AN ILLUSTRATED THEORY OF NUMBERS

MARTIN H. WEISSMAN

AMERICAN MATHEMATICAL SOCIETY
Providence, Rhode Island

2010 *Mathematics Subject Classification*. Primary 11A05, 11A07, 11A15, 11A41, 11A51, 11D04, 11D09, 11E16, 11E41.

For additional information and updates on this book, visit
www.ams.org/bookpages/mbk-105

Library of Congress Cataloging-in-Publication Data

Names: Weissman, Martin H., 1976-
Title: An illustrated theory of numbers / Martin H. Weissman.
Description: Providence, Rhode Island : American Mathematical Society, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2017003379 | ISBN 9781470434939 (alk. paper)
Subjects: LCSH: Number theory. | AMS: Number theory – Elementary number theory – Multiplicative structure; Euclidean algorithm; greatest common divisors. msc | Number theory – Elementary number theory – Congruences; primitive roots; residue systems. msc | Number theory – Elementary number theory – Power residues, reciprocity. msc | Number theory – Elementary number theory – Primes. msc | Number theory – Elementary number theory – Factorization; primality. msc | Number theory – Diophantine equations – Linear equations. msc | Number theory – Diophantine equations – Quadratic and bilinear equations. msc | Number theory – Forms and linear algebraic groups – General binary quadratic forms. msc | Number theory – Forms and linear algebraic groups – Class numbers of quadratic and Hermitian forms. msc
Classification: LCC QA241 .W354 2017 | DDC 512.7–dc23 LC record available at https://lccn.loc.gov/2017003379

♾ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 22 21 20 19 18 17

I PROPOUND THIS EASY PROCESS OF COMPUTATION, DELIGHTFUL BY ITS ELEGANCE, PERSPICUOUS WITH WORDS CONCISE, SOFT AND CORRECT, AND PLEASING TO THE LEARNED.

BHASKARA II, FROM *Lilavati*, 1150

IT IS CHARACTERISTIC OF HIGHER ARITHMETIC THAT MANY OF ITS MOST BEAUTIFUL THEOREMS CAN BE DISCOVERED BY INDUCTION WITH THE GREATEST OF EASE BUT HAVE PROOFS THAT LIE ANYWHERE BUT NEAR AT HAND AND ARE OFTEN FOUND ONLY AFTER MANY FRUITLESS INVESTIGATIONS WITH THE AID OF DEEP ANALYSIS AND LUCKY COMBINATIONS.

CARL FRIEDRICH GAUSS, 1817

I HOPE THAT POSTERITY WILL JUDGE ME KINDLY, NOT ONLY AS TO THE THINGS WHICH I HAVE EXPLAINED, BUT ALSO AS TO THOSE WHICH I HAVE INTENTIONALLY OMITTED SO AS TO LEAVE TO OTHERS THE PLEASURE OF DISCOVERY.

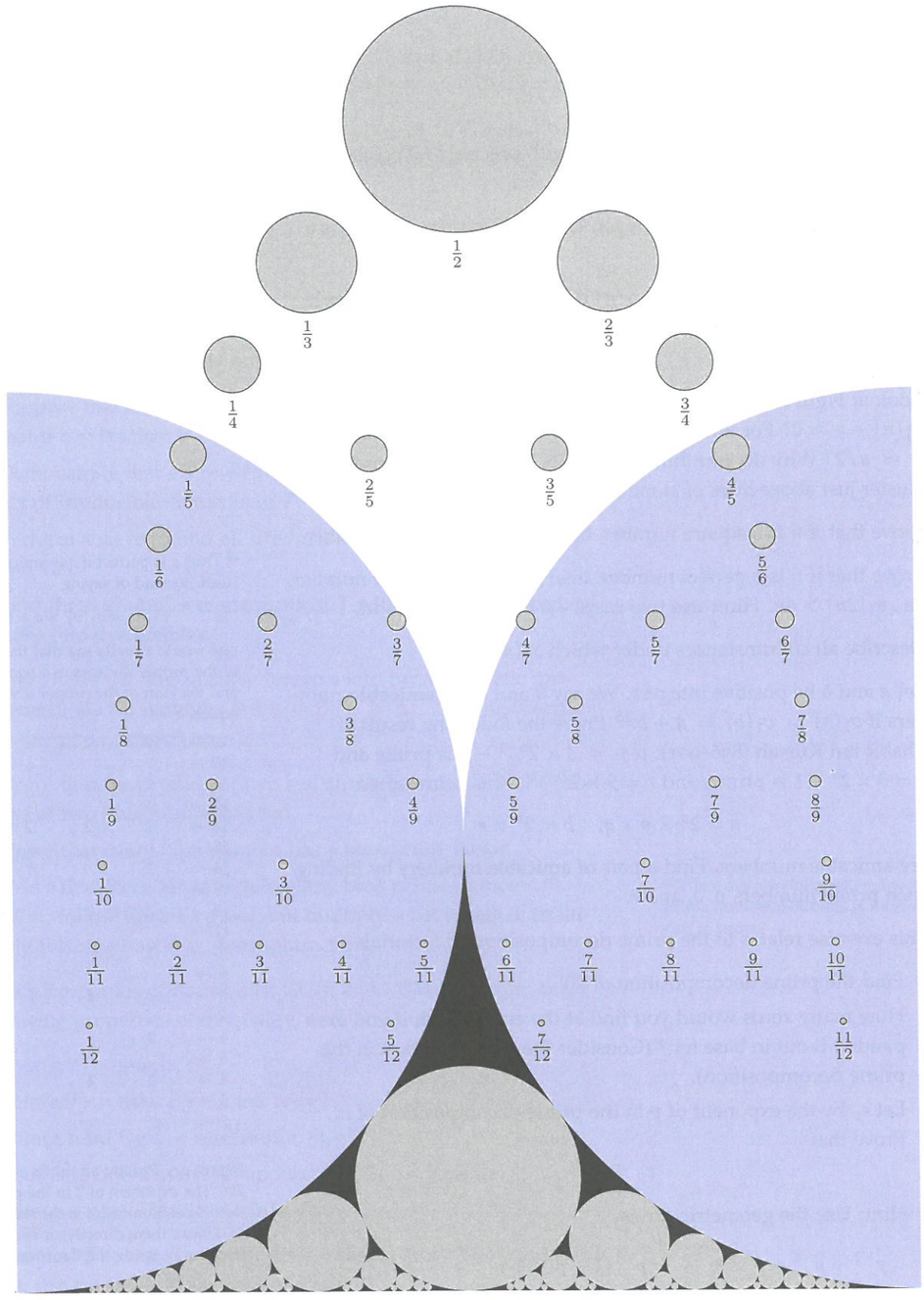
RENÉ DESCARTES, FROM *La Géométrie*, 1637.

