

Overview (Divine Proportions)

1.1 Introducing quadrance and spread

Trigonometry is the *measurement of triangles*. In classical trigonometry, measurement uses *distance* and *angle*, while in rational trigonometry measurement uses *quadrance* and *spread*. To appreciate the difference, let's consider a specific triangle in the decimal number plane from both points of view.

Classical measurements

The triangle with side lengths

$$d_1 = 4 \quad d_2 = 7 \quad d_3 = 5$$

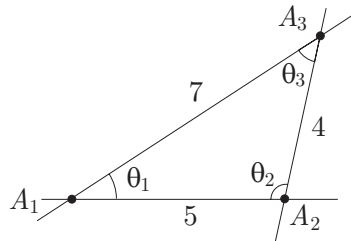
has respective angles (approximately)

$$\theta_1 \approx 33.92^\circ \quad \theta_2 \approx 102.44^\circ \quad \theta_3 \approx 43.64^\circ.$$

There are numerous classical relations between these six quantities, such as the Sums of angles law

$$\theta_1 + \theta_2 + \theta_3 = 180^\circ,$$

along with the Cosine law, the Sine law, and others. Typically these laws involve the trigonometric functions $\cos \theta$, $\sin \theta$ and $\tan \theta$, and implicitly their inverse functions $\arccos x$, $\arcsin x$ and $\arctan x$, all of which are difficult to define precisely without calculus.



Rational measurements

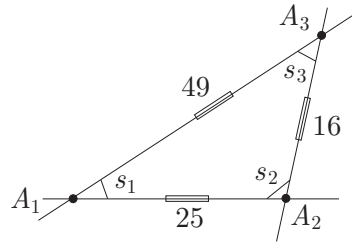
In rational trigonometry the most important measurements associated with the same triangle are the *quadrances* of the sides,

$$Q_1 = 16 \quad Q_2 = 49 \quad Q_3 = 25$$

and the *spreads* at the vertices

$$s_1 = 384/1225 \quad s_2 = 24/25 \quad s_3 = 24/49.$$

Let's informally explain the meaning of these terms, and show how these numbers were obtained.



Quadrance

Quadrance measures the separation of two points. The easiest definition is that *quadrance is distance squared*. Of course this assumes that you already know what distance is. A point in the decimal number plane can be specified by its x and y coordinates with respect to a fixed pair of rectangular axes, and the usual definition of the distance $|A_1A_2|$ between the points $A_1 \equiv [x_1, y_1]$ and $A_2 \equiv [x_2, y_2]$ is

$$|A_1A_2| \equiv \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

So the **quadrance** $Q(A_1, A_2)$ between the points is

$$Q(A_1, A_2) \equiv (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

From this point of view, quadrance is the more fundamental quantity, since it does not involve the square root function. The relationship between the two notions is perhaps more accurately described by the statement that *distance is the square root of quadrance*.

In diagrams, small rectangles along the sides of a triangle indicate that quadrance, not distance, is being measured, a convention maintained throughout the book. Occasionally when this is inconvenient, a quadrance is enclosed in a rectangular box.

Spread

Spread measures the separation of two lines. This turns out to be a much more subtle issue than the separation of two points. Given two intersecting lines such as l_1 and l_2 in Figure 1.1, we would like to define a number that quantifies how 'far apart' the lines are spread. Historically there are a number of solutions to this problem.

The most familiar is the notion of *angle*, which roughly speaking can be described as follows. Draw a circle of radius one with center the intersection A of l_1 and l_2 . Let this circle intersect the lines l_1 and l_2 respectively at points B_1 and B_2 . Then ‘define’ the angle θ between the lines to be the length of the circular arc between B_1 and B_2 , as in the first diagram in Figure 1.1.

If the lines are close to being parallel, the angle is close to zero, while if the lines are perpendicular, the angle turns out to be, after a highly non-trivial calculation that goes back to Archimedes, a number with the approximate value $1.570\,796\,326\dots$

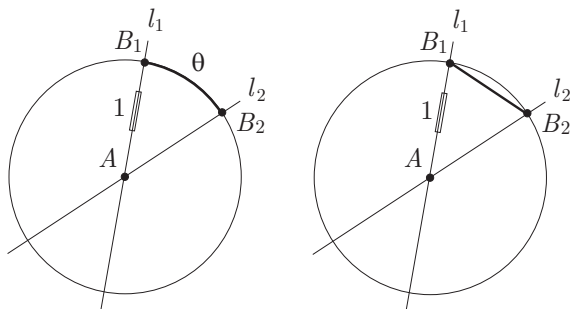
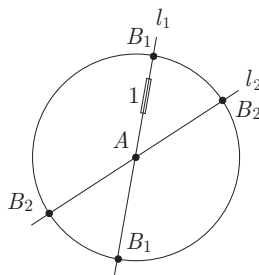


Figure 1.1: Separation of two lines: two dubious approaches

Immediately one observes some difficulties with this ‘definition’.

