algebra and Geometry Via the Theory of Proportions”, *Revista Portuguesa de Filosofia, T. 73, Fasc. 3/4, Ciências Formais e Filosofia: Lógica e Matemática / Formal Sciences and Philosophy: Logic and Mathematics* (2017), pp. 1239-1258

22. Crocker, R. L. (1963), “Pythagorean Mathematics and Music”, *The Journal of Aesthetics and Art Criticism*, Vol. 22, No. 2 (Winter, 1963), pp. 189-198.

23. Dawson, B. (2019), “0.999… = 1: An infinitesimal explanation”, Math horizons, Vol 24, No 2 (Nov 2016), pp 5-7.

24. Dennis, D. (1997), “Rene Descartes' Curve-Drawing Devices: Experiments in the Relations Between Mechanical Motion and Symbolic Language”, *Mathematics Magazine*, Jun., 1997, Vol. 70, No. 3, pp. 163-174.

25. Descartes, R. (1954). *The Geometry of René Descartes: Translated from French and Latin by David Eugene Smith and Marcia L. Latham*. New York: Dover.

26. Earman, J. (1975), “Infinities, infinitesimals, and indivisibles: The Leibnizian labyrinth”, *Studia Leibnitia*, 1975, Bd, 7, H. 2, pp236-251

27. Ebbinghaus, H. -D. et al. (1991), *Numbers*, Graduate texts in mathematics 123, Springer

28. Edwards, C.H. (1979), “The Historical Development of the Calculus”, Springer-Verlag, New York.

29. Eves, H. (1969), *An introduction to the history of mathematics (third edition)*, Holt, Rinehart, and Winston.

30. Fowler, D. H. F. (1994), "The Story of the Discovery of Incommensurability, Revisited," in *Trends in the Historiography of Science*, Boston, Gavroglu, K., Christianidis, J., Nicolaidis, E. (eds.), Kluwer.

31. Fritz, K. (1945), "The Discovery of Incommensurability by Hippasus of Metapontum", *Annals of Mathematics*, Apr., 1945, Second Series, Vol. 46, No. 2 (Apr., 1945).

32. Gonzalez, M. O. and Mancill, J. D. "On the System of Natural Numbers", *The American Mathematical Monthly*, Feb., 1950, Vol. 57, No. 2 (Feb., 1950), pp.104-112

33. Grabiner, J. (1983), "The Changing Concept of Change: The Derivative from Fermat to Weierstrass", *Mathematics Magazine*, Sep., 1983, Vol. 56, No. 4 (Sep., 1983), pp. 195-206

34. Grabiner, J. (1995), "Descartes and Problem-Solving", *Mathematic Magazine*, April 1995, vol 68, No 2, pp88-97,

35. Grattan-Guinness, I. (Ed) (1980), "From the Calculus to Set Theory 1630-1910: An Introductory History", Princeton university press.

36. Hawking, S. (2007), "God created the integers: The mathematical breakthroughs that changed history", Running press.

37. Heath, T. (1921), A history of Greek mathematics, Vol 1, Theommes Press, 1993.

38. Hight, D. W., (1977), *A concept of limits*, Dover.

39. Hofmann, J. E. (1974), *Leibniz in Paris: 1672 – 1676: His growth to mathematical maturity*, Cambridge University Press

40. Hogben, L. (1936), *Mathematics for the million*, George Allen & Unwin.

41. Hopgood, K. (date unknown), downloaded from http://jwilson.coe.uga.edu/EMAT6680Fa06/Hobgood/Pythagoras.html , August 2021.

42. Horváth, M. (1986), "On the Attempts made by Leibniz to Justify his Calculus", *Studia Leibnitiana* , 1986, Bd. 18, H. 1 (1986), pp. 60-71

43. Huffman, C. A. (Editor) (2014), *A history of Pythagoreanism*, Cambridge University Press

44. Karpinski, L. C. (1925), *The history of arithmetic*, Rand McNally & Company.

45. Katz, M. G., and Sherry, D. M., "Leibniz's Laws of Continuity and Homogeneity," *Notices of the AMS*, Volume 59, Number 11.

46. Katz, M. G., and Sherry, D. M. (2013), "Leibniz's Infinitesimals: Their Fictionality, Their Modern Implementations, and Their Foes from Berkeley to Russell and Beyond", Mikhail G. Katz and David Sherry, *Erkenntnis*, June 2013, Vol. 78, No. 3, pp. 571-625.

47. Kent, D.A., Muraki, D. J., (2016), "A Geometric Solution of a Cubic by Omar Khayyam … in Which Colored Diagrams Are Used Instead of Letters for the Greater Ease of Learners", *The American Mathematical Monthly*, Vol. 123, No. 2 (February 2016), pp. 149-160.

48. Kidron , I., and Tall, D., (2015), "The roles of visualization and symbolism in the potential and actual infinity of the limit process", *Educational Studies in Mathematics*, February 2015, Vol. 88, No. 2 (February 2015), pp. 183-199

49. Kitcher, P., (1973), "Fluxions, Limits, and Infinite Littlenesse. A Study of Newton's Presentation of the Calculus", *Isis* , Mar., 1973, Vol. 64, No. 1 (Mar., 1973), pp. 33-49

50. Kleiner, I. (2001), "History of the Infinitely Small and the Infinitely Large in Calculus", *Educational Studies in Mathematics*, 2001, Vol. 48, No. 2/3, Infinity: The Never-Ending Struggle (2001), pp. 137-174 Carnot

51. Kline, M. (1972), "Mathematical thought from ancient to modern times", Oxford University Press, New York

52. Knobloch, E., (2002), "Leibniz's rigorous foundation of infinitesimal geometry by means of Riemannian sums", 133: 59–73, 2002

53. Knorr, W. R. (1975), "The evolution of the Euclidean elements", D. Reidel Publishing

54. Knott., R. (1979), "The Development of Number Systems", *Mathematics in School*, Vol. 8, No. 4 (Sep., 1979), pp. 23-25.