THUS, EVEN IF YOUR PROBLEM IS NOT A PROBLEM OF GEOMETRY, YOU MAY TRY TO *draw a figure*. TO FIND A LUCID GEOMETRIC REPRESENTATION FOR YOUR NONGEOMETRICAL PROBLEM COULD BE AN IMPORTANT STEP TOWARD THE SOLUTION.

GEORGE PÓLYA, FROM *How to Solve It*, 1946.

THAT THE GEOMETER'S MIND IS NOT LIKE THE PHYSICIST'S OR THE NATURALIST'S, ALL THE WORLD WOULD AGREE; BUT MATHEMATICIANS THEMSELVES DO NOT RESEMBLE EACH OTHER; SOME RECOGNIZE ONLY IMPLACABLE LOGIC, OTHERS APPEAL TO INTUITION AND SEE IN IT THE ONLY SOURCE OF DISCOVERY.

HENRI POINCARÉ, FROM *The Value of Science*, 1905



3
Rational and Constructible Numbers

THE NATURAL NUMBERS are 0, 1, 2, 3, 4, ... The integers include natural numbers and their negatives: $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. Within the integers we can always add, subtract, and multiply.

The rational numbers are those numbers[1] which are quotients of integers (without dividing by zero). Every integer is a rational number: If n is an integer, then $n = n/1$ is a rational number. Within the rational numbers, we can always add, subtract, and multiply, and moreover we can divide, except by zero.

It is a bit difficult to place rational numbers. Find $2/5$ on the number line. Now locate $3/7$. Which one is to the right and which one is to the left?

[1] The reader might notice that throughout the book we are assuming a large set of numbers, with familiar operations and properties. This book does not contain an axiomatic construction of numbers. Instead, we take for granted the real numbers, all properties of arithmetic operations therein, equality and order, and the intermediate value theorem for polynomials (for the sake of square roots).

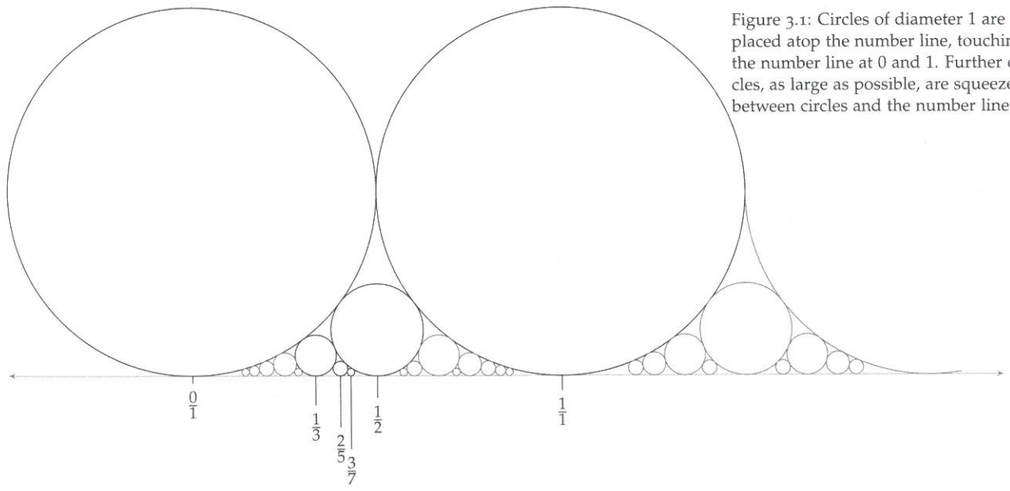


Figure 3.1: Circles of diameter 1 are placed atop the number line, touching the number line at 0 and 1. Further circles, as large as possible, are squeezed between circles and the number line.