First of all there are two possible choices for B_1 , and also two possible choices for B_2 . There are then four possible pairs $[B_1, B_2]$ to consider. Each such pair divides the circle into *two* circular arcs. In general there are eight possible circular arcs to measure, and four possible results of those measurements. Furthermore, defining the length of a circular arc is not at all straightforward, a point that will be returned to later.

There are other approaches to the question of how to measure the separation of two lines. For example, with the circle of radius one as above, and the same choice of points B_1 and B_2 , you could consider the length $|B_1B_2|$ as the main object of interest, as in the second diagram in Figure 1.1. When the lines are close to parallel, with the right choices of B_1 and B_2 , $|B_1B_2|$ is close to zero, and when the lines are perpendicular $|B_1B_2|$ is $\sqrt{2}$, independent of the choices.



However in general there are still two different values that can be obtained. This is less than optimal, but it is already an improvement on the first method. An even better choice is the quadrance $Q(B_1, B_2)$, but the two-fold ambiguity in values remains.

A completely different tack is to give up on the separation between lines, and measure only the separation between *rays*. But a line is a more elementary, fundamental and general notion than a ray, so this is capitulation.

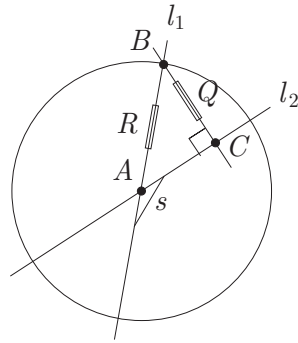


Figure 1.2: Separation of two lines: the right approach

The right idea is remarkably simple, and is shown in Figure 1.2. Take any point $B \neq A$ on one of the lines, and then let C be the foot of the altitude, or perpendicular, from B to the other line. Then define the **spread** $s(l_1, l_2)$ between the two lines l_1 and l_2 to be the ratio of quadrances

$$s(l_1, l_2) \equiv \frac{Q(B, C)}{Q(A, B)} = \frac{Q}{R}. \quad (1.1)$$

This number $s \equiv s(l_1, l_2)$ is independent of the choice of first line, or the choice of the point B on it. It is a unique number, somewhere between 0 and 1, which measures the separation of two lines unambiguously. Note that the circle, which played an important role in the previous constructions, is now essentially irrelevant.

In diagrams, spreads are placed adjacent to small straight line segments joining the relevant lines, instead of the usual circular arcs used to denote angles. There are four possible places to put this spread, each equivalent, as in Figure 1.3.

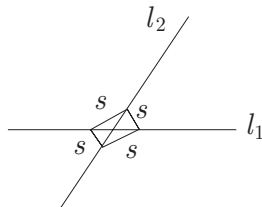


Figure 1.3: Four possible labellings

Why do we not consider the spread to be the ratio of lengths $|B, C| / |A, B|$? This becomes clearer when the spread is expressed in terms of the coordinates of the lines.

Spread in terms of coordinates

A line in the plane can be specified by an equation in the form $y = mx + b$, or the more general form $ax + by + c = 0$. The latter is preferred in this text, for a variety of good reasons, among them being that it includes vertical lines, generalizes well to projective geometry, and is better suited for the jump to linear algebra and higher dimensions.

The equation $ax + by + c = 0$ of a line l is not unique, as you can multiply by an arbitrary non-zero number. This means that it is actually the *proportion* $a : b : c$ that specifies the line, and later on we *define* the line in terms of this proportion.

Two lines l_1 and l_2 with respective equations $a_1x + b_1y + c_1 = 0$ and $a_2x + b_2y + c_2 = 0$ are **parallel** precisely when

$$a_1b_2 - a_2b_1 = 0$$

and **perpendicular** precisely when

$$a_1a_2 + b_1b_2 = 0.$$

The spread between them is unchanged if the lines are moved while remaining parallel, so you may assume that the lines have equations $a_1x + b_1y = 0$ and $a_2x + b_2y = 0$ with intersection the origin $A \equiv [0, 0]$. Let's calculate the spread between these two lines.