55. Lenoir, T. (1979), "Descartes and the geometrization of thought: The methodological background of Descartes' Géométrie", *Historia Mathematics*, vol 6 , 355-379.

56. Lolli, G., (2012), "Infinitesimals and infinites in the history of mathematics: A brief survey", *Applied Mathematics and Computation* vol 218, p7979–7988

57. Lovitt, W. V. (1960), *Elementary theory of equations*, Prentice Hall.

58. Gardiner, A. (2002), *Understanding infinity: The mathematics of infinite processes*, Dover Publication.

59. Gouvea, Fernando Q. (2010), "From numbers to number systems – Part II: the origins of modern mathematics", pp. 77–82 of *The Princeton Companion to Mathematics*

60. Janke, Hans N. (Editor) (2003), *A history of analysis*, History of mathematics, Volume 24, American Mathematical Society,

61. Katz, K. U. and Katz, M, "When is 0.999… less than 1", *The Montana Mathematics Enthusiast*, Vol 7, No 1.

62. Katz, M., Tall, D., "The tension between intuitive infinitessimals and formal mathematical analysis", Downloaded from ResearchGate.

63. Katz, V. J., Hunger Parshall. K., (2014), *Taming the Unknown: A History of Algebra from Antiquity to the Early Twentieth Century*, Princeton university press

64. Knott, R. (1979), "The Development of Number Systems", *Mathematics in School* , Sep., 1979, Vol. 8, No. 4 (Sep., 1979), pp. 23-25.

65. Malet, A, (1996), "From indivisibles to infnitesimals: Studies on seventeenth century mathematizations of infinitely small quantities", Monografies 6, Enrahonar.

66. Malik, S. C. and Arora, S. (1982), *Mathematical Analysis (4th Ed.)*, New Age International Publishers.

67. Martinez, Alberto A. (2012), *The Cult of Pythagoras: Math and Myths*, University of Pittsburgh Press.

68. Merzbach, U. C., and Boyer, C. B. (2011), *A history of mathematics (3rd Ed.)*, Wiley & Sons.

69. Moore, A. W. (1990), *The infinite*, Routledge.

70. Nickalls, R. W. D., (2006), "Viète, Descartes and the cubic equation", *The Mathematical Gazette* 2006; 90 (July; No. 518), 203–208.

71. Nillsen, R. (2021), "Infinitesimal knowledges", *Axiomathes*, published online 21 March 2021.

72. Pappus (1986), *Pappus of Alexandria – Book 7th of the Collection*, trans. Alexander Jones, Dordrecht Heidelberg, New York London: Springer, 1986

73. Ramati, A., (2001), "The Hidden Truth of Creation: Newton's Method of Fluxions", *The British Journal for the History of Science*, Dec., 2001, Vol. 34, No. 4 (Dec., 2001), pp. 417-438

74. Richeson, D. S. (2019), *Tales of Impossibility: The 2000-Year Quest to Solve the Mathematical Problems of Antiquity*, Princeton University Press

75. Rucker, R. (2019), *Infinity and the Mind: The Science and Philosophy of the Infinite*, Princeton University Press

76. Sagal, Paul T. (1973), "How many numbers are there?", *Philosophia Mathematica*, Volume s1-10, issue 2, Winter 1973.

77. Saito, K., (2013), "Archimedes and double contradiction proof", *Lettera Matematica* (2013) 1(3):p97–104.

78. Sardelis, D. A., Valahas, T. M. (1998), *Elements of Pythagorean Arithmetic*, Experimental mathematics series, Issue 2, The American College of Greece.

79. Seig, W. (2013), *Hilbert's program and beyond*, Oxford University Press.

80. Schrader, D. V. (1968), "DE ARITHMETICA, Book I, of Boethius", *The Mathematics Teacher*, Vol. 61, No. 6, pp. 615–628.

81. Sherry, D., and Katz, M. (2012), "Infinitesimals, imaginaries, ideals, and fictions", *Studia Leibnitiana*, 2012, Bd 44, H. 2(2012), p166-192.

82. Silver, Kenneth (2017), "Pythagorean philosophy", in *Alexandria and Qumran: Back to the beginning*, Archeopress.

83. Spivak, M. (1994), *Calculus, 3rd edition*, Cambridge University Press.

84. Stein, S., (1999), *Archimedes: What did he do besides cry Eureka?*, The Mathematical Association of America.

85. Stillwell, J., (2002), *Mathematics and its history (second edition)*, Springer

86. Stillwell, J., (2013), *The real numbers – An introduction to set theory and analysis*, Springer

87. Stillwell, J., (2016), *Elements of Mathematics: From Euclid to Gödel*, Princeton University Press.

88. Tattersall, J. J. (1999), *Elementary number theory in nine chapters*, Cambridge University Press.

89. Taylor, T. (1816), *Theoretic arithmetic*, , A. J. Valpy, Chancery Lane

90. Waismann, F. (1951), *Introduction to mathematical thinking*, Frederick Ungar Publishing.

91. Windred, G. (1929), "History of the Theory of Imaginary and Complex Quantities Author", *The Mathematical Gazette* , Oct., 1929, Vol. 14, No. 203 (Oct., 1929), pp. 533-541.

92. Wren, F. L., Garrett, J. A., (1933), "The Development of the Fundamental Concepts of Infinitesimal Analysis", *The American Mathematical Monthly*, May, 1933, Vol. 40, No. 5 (May, 1933), pp. 269-281