EVERY RATIONAL NUMBER can be expressed as a/b for some integers a and b (with b nonzero). An expression a/b, with a and b integers, is called a **fraction**. The relationship[2] between fractions and rational numbers is the relationship between a word and what it means. A *fraction* is a way of writing a *rational number*. For what comes later, it is convenient to allow fractions like $1/0$ and $-17/0$, even though they do not represent rational numbers.

For example, $2/3$ and $4/6$ are *different* fractions, but they represent the *same* rational number. The *fraction* $1/0$ does not represent any rational number; only fractions with nonzero denominators represent rational numbers.

We learn in school that the different fractions $2/3$ and $4/6$ are *equal* as rational numbers. More generally, equality of rational numbers is defined by the rule

$$\frac{a}{b} = \frac{c}{d}, \text{ if } ad = bc.$$

Using this rule, we find that for nonzero integers n,

$$\frac{a}{b} = \frac{n \cdot a}{n \cdot b}, \text{ since } a(nb) = b(na).$$

We call a fraction a/b **reduced** if $\mathrm{GCD}(a,b) = 1$ and $b > 0$. It is also sometimes convenient[3] to call the fraction $1/0$ reduced, though it represents no rational number.

Theorem 3.1 (Reduction of fractions) *If a/b is a fraction with $b \neq 0$, then there exists a unique reduced fraction c/d such that $a/b = c/d$. Moreover, there exists a nonzero integer n such that $a = nc$ and $b = nd$.*

PROOF: Since $b \neq 0$, $\mathrm{GCD}(a,b) \neq 0$. If $b < 0$, let $n = -\mathrm{GCD}(a,b)$; if $b > 0$, let $n = \mathrm{GCD}(a,b)$. In either case, we have $n \mid a$ and $n \mid b$. Let $c = a/n$ and $d = b/n$. By construction,

$$a = nc \text{ and } b = nd \text{ and } a/b = c/d \text{ and } d > 0 \text{ and } \mathrm{GCD}(c,d) = 1.$$

The fact that $\mathrm{GCD}(c,d) = 1$ follows from Corollary 1.29.

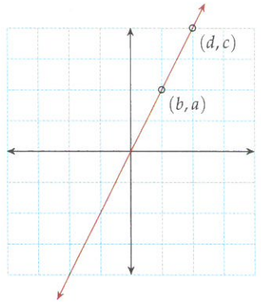
For uniqueness, suppose that $a/b = c/d = e/f$ for another reduced fraction e/f. Then

$$cf = de \text{ and } \mathrm{GCD}(c,d) = \mathrm{GCD}(e,f) = 1 \text{ and } d > 0 \text{ and } f > 0.$$

Since $c \mid cf = de$, Euclid's Lemma (Lemma 2.12) implies $c \mid e$. Similarly, $e \mid de = cf$, and by Euclid's Lemma, $e \mid c$. Likewise, we find $d \mid f$ and $f \mid d$. Therefore $c = \pm e$ and $d = \pm f$. Since $d > 0$ and $f > 0$, we have $d = f$. Since $cf = de$ and $d = f$, it follows quickly that $c = e$ as well. ∎

[2] **Semiotics** is the study of signs and what they signify; it is crucial to have different language to describe expressions – arrangements of symbols on a page – and their numerical meaning.

Since a fraction is just an ordered pair of integers – just written a/b – one associate to a fraction a grid-point in the plane.



The fractions a/b and c/d are thought of as grid-points (b,a) and (d,c). The equality of rational numbers $a/b = c/d$ reflects the fact that both lie on a line of the same slope through the origin. The order, from a/b to (b,a), reverses since slope is defined as "rise over run," so a must correspond to the rise and b to the run. The rational number *is* the slope of the line.

[3] We say that a fraction of the form $a/0$ represents ∞ ("infinity"), if $a \neq 0$. We think of $1/0$ as the reduced fraction representation of ∞. Note that both $1/0$ from $-1/0$ represent the same ∞; there is no "negative infinity" here.

PRIME DECOMPOSITION can be extended from positive integers to all nonzero rational numbers. The extension to nonzero integers is ad hoc; every nonzero integer a has a unique expression of the form

$$a = \pm 2^{e_2} 3^{e_3} 5^{e_5} 7^{e_7} \cdots,$$

in which the exponents $e_2, e_3, e_5, e_7, \ldots$ are natural numbers. Positive integers have sign $+$, and negative integers have sign $-$ in the above decomposition.[4]

The prime decomposition of rational numbers permits not just natural exponents but integer exponents.