A point B on l_1 is $B \equiv [-b_1, a_1]$. An arbitrary point on l_2 is of the form $C \equiv [-\lambda b_2, \lambda a_2]$. The quadrances of the triangle \overline{ABC} are then

$$Q(A, B) = b_1^2 + a_1^2 \quad (1.2)$$

$$Q(A, C) = \lambda^2 (b_2^2 + a_2^2) \quad (1.3)$$

$$Q(B, C) = (b_1 - \lambda b_2)^2 + (\lambda a_2 - a_1)^2. \quad (1.4)$$

Pythagoras' theorem asserts that the triangle \overline{ABC} has a right vertex at C precisely when

$$Q(A, C) + Q(B, C) = Q(A, B).$$

This is the equation

$$\lambda^2 (b_2^2 + a_2^2) + (b_1 - \lambda b_2)^2 + (\lambda a_2 - a_1)^2 = b_1^2 + a_1^2.$$

After expansion and simplification you get

$$2\lambda (a_1a_2 + b_1b_2 - \lambda (a_2^2 + b_2^2)) = 0.$$

Thus $\lambda = 0$ or

$$\lambda = \frac{a_1a_2 + b_1b_2}{a_2^2 + b_2^2}. \quad (1.5)$$

But $\lambda = 0$ precisely when the lines are perpendicular, in which case (1.5) still holds. So substitute (1.5) back into (1.4) and use some algebraic manipulation to show that

$$Q(B, C) = \frac{(a_1b_2 - a_2b_1)^2}{a_2^2 + b_2^2}.$$

Thus

$$s(l_1, l_2) = \frac{Q(B, C)}{Q(A, B)} = \frac{(a_1 b_2 - a_2 b_1)^2}{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}.$$

Now it should be clearer why the definition (1.1) of the spread using a ratio of quadrances is a good idea—the final answer is a *rational expression in the coordinates of the two lines*. If we had chosen to define the spread as a ratio of lengths, then square roots would appear at this point. Square roots are to be avoided whenever possible.

Calculating a spread

To calculate the spread s_1 in the triangle $\overline{A_1 A_2 A_3}$ with side lengths $|A_2, A_3| = 4$, $|A_1, A_3| = 7$ and $|A_1, A_2| = 5$ let B be the foot of the altitude from A_2 to the line $A_1 A_3$, and let $a \equiv |A_1 B|$ and $b \equiv |A_2 B|$ as in Figure 1.4.

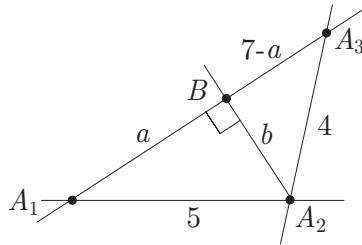


Figure 1.4: Calculating a spread

In the right triangles $\overline{A_1 A_2 B}$ and $\overline{A_3 A_2 B}$, Pythagoras theorem gives

$$a^2 + b^2 = 25 \quad \text{and} \quad (7 - a)^2 + b^2 = 16.$$

Take the difference between these equations to get

$$14a - 49 = 9$$

so that

$$a = 58/14 = 29/7.$$

Then

$$b^2 = 25 - a^2 = 25 - (29/7)^2 = 384/49$$

and so

$$s_1 = \frac{Q(A_2, B)}{Q(A_1, A_2)} = \frac{384/49}{25} = \frac{384}{1225}.$$

Exercise 1.1 Verify also that $s_2 = 24/25$ and $s_3 = 24/49$. \diamond

1.2 Laws of rational trigonometry

This section introduces the main concepts and basic laws of rational trigonometry. So far the discussion has been in the decimal number plane. The algebraic nature of quadrance and spread have the important consequence that the basic definitions and laws may be formulated to hold with coefficients from an arbitrary field. That involves rethinking the above discussion so that no mention is made of distance or angle.

The laws given below are polynomial, and can be derived rigorously using only algebra and arithmetic, in a way that high school students can follow. This is shown in Part II of the book. If you work with geometrical configurations whose points belong to a particular field, then the solutions to these equations also belong to this field.

So these laws are valid in great generality, and it is now possible to investigate Euclidean (and non-Euclidean) geometry over general fields. This is the beginning of *universal geometry*.

A **point** A is an ordered pair $[x, y]$ of numbers. The **quadrance** $Q(A_1, A_2)$ between points $A_1 \equiv [x_1, y_1]$ and $A_2 \equiv [x_2, y_2]$ is the number

$$Q(A_1, A_2) \equiv (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

A **line** l is an ordered proportion $\langle a : b : c \rangle$, representing the equation $ax + by + c = 0$. The **spread** $s(l_1, l_2)$ between lines $l_1 \equiv \langle a_1 : b_1 : c_1 \rangle$ and $l_2 \equiv \langle a_2 : b_2 : c_2 \rangle$ is the number

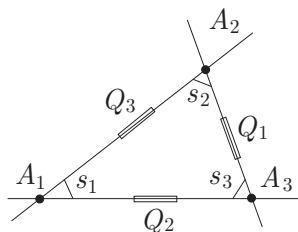
$$s(l_1, l_2) \equiv \frac{(a_1 b_2 - a_2 b_1)^2}{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}.$$

For distinct points A_1 and A_2 the unique line passing through them both is denoted $A_1 A_2$. Given three distinct points A_1, A_2 and A_3 , define the quadrances

$$\begin{aligned} Q_1 &\equiv Q(A_2, A_3) \\ Q_2 &\equiv Q(A_1, A_3) \\ Q_3 &\equiv Q(A_1, A_2) \end{aligned}$$

and the spreads

$$\begin{aligned} s_1 &\equiv s(A_1 A_2, A_1 A_3) \\ s_2 &\equiv s(A_2 A_1, A_2 A_3) \\ s_3 &\equiv s(A_3 A_1, A_3 A_2). \end{aligned}$$



Here are the five main laws of rational trigonometry.

Triple quad formula The three points A_1, A_2 and A_3 are collinear (meaning they all lie on a single line) precisely when

$$(Q_1 + Q_2 + Q_3)^2 = 2(Q_1^2 + Q_2^2 + Q_3^2).$$

Pythagoras' theorem The lines A_1A_3 and A_2A_3 are perpendicular precisely when

$$Q_1 + Q_2 = Q_3.$$

Spread law For any triangle $\overline{A_1A_2A_3}$

$$\frac{s_1}{Q_1} = \frac{s_2}{Q_2} = \frac{s_3}{Q_3}.$$

Cross law For any triangle $\overline{A_1A_2A_3}$ define the **cross** $c_3 = 1 - s_3$. Then

$$(Q_1 + Q_2 - Q_3)^2 = 4Q_1Q_2c_3.$$

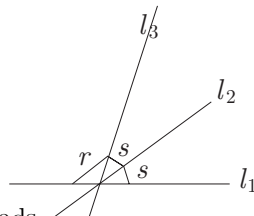
Triple spread formula For any triangle $\overline{A_1A_2A_3}$

$$(s_1 + s_2 + s_3)^2 = 2(s_1^2 + s_2^2 + s_3^2) + 4s_1s_2s_3.$$

These formulas are related in interesting ways, and deriving them is reasonably straightforward. There are also numerous alternative formulations of these laws, as well as generalizations to four quadrances and spreads, which become important in the study of quadrilaterals. The difference between the sides of the Triple quad formula is used to define the *quadrea* of a triangle, which is the single most important number associated to a triangle.

You also need to understand the *spread polynomials*. If l_1, l_2 and l_3 are lines making spreads $s(l_1, l_2) = s(l_2, l_3) \equiv s$, then the Triple spread formula shows that the spread $r \equiv s(l_1, l_3)$ is either 0, in which case l_1 and l_3 are parallel, or $4s(1 - s)$.

The polynomial function $S_2(s) \equiv 4s(1 - s)$ figures prominently in chaos theory, where it is called the *logistic map*.



Generalizing this, $S_n(s)$ is the spread made by n successive spreads of s , which can be formulated in terms of repeated reflections of lines in lines. It turns out that the *spread polynomial* $S_n(s)$ is of degree n in s , and in the decimal number field is closely related to the classical Chebyshev polynomial $T_n(x)$.

In rational trigonometry the remarkable properties of these polynomials extend to general fields, suggesting new directions for the theory of special functions. There are also analogous *cross polynomials*. Together with *oriented triangles*, *signed areas* and *turns*, these are the basic ingredients for rational trigonometry.

1.3 Why classical trigonometry is hard

For centuries students have struggled to master angles, trigonometric functions, and their many intricate relations. Those who learn how to apply the formulas correctly often don't know *why* they are true. Such difficulties are to an extent the natural reflection of an underlying ambiguity at the heart of classical trigonometry. This manifests itself in a number of ways, but can be boiled down to the single critical question:

What precisely is an angle??

The problem is that defining an angle correctly *requires calculus*. This is a point implicit in Archimedes' derivation of the length of the circumference of a circle, using an infinite sequence of successively refined approximations with regular polygons. It is also supported by the fact that *The Elements* [Euclid] does not try to measure angles, with the exception of right angles and some related special cases. Further evidence can be found in the universal reluctance of traditional texts to spell out a clear definition of this supposedly 'basic' concept.

Exercise 1.2 Open up a few elementary and advanced geometry books and see if this claim holds. \diamond

Let's clarify the point with a simple example. The rectangle \overline{ABCD} in Figure 1.5 has side lengths $|AB| = 2$ and $|BC| = 1$. What is the angle θ between the lines AB and AC in degrees to four decimal places? (Such accuracy is sometimes needed in surveying.)

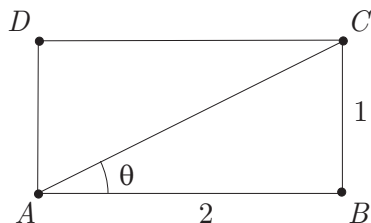


Figure 1.5: A simple problem?

Without tables, a calculator or calculus, a student has difficulty in answering this question, because the usual definition of an angle (page 11) is not precise enough to show how to calculate it. But how can one claim understanding of a mathematical concept without being able to compute it in simple situations? If the notion of an angle θ cannot be made completely clear from the beginning, it cannot be fundamental.