Proposition 3.2 *If x is a nonzero rational number, then there exist unique **integers** e_2, e_3, e_5, e_7, etc., such that all but finitely many are zero and*

$$x = \pm 2^{e_2} 3^{e_3} 5^{e_5} 7^{e_7} \cdots.$$

PROOF: Every nonzero rational number can be uniquely expressed as a reduced fraction $x = a/b$. Dividing the prime decomposition of the nonzero integer a by that of the positive integer b, prime by prime, leads to a decomposition of the form above.[5]

For uniqueness, consider a prime decomposition

$$x = \pm 2^{e_2} 3^{e_3} 5^{e_5} 7^{e_7} \cdots.$$

Consider the set $\{p_1, \ldots, p_s\}$ of primes with positive exponent and the set $\{q_1, \ldots, q_t\}$ of primes with negative exponent, in the prime decomposition above. Define integers

$$c = \pm p_1^{e_{p_1}} \cdots p_s^{e_{p_s}} \text{ and } d = q_1^{-e_{q_1}} \cdots q_t^{-e_{q_t}}.$$

Since the sets of primes are disjoint, $\mathrm{GCD}(c,d) = 1$. Moreover, $d > 0$ and $x = c/d$. Since x has a unique expression as a reduced fraction, $c = a$ and $d = b$. Theorem 2.15 implies that the decompositions of c and d above are the unique prime decompositions of a and b, respectively. Hence the only prime decomposition of $x = a/b$ is that which arises from the unique prime decompositions of a and b. ∎

In practice, the prime decomposition of rational numbers is no harder than the prime decomposition of numerator and denominator.

$$\frac{3}{4} = \frac{3^1}{2^2} = 2^{-2}3^1,$$
$$\frac{11}{6} = \frac{11}{2^1 3^1} = 2^{-1}3^{-1}5^0 7^0 11^1.$$

As for positive integers, prime decomposition transforms multiplication of rational numbers into addition of exponents.

$$\frac{3}{4} \cdot \frac{11}{6} = 2^{-2-1} \cdot 3^{1-1} \cdot 5^0 \cdot 7^0 \cdot 11^1 = 2^{-3}3^0 5^0 7^0 11^1.$$

[4] One might even consider (-1) to be another prime; this perspective has been encouraged by John H. Conway, but adoption has not been widespread.

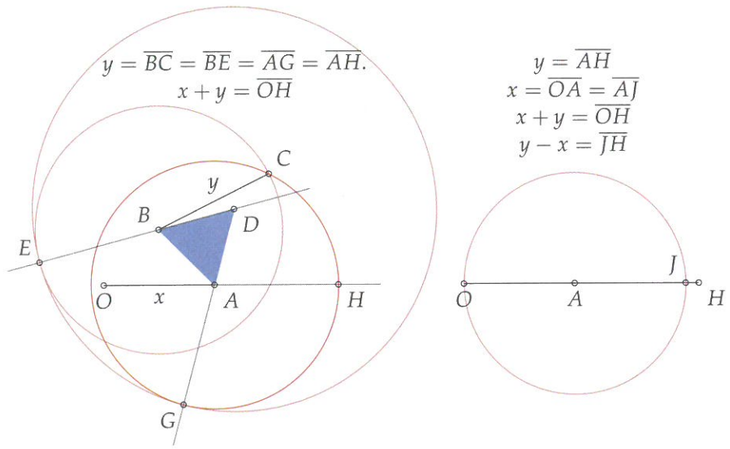
[5] Division yields subtraction of exponents:

$$\frac{p^{f_p}}{p^{g_p}} = p^{f_p - g_p}.$$

Thus the exponent of p in the decomposition of a/b is the difference of exponents in the decompositions of a and b.

René Descartes began his *Géométrie*[6] by describing the geometry of arithmetic operations – addition, subtraction, multiplication, division, and extraction of roots. To understand questions of rationality and irrationality, we present some geometric constructions in the spirit of Descartes (1596–1650).

In Propositions I.2 and I.3, of the *Elements*, Euclid constructs the sum and difference of two line segments $x = \overline{OA}$ and $y = \overline{BC}$.



[6] See "The Geometry of René Descartes," translated from the French and Latin by D.E. Smith and M.L. Latham, Dover Publications, 1954.

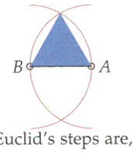


Figure 3.2: Euclid's steps are, in order: connect A to B by a line segment, and construct an equilateral triangle ABD on that line segment (see the figure above for the method). Extend the lines BD and AD indefinitely. Construct the circle at center B, with radius BC. Let E be the intersection of that circle with the line extending BD. Now construct the circle at center D with radius DE. Let G be the intersection of this circle with the line extending AD. Finally, construct the circle with center A and radius AG. Let H be the intersection of this circle with the line extending OA. Then AH has length y, and OH has length $x + y$.

To find the difference $y - x$, see the figure on the right. Draw a circle of radius x with center A. Its intersection with AH is a point J. Since AH has length y, and AJ has length x, we find that JH has length $y - x$.

Descartes uses similar triangles to construct the product and quotient of two lengths x and y, relative to a fixed unit of measure.