Of course there is one feature about angles that is very useful—they transform the essentially nonlinear aspect of a circle into something linear, so that for example the sum of adjacent angles can be computed using only addition, at least if the angles are small. However the price to be paid for this convenience is high.

One difficulty concerns units. Most scientists and engineers work with degrees, the ancient Babylonian scale ranging from 0° to 360° around a full circle, although in Europe the grad is also used—from 0 to 400. Mathematicians officially prefer radian measure, in which the full circle gets the angle of $6.283185\dots$, also known as 2π . The radian system is theoretically convenient, but practically a nuisance. Whatever scale is chosen, there is then the ambiguity of adding ‘multiples of a circle’ to any given angle, and of deciding whether angles are to have signs, or to be always positive.

Angles in the range from 90° to 180° are awkward and students invariably have difficulty with them. For example the angle 156° is closely related to the angle 24° , and a question involving the former can be easily converted into a related question about the latter. There is an unnecessary duplication of information in having separate measurements for an angle and its supplement. For angles in the range from 180° to 360° similar remarks apply. Indeed, such angles are rarely necessary in elementary geometry.

Another consequence of relying on angles is a surprisingly limited range of configurations that can be analysed in classical trigonometry using elementary techniques. Students are constantly given examples that deal essentially with $90^\circ/60^\circ/30^\circ$ or $90^\circ/45^\circ/45^\circ$ triangles, since these are largely the only ones for which they can make unassisted calculations.

Small wonder that the trigonometric functions $\cos \theta$, $\sin \theta$ and $\tan \theta$ and their inverse functions cause students such difficulties. Although pictures of unit circles and ratios of lengths are used to ‘define’ these in elementary courses, it is difficult to understand them correctly without calculus. For example, the function $\tan \theta$ is given, for θ in a suitable range, by the infinite series

$$\tan \theta \approx \theta + \frac{1}{3}\theta^3 + \frac{2}{15}\theta^5 + \frac{17}{315}\theta^7 + \frac{62}{2835}\theta^9 + \dots$$

while the inverse function $\arccos x$ is given by

$$\arccos x \approx \frac{1}{2}\pi - x - \frac{1}{6}x^3 - \frac{3}{40}x^5 - \frac{5}{112}x^7 - \frac{35}{1152}x^9 + \dots$$

and in the case of $\tan \theta$ the coefficients defy a simple closed-form expression. This kind of subtlety is largely hidden from beginning students with the reliance on the power of calculators.

If the foundations of a building are askew, the entire structure is compromised. The underlying ambiguities in classical trigonometry revolving around the concept of angle are an impediment to learning mathematics, weaken its logical integrity, and introduce an unnecessary element of approximation and inaccuracy into practical applications of the subject.

1.4 Why rational trigonometry is easier

The key concepts of rational trigonometry are simpler, and mathematically more natural, than those of classical trigonometry. *Quadrance* is easier to work with than *distance* (as most mathematicians already know) and a *spread* is more elementary than an *angle*. The spread between two lines is a dimensionless quantity, and in the rational or decimal number fields takes on values between 0 and 1, with 0 occurring when lines are parallel and 1 occurring when lines are perpendicular. Forty-five degrees becomes a spread of $1/2$, while thirty and sixty degrees become respectively spreads of $1/4$ and $3/4$. What could be simpler than that?

Another advantage with spreads is that the measurement is taken between lines, not rays. As a consequence, the two range of angles from 0° to 90° and from 90° to 180° are treated symmetrically. For example, the spreads associated to 24° and 156° are identical (namely $s = 0.165434\dots$). If one wishes to distinguish between these two situations in the context of rays, a single binary bit of additional information is required, namely the choice between *acute* (*ac*) and *obtuse* (*ob*). The *Triangle spread rules* described in the applications part of the book deal with this additional information.

There are many triangles that can be analysed completely by elementary means using rational trigonometry, giving students exposure to a wider range of examples.

The straightforwardness of rational trigonometry is also evident from the polynomial form of the basic laws, which do not involve any transcendental functions, rely only on arithmetical operations, and are generally quadratic in any one variable. As a consequence, tables of values of trigonometric functions, or modern calculators, are not necessary to do trigonometric calculations. Computations for simple problems can be done by hand, more complicated problems can use computers more efficiently.

With the introduction of rational polar and spherical coordinates in calculus, this simplicity can be put to work in solving a wide variety of sophisticated problems. Computations of volumes, centroids of mass, moments of inertia and surface areas of spheres, paraboloids and hyperboloids become in many cases more elementary. This simplicity extends to higher dimensional spaces, where the basic algebraic relations reduce the traditional reliance on pictures and argument by analogy with lower dimensions.