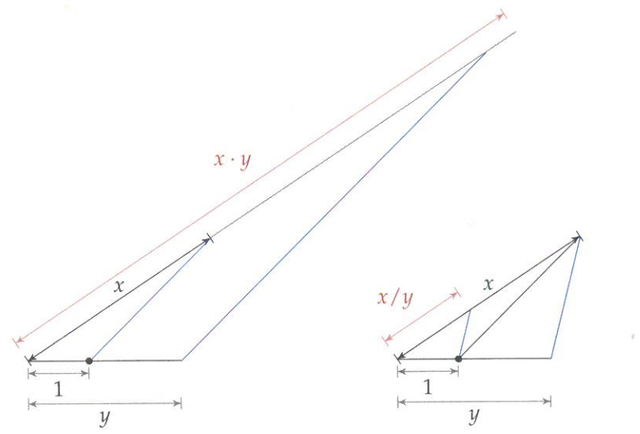
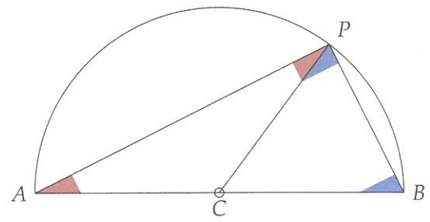
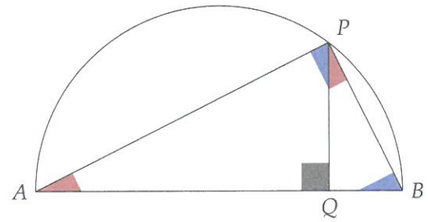


Figure 3.3: Here we arrange the segments of lengths 1, x, and y, using Euclidean constructions as above. Parallel lines (here, in blue) are constructed, yielding pairs of similar triangles. Similarity demonstrates that the red line segments have lengths $x \cdot y$ and x/y as displayed.

Following Euclid,[7] Descartes also gives a geometric construction of square roots. It is based directly on Euclid's construction of the mean proportional, and the figures below demonstrate Euclid's ideas.

[7] Compare to the *Elements*, Propositions VI.13 and III.31.





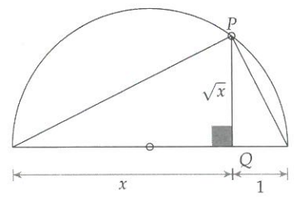
On the left, C is the center of the circle and all radii are equal, so triangles ACP and BCP are isosceles; it follows that the two red angles equal each other, and the two blue angles equal each other. The sum of two red angles and two blue angles equals 180°, the total of the angles of triangle ABP. Hence one red angle and one blue angle sum to 90°, a right angle. The angle APB at the vertex P is a right angle.

On the right, drop a perpendicular from P to the line AB. The red angles in this figure are equal, since they are both complements[8] of the same blue angle APQ. Similarly, the blue angles in this figure are equal, since they are both complements of the same red angle QPB. Therefore the right triangles PQA and BQP are similar triangles.

[8] **Complementary angles** are those that add to 90°.

By similarity, we find a proportion

$$\frac{\text{length}(BQ)}{\text{length}(PQ)} = \frac{\text{length}(PQ)}{\text{length}(QA)}.$$



If we have lengths 1 and x already constructed, this proportion gives a construction of $\sqrt{x}$. Produce a line segment of length $1 + x$, and draw a semicircle whose diameter is the line segment. The altitude will be[9] $\sqrt{x}$ as in the figure on the right.

Constructible numbers are those lengths which can be obtained, beginning with a unit line segment, and carrying out only straightedge and compass drawings; one may extend lines through any two points, draw circles whose radius is an existing segment, and that is all. The constructions here demonstrate that constructible numbers include all those which can be obtained, beginning with only 1, by addition, subtraction ($y - x$ when $y > x$), multiplication, division (not by zero, of course), and square roots. Constructible numbers include, therefore, all positive rational numbers. In addition, constructible numbers include $\sqrt{2}$ and more exotic numbers like $\sqrt{8 - \sqrt{7}}$.

[9] Indeed, the proportion gives

$$\frac{1}{\text{length}(PQ)} = \frac{\text{length}(PQ)}{x},$$

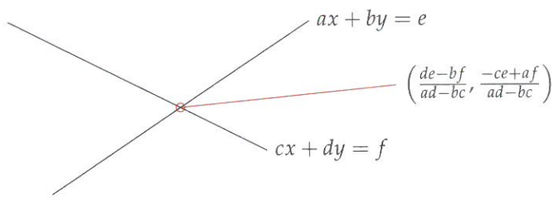
which implies

$$\text{length}(PQ)^2 = x,$$

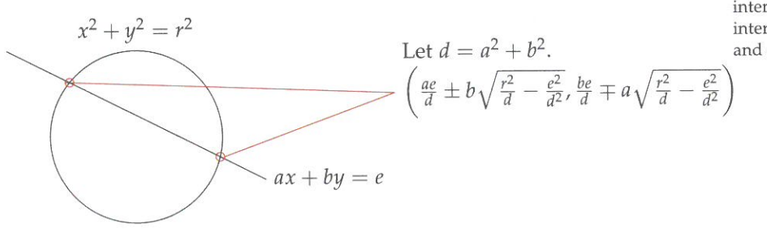
so the altitude PQ has length $\sqrt{x}$.

Theorem 3.3 (Algebraic characterization of constructible numbers) *Addition, subtraction, multiplication, division, and square roots suffice*[10] *to describe every constructible number.*

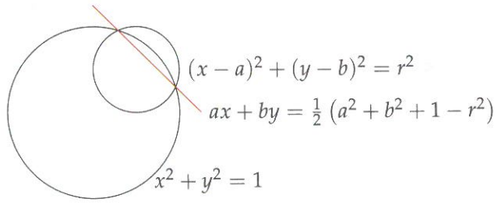
PROOF: When $ax + by = e$ and $cx + dy = f$ describe two lines, their intersection point[11] is given by the formula below.