Rational trigonometry works over any field. So the difficulties inherent in the decimal and 'real number' fields can be avoided. It is not necessary to have a prior model of the continuum before one begins geometry. Furthermore many calculations become much simpler over finite fields, which can be a significant advantage.

Even when extended to *spherical trigonometry* and *hyperbolic geometry*, the theory's basic formulas are polynomial, as will be shown in a future volume. The concepts of rational trigonometry also have advantages when working statistically, but this point will not be developed here.

1.5 Comparison example

Here is an illustration of how rational trigonometry allows you to solve explicit practical problems, and why this method is superior to the classical approach. There is nothing particularly special about this example or the numbers appearing in it.

Problem 1 The triangle $\overline{A_1A_2A_3}$ has side lengths $|A_1A_2| = 5$, $|A_2A_3| = 4$ and $|A_3A_1| = 6$. The point B is on the line A_1A_3 with the angle between A_1A_2 and A_2B equal to 45° . What is the length $d \equiv |A_2B|$?

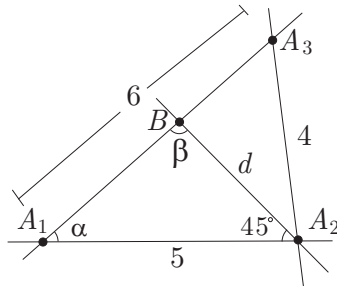


Figure 1.6: Classical version

Classical solution

Let the angles at A_1 and B be α and β respectively, as shown in Figure 1.6. The Cosine law in the triangle $\overline{A_1A_2A_3}$ gives

$$4^2 = 5^2 + 6^2 - 2 \times 5 \times 6 \times \cos \alpha$$

so that using a calculator

$$\alpha = \arccos \frac{3}{4} \approx 41.4096^\circ.$$

Since the sum of the angles in $\overline{A_1A_2B}$ is 180° ,

$$\beta \approx 180^\circ - 45^\circ - 41.4096^\circ \approx 93.5904^\circ.$$

Now the Sine law in $\overline{A_1A_2B}$ states that

$$\frac{\sin \alpha}{d} = \frac{\sin \beta}{5}$$

so that again using the calculator

$$d \approx \frac{5 \sin 41.4096^\circ}{\sin 93.5904^\circ} \approx 3.3137.$$

Rational solution

To apply rational trigonometry, first convert the initial information about lengths and angles into quadrances and spreads. The three quadrances of the triangle are the squares of the side lengths, so that $Q_1 = 16$, $Q_2 = 36$ and $Q_3 = 25$.

The spread corresponding to the angle 45° is $1/2$, as can be easily seen from Figure ??.

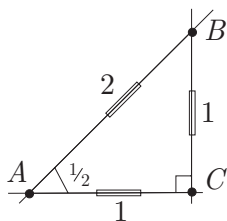


Figure 1.7: Spread of $1/2$

Let s be the spread between the lines A_1A_2 and A_1A_3 , and let r be the spread between the lines BA_1 and BA_2 . Let $Q \equiv Q(B, A_2)$. This yields Figure 1.8 involving only quantities from rational trigonometry.

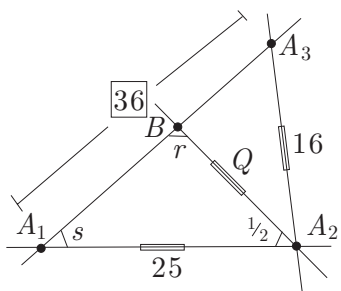


Figure 1.8: Rational version

Let's follow the use of the basic laws to first find s , then r , and then Q . Use the Cross law in $\overline{A_1A_2A_3}$ to get

$$(25 + 36 - 16)^2 = 4 \times 25 \times 36 \times (1 - s)$$

so that $s = 7/16$.

Use the Triple spread formula in $\overline{A_1A_2B}$ to obtain for r the quadratic equation

$$\left(\frac{7}{16} + \frac{1}{2} + r\right)^2 = 2 \left(\frac{49}{256} + \frac{1}{4} + r^2\right) + 4 \times \frac{7}{16} \times \frac{1}{2} \times r.$$

This simplifies to

$$r^2 - r + \frac{1}{256} = 0$$

so that

$$r = \frac{1}{2} \pm \frac{3}{16}\sqrt{7}.$$

For each of these values of r , use the Spread law in $\overline{A_1A_2B}$

$$\frac{r}{25} = \frac{s}{Q}$$

and solve for Q , giving values

$$Q_1 = 1400 - 525\sqrt{7}$$

$$Q_2 = 1400 + 525\sqrt{7}.$$

To convert these answers back into distances, take square roots

$$d_1 = \sqrt{Q_1} \approx 3.3137\dots$$

$$d_2 = \sqrt{Q_2} \approx 264.056\dots$$

Clearly the answer is d_1 . Where does d_2 come from? The initial information (apart from the picture) describes two different possibilities. The second one is that the line A_2B makes a spread of $1/2$ with A_1A_2 as shown in Figure 1.9, with the intersection B off the page in the direction shown, since $s = 7/16$ is less than $1/2$.