When $ax + by = e$ describes a line, and $x^2 + y^2 = r^2$ describes a circle, the intersection point(s)[12] are described by the formula below.



When $x^2 + y^2 = 1$ and $(x-a)^2 + (y-b)^2 = r^2$ describe two circles, the line through their intersection points (if they intersect[13]) is given by the formula below:



Points that arise from straightedge-compass constructions arise from these three kinds of intersections, together with shifting and scaling.[14] Only addition, subtraction, multiplication, division, and square roots are ever needed to express coordinates of constructible points. By the Pythagorean theorem, only these operations are needed to express distances between constructible points.

[10] The previous page demonstrated that numbers obtained by these operations were constructible; this page demonstrates that all constructible numbers can be given by these operations.

[11] The lines are parallel if and only if $ad - bc = 0$, in which case the formula for their intersection point does not make sense.

[12] If the quantity under the square root is positive, then there are two intersection points; if zero, there is one intersection point; if negative, the line and circle do not intersect.

[13] If the circles do not intersect, the line still makes sense but passes between the circles rather than through them. When the two circles intersect, their intersection points may be found as the intersection of the line with either circle, using the formula from the previous figure.

[14] Shifting corresponds to addition and subtraction of coordinates, and scaling corresponds to multiplication of coordinates.

Our algebraic formulae for intersections of lines and circles yields a complete description of the constructible numbers. Intersecting lines and circles also yields an old solution to an older problem.

Theorem 3.4 (Infinitude of Pythagorean triples) *There are infinitely many primitive*[15] ***Pythagorean triples.*** *In other words, there are infinitely many triples (a, b, c) of integers, for which $\mathrm{GCD}(a, b, c) = 1$ and $a^2 + b^2 = c^2$.*

PROOF: Consider the unit circle centered at the origin, and a line through $(0, 1)$ of nonzero rational slope m.

The line intersects the circle at two points; $(0, 1)$ and (u, v) with $u \neq 0$. As $u^2 + v^2 = 1$ and $v = 1 + mu$, substituting yields

$$u^2 + (1 + mu)^2 = 1.$$

Expanding and simplifying yields $u \cdot \left((1 + m^2)u + 2m\right) = 0$. As $u \neq 0$, we find

$$u = -\frac{2m}{1 + m^2}, \quad v = \frac{1 - m^2}{1 + m^2}.$$

In this way, every nonzero rational number m gives a pair (u, v) of rational numbers such that $u^2 + v^2 = 1$. Let a/c be the reduced expression for u and b/d the reduced expression for v. Then

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{d}\right)^2 = 1, \quad \mathrm{GCD}(a, c) = \mathrm{GCD}(b, d) = 1, \quad c > 0, d > 0.$$

Expanding and multiplying through by cd yields

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

The two out of three principle implies $c^2 \mid a^2d^2$ and $d^2 \mid b^2c^2$. By Euclid's Lemma (Lemma 2.12), $c^2 \mid d^2$ and $d^2 \mid c^2$, and so $c = d$ (as both are positive). We may now write the equality $x^2 + y^2 = 1$ instead in the form

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1, \quad \mathrm{GCD}(a, c) = \mathrm{GCD}(b, c) = 1, \quad c > 0.$$

Multiplying through by c^2 yields a primitive Pythagorean triple,

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

The steps in this process may be reversed easily enough,[16] and this sets up a one to one correspondence between two sets:

- nonzero rational numbers (the slopes m);
- primitive Pythagorean triples (a, b, c) with $c > 0$ and $a \neq 0$.

Since the first set is infinite, so too is the second. ∎

[15] Without the primitive condition, $\mathrm{GCD}(a, b, c) = 1$, the theorem is not so interesting. Beginning with a Pythagorean triple like $(3, 4, 5)$, one can scale to form infinitely many nonprimitive triples, like $(6, 8, 10)$ and $(9, 12, 15)$ and $(12, 16, 20)$, etc.

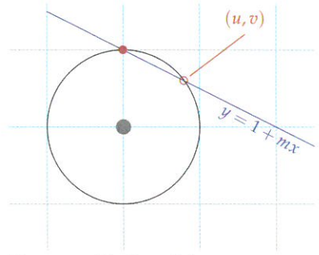


Figure 3.4: The line of slope

$$m = -1/2$$

intersects the circle at the point

$$(u, v) = \left(\frac{4}{5}, \frac{3}{5}\right).$$

As these fractions are reduced, they directly yield the primitive Pythagorean triple

$$(a, b, c) = (4, 3, 5).$$

In this two-step process, every nonzero rational slope yields a primitive Pythagorean triple.

[16] Begin with a primitive Pythagorean triple (a, b, c) and let $u = a/b$ and let $v = b/c$. Then draw a line from $(0, 1)$ to (u, v); the slope is certainly rational.

WHICH constructible numbers are rational? From Euclid's perspective,[17] this is a question of commensurability. Two lengths x and y are said to be **commensurable** if there exists a length z and integers m and n, such that $x = mz$ and $y = nz$. In other words, two lengths are commensurable if they can be both be measured precisely by the same measuring rod.