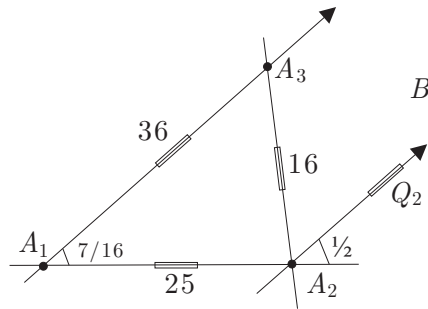


Figure 1.9: Alternate possibility

Comparison

Clearly the solution using rational trigonometry is more accurate, and reveals that the irrational number $\sqrt{7}$ is intimately bound up in this problem, which is certainly not obvious from the initial data, nor apparent from the classical solution.

Perhaps it appears that the classical solution is however shorter. That is because the main computations of $\arccos 3/4$, $\sin \alpha$ and $\sin \beta$ were done by the calculator. The

rational solution can be found entirely by hand. To be fair, it should be added that classical trigonometry *is* capable of getting the exact answers obtained by the rational method here, but a more sophisticated approach is needed. One can simulate rational trigonometry from within classical trigonometry, essentially by never calculating angles.

The rational method has solved two problems at the same time. This phenomenon is closely related to the fact that spreads, unlike angles, do not distinguish between the concepts of *acute* and *obtuse*, although in the applications section these concepts do appear in the context of the Triangle spread rules.

Since rational trigonometry is so much simpler than the existing theory, why was it not discovered a long time ago? In fact bits and pieces of rational trigonometry have surfaced throughout the history of mathematics, and have also been used by physicists.

It is well known, for example, that many trigonometric identities and integrals can be easily verified and solved by appropriate rational substitutions and/or use of the complex exponential function. The ubiquitous role of squared quantities in Euclidean geometry is also familiar, as is the insight of linear algebra to avoid angles as much as possible by using dot products and cross products. Modern geometry is very aware of the central importance of quadratic forms. In Einstein's theory of relativity the quadratic 'interval' plays a crucial role.

Perhaps these different clues have not been put together formerly because of the strength of established tradition, and in particular the reverence for Greek geometry.

1.6 Ancient Greek triumphs and difficulties

The serious study of geometry began with the ancient Greeks, whose mathematical heritage is one of the most remarkable and sublime achievements of humanity. Nevertheless, they encountered some difficulties, some of which persist today and interfere with our understanding of geometry.

A Pythagorean dilemma

The Pythagoreans believed that the workings of the universe could ultimately be described by *proportions between whole numbers*, such as 4 : 5. This was not necessarily a mistake—the future may reveal the utility of such an idea. After early successes in applying this principle to music and geometry, however, they discovered a now famous dilemma. The length of the diagonal to the length of the side of a square, namely $\sqrt{2} : 1$, could not be realized as a proportion between two whole numbers.

Unfortunately, this difficulty was something of a red herring. Had they stuck with their original beliefs in the workings of the Divine Mind, and boldly concluded that the *squares* of the lengths ought to be more important than the lengths themselves, then the history of mathematics would look quite different.

Circles and trigonometric functions

Points, lines and circles are the most important objects in Euclidean geometry. The ancient Greeks were much enamoured with the elegance of the circle, and chose to give the compass equal importance to the straightedge in their geometry. Following their lead, mathematics courses still rely heavily on the circle and its properties to introduce the basic concepts of classical trigonometry. Most high school students will be familiar with the unit circle used to define both angles, as on page 11, and the trigonometric ratios $\cos \theta$, $\sin \theta$ and $\tan \theta$.

This thinking is unfortunate. A circle is a more complicated object than a line. In Cartesian coordinates it is described by a second degree equation, while a line is described by a first degree equation. Defining arc lengths of curves other than line segments is quite sophisticated, even for arcs of circles. So it does not make mathematical sense to treat circles on a par with lines, or to attempt to use circles to define the basic measurement between lines.

In rational trigonometry *points and lines are more fundamental than circles*. The essential formulas involve only points and lines, while circles appear later as particular kinds of conics in universal geometry, where they can be studied purely in the rational context without the use of trigonometric functions. Even rotations can be investigated solely within universal geometry.

The essential roles of $\cos \theta$, $\sin \theta$ and $\tan \theta$ arise rather when considering *uniform motion* around a circle, which has more to do with mechanics than geometry. Trigonometric functions also occur in harmonic analysis and complex analysis, where they are however secondary to the complex exponential function, whose theory is simpler and more fundamental.

And what about all those engineering problems with spherical or cylindrical symmetry? Rational trigonometry solves many of these problems in a more direct fashion, as shown by the last chapter in this book. The reality is that *the trigonometric functions are overrated*, and their intensive study in high schools is for many students an unnecessary complication.

Euclid's ambiguities

To develop a general and powerful geometry, it is critical to ensure that the basic definitions are completely unambiguous, and that circular reasoning plays no role.

Unfortunately, it is too easy to loosely convey a geometric idea with a picture. Students can seemingly learn concepts—such as parallel and perpendicular lines, triangles, quadrilaterals, tangent lines to a circle, angles and many others—by being shown a sufficient number of illustrative examples. However, when they want to use these ideas in proofs or precise calculations, the lack of clear definitions becomes a burden, not a convenience.

One of the appeals of mathematics is that one can understand completely the basic concepts and then logically derive the entire theory, step by step, using only valid logical arguments. When students meet mathematics in which the logical structure is not clearly evident, their confidence and understanding suffers. This applies to all students, even—perhaps *especially*—to weaker ones.

The extraordinary work of Euclid, *The Elements*, which set the tone of mathematics for two thousand years, is guilty of ambiguity in its initial definitions, despite the otherwise generally high standard of logic and precision throughout. The historical dominance of Euclid's work, with its supposedly axiomatic framework, has coloured mathematical thinking to this day, not altogether in a positive way.

According to Euclid, a *point* is that which 'has no parts or magnitude', a *line* is 'length without breadth', and *perpendicularity* of lines is defined as 'lying evenly between each other'. For someone who already has an intuitive idea of what a point, line and perpendicular mean, perhaps by having been shown many relevant pictures from an early age, these definitions may seem reasonable, but to someone without prior experience of geometry they are surely quite unintelligible.

For many centuries it appeared that there was no alternative to the geometry of Euclid. But in the seventeenth century the work of Fermat, Descartes and others opened up a dramatic new possibility. The Cartesian idea of defining a point as an ordered pair of numbers, and then analysing curves by algebraic manipulations of their defining equations is very powerful. It reduces the foundations of geometry to that of arithmetic and algebra, and allows the formulation of many geometrical questions over general fields.

Rational trigonometry adopts the logical and precise Cartesian approach of Descartes, but it adds a metrical aspect with the notions of quadrance and spread, still in the purely algebraic setting. This allows Euclidean geometry to be put on a firm and proper foundation.

1.7 Modern ambiguities

In the nineteenth century some of the difficulties and limitations with Euclid's geometry became more recognized, and for this reason and others, the twentieth century saw a slow but steady decline in the importance of Euclid as the basis of education in geometry. Unfortunately, there was no successful attempt to put into place an alternative framework.

One of those who realized the importance of the task was D. Hilbert, whose *Foundations of Geometry* [Hilbert] was an attempt to shore up the logical deficiencies of Euclid. This approach was not universally accepted, and the thought arose that perhaps the task was intrinsically impossible, and that a certain amount of ambiguity could not be removed from the subject.

Twentieth century mathematics largely resigned itself to an acceptance of vagueness in the foundations of mathematics. It became accepted wisdom that any treatment of Euclidean geometry had to incorporate the philosophical aspects of the ‘construction of the continuum’ and Cantor’s theory of ‘infinite sets’. Distance, angle and the trigonometric functions required the usual framework of modern analysis, along with its logical deficiencies.

The common feeling was that the best solution was to proceed along the lines of an *axiomatic system*, involving formal manipulations of underlying concepts that are not given meaning. It was often supposed that this is what Euclid had in mind.

But this is *not* what Euclid had in mind. When he stated his ‘axioms’, Euclid wanted only to clarify which facts he was going to regard as obvious, before deriving all other facts using deduction. The meaning of these basic facts was never in question. Euclid’s work in this sense is quite different from formalist systems which came to be modelled on it.

When we begin a study of universal geometry, these preconceived twentieth century notions must be put aside. It *is* possible to start from the beginning, and to aspire to give a complete and precise account of geometry, without any missing gaps. The logical framework should be clearly visible at all times. The basic concepts should be easy to state to someone beginning the subject. Pulling in theorems from the outside is not allowed. The ‘continuum’ need not be understood. ‘Infinite set theory’ and its attendant logical difficulties—deliberately swept under the carpet by modern mathematics—should be avoided.

With such an approach the logical coherence and beauty of mathematics become easier to verify and appreciate directly, without appeals to authority and convention. To build up mathematics properly, axioms are not necessary. You do not have to engage in philosophy to do geometry, nor require sophisticated modern logic to understand foundations. Clear thinking, careful definitions and an interest in applications should suffice.

So let’s begin the story, starting with an informal review of the basic assumed knowledge from algebra and arithmetic. Then let’s build up geometry, one step at a time.