The figure in the margin can be used to demonstrate that 1 and $\sqrt{2}$ are incommensurable. Or geometrically, the leg and the hypotenuse of a 45-45-90 triangle are incommensurable. Indeed, any measuring rod that can measure both leg ℓ and hypotenuse h can measure the differences $h - \ell$ and $2\ell - h$. These are the leg and hypotenuse of a smaller 45-45-90 triangle. Repeating this, the measuring rod must would be able to measure the sides of smaller and smaller triangles, ad infinitum. This is a contradiction – the triangles must eventually be smaller than the measuring rod itself!

In modern times, commensurability is usually rephrased in terms of rationality. Namely, if $x = mz$ and $y = nz$ for two positive integers m and n, then $x/y = mz/nz = m/n$ is a rational number. Hence commensurable lengths are those with rational quotient.

Square, cube, and higher roots[18] are irrational, except when they are almost obviously rational. Reduction of fractions provides a powerful tool to prove irrationality; this has deprecated Euclid's geometric proofs of incommensurability.

Proposition 3.5 (Irrationality of surds) *Let a/b be a reduced fraction (with $b > 0$). Then $\sqrt[n]{a/b}$ is a rational number if and only if a and b are n^{th} powers of integers.*

PROOF: If $\sqrt[n]{a/b}$ is a rational number, then $\sqrt[n]{a/b} = c/d$ for some reduced fraction c/d. It follows that $a/b = (c/d)^n = c^n/d^n$. Since $\mathrm{GCD}(c,d) = 1$, Corollary 2.25 implies $\mathrm{GCD}(c^n, d^n) = 1$. Therefore, the fraction c^n/d^n is reduced. By the uniqueness of reduced representatives, the equality $a/b = c^n/d^n$ implies that $a = c^n$ and $b = d^n$. Hence a and b are n^{th} powers of integers.

Conversely, if a and b are n^{th} powers of integers, then $a = c^n$ and $b = d^n$ for some integers c and d. Hence $a/b = c^n/d^n = (c/d)^n$, so $\sqrt[n]{a/b} = c/d$ is a rational number. ∎

Problem 3.6 Demonstrate that $\sqrt{17}$ is irrational.

SOLUTION: Since $17 = 17/1$, a reduced fraction, and 17 is not a square number (16 and 25 are squares of 4 and 5, and no square number lies between 16 and 25), Proposition 3.5 implies that $\sqrt{17/1}$ is irrational. ✓

[17] Book X of the *Elements*, the longest of the 13 books, studies questions of commensurability.

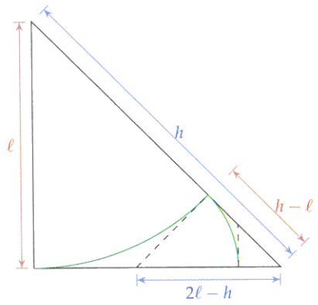


Figure 3.5: A 45-45-90 triangle. The measurement of the length $2\ell - h$ can be derived from the Pythagorean theorem: If $h^2 = 2\ell^2$ then $(2\ell - h)^2 = 2(h - \ell)^2$. The incommensurability of a leg with the hypotenuse is equivalent to the irrationality of the quotient $\sqrt{2}/1 = \sqrt{2}$.

[18] We follow the convention that $\sqrt[n]{x}$ means the *positive* n^{th} root of x, when n is even and x is positive. When n is odd, every real number x has a unique real n^{th} root. When n is even and x is negative, the number x has no real n^{th} root.

Problem 3.7 Are the numbers $\sqrt[3]{2}$ and $3\sqrt[4]{3}$ commensurable?

SOLUTION: Two numbers are commensurable if their quotient is rational. The quotient is $\sqrt[3]{2}/3\sqrt[4]{3}$, and

$$\frac{\sqrt[3]{2}}{3\sqrt[4]{3}} = \frac{1}{3}\frac{\sqrt[12]{16}}{\sqrt[12]{27}} = \frac{1}{3}\sqrt[12]{\frac{16}{27}}.$$

Since $16/27$ is a reduced fraction, and neither 16 nor 27 is a perfect twelfth power, we find that $\sqrt[12]{16/27}$ is irrational. Hence $\sqrt[3]{2}$ and $3\sqrt[4]{3}$ are not commensurable. ✓

What about more complicated constructible numbers? Numbers like $\sqrt{8 - \sqrt{7}}$, for example? For **algebraic**[19] numbers, the easiest irrationality proofs use the rational root theorem.

Theorem 3.8 (Rational Root Theorem) *Let a/b be a reduced fraction, representing a rational number x. Let $c_0, \ldots, c_d$ be positive integers. Then*

$$c_0 + c_1 x + \cdots + c_{d-1}x^{d-1} + c_d x^d = 0$$

implies $a \mid c_0$ and $b \mid c_d$.

PROOF: Noting that $x = a/b$, substitute to find

$$c_0 + c_1\frac{a}{b} + \cdots + c_{d-1}\frac{a^{d-1}}{b^{d-1}} + c_d\frac{a^d}{b^d} = 0.$$

Multiply through by b^d to obtain

$$c_0 b^d + c_1 a b^{d-1} + \cdots + c_{d-1}a^{d-1}b + c_d a^d = 0.$$

Observe that a divides all terms in the red box; hence[20] a divides the remaining term $c_0 b^d$. Similarly, b divides all terms in the blue box; hence b divides the remaining term $c_d a^d$. We have found

$$a \mid c_0 b^d \text{ and } b \mid c_d a^d.$$

Since $\mathrm{GCD}(a,b) = 1$, Corollary 2.25 implies $\mathrm{GCD}(b, a^d) = 1$ and $\mathrm{GCD}(a, b^d) = 1$, and Euclid's Lemma (Lemma 2.12) implies

$$a \mid c_0 \text{ and } b \mid c_d.$$

∎

From this theorem, one may hunt for rational roots to a polynomial systematically. There are an infinite number of rational numbers, but one needs only to search the divisors of the coefficients c_0 and c_d in order to find the numerators and denominators of rational roots.

[19] An algebraic number is a number x which is a root of a nonzero polynomial with integer coefficients. For example, $\sqrt{2}$ is algebraic, since $x = \sqrt{2}$ satisfies $x^2 - 2 = 0$. All rational numbers are algebraic, since if $x = a/b$ is a reduced fraction, then $bx - a = 0$. All constructible numbers are algebraic, but there are many more algebraic numbers. For example, the real number $\sqrt[5]{13}$ is algebraic (since $x = \sqrt[5]{13}$ satisfies $x^5 - 13 = 0$) but not constructible. Some algebraic numbers are hard to describe without giving a polynomial; for example, if x is the unique real number satisfying $x^5 - 2x + 2 = 0$, then there is no better way to describe the algebraic number x. Non-algebraic numbers are called **transcendental**. Famous transcendental numbers include e, π, and $\log(2)$.

[20] Here we use the two-out-of-three principle, when one term is zero, to give a one-out-of-two principle: if $x + y = 0$ and $a \mid y$ then $a \mid x$.

We demonstrate how to use the rational root theorem to prove the irrationality of many numbers. Let us first reinterpret the rational root theorem a bit. As stated in Theorem 3.8, if x is a rational number, $x = a/b$ as a reduced fraction, and

$$c_0 + c_1 x + c_2 x^2 + \cdots + c_d x^d = 0, \tag{3.1}$$

then $a \mid c_0$ and $b \mid c_d$. If no such rational number exists, satisfying Equation (3.1) and these divisibility conditions, then any number satisfying Equation (3.1) must be irrational.

Problem 3.9 Prove that $\sqrt{2}+\sqrt{3}$ is irrational.

SOLUTION: Let $x = \sqrt{2}+\sqrt{3}$. The square of x may be computed

$$x^2 = (\sqrt{2}+\sqrt{3})^2 = 2 + 2\sqrt{2}\sqrt{3} + 3 = 5 + 2\sqrt{6}.$$

The fourth power of x may be computed:

$$x^4 = x^2 \cdot x^2 = (5+2\sqrt{6})^2 = 25 + 20\sqrt{6} + 24 = 49 + 20\sqrt{6}.$$

Now observe that one can "cancel" the $\sqrt{6}$ terms by subtracting suitable multiples:

$$x^4 - 10x^2 = 49 + 20\sqrt{6} - 10(5+2\sqrt{6}) = 49 + 20\sqrt{6} - 50 - 20\sqrt{6} = -1.$$

We have found that if $x = \sqrt{2}+\sqrt{3}$ then

$$x^4 - 10x^2 + 1 = 0.$$

This allows us to apply the rational root theorem.

For if $x = a/b$ were a rational number, with a/b a reduced fraction, then $a \mid 1$ and $b \mid 1$. This would imply that $x = 1$ or $x = -1$. But neither of these numbers satisfies the polynomial equation $x^4 - 10x^2 + 1 = 0$. Hence x cannot be a rational number. ✓

This problem might make the reader think something like "if x doesn't look like a rational number, then x isn't a rational number." But not all rational ducks quack like a rational duck. Here is a very unusual-looking number, mentioned first by Daniel Shanks:[21]

$$\sqrt{5+\sqrt{22+2\sqrt{5}}-\sqrt{11+2\sqrt{29}}-\sqrt{16-2\sqrt{29}+2\sqrt{55-10\sqrt{29}}}}.$$

This beast of a constructible number certainly looks irrational. But it is a most rational number; it equals zero.

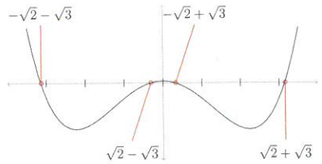


Figure 3.6: The graph of $y = x^4 - 10x^2 + 1$ intersects the x-axis at four points. The rightmost point is $\sqrt{2}+\sqrt{3}$. The rational root theorem shows that all four intersection points $\pm\sqrt{2}\pm\sqrt{3}$ are irrational, since they do not equal ± 1.

[21] See "Incredible Identities," by D. Shanks, in the *Fibonacci Quarterly*, **12**, pp. 271–281 (1974).

Before returning to rational numbers, we give one more connection between straightedge-compass constructions and number theory.

Proposition I.1, of the *Elements* constructs an equilateral triangle. But for another approach, begin with a unit length 1 and some other length x, less than 1. Then try to construct a polygon, by tracing chords inside a unit circle, at distance x from the center:

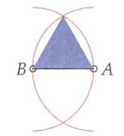


Figure 3.7: Euclid's construction of an equilateral triangle on the segment AB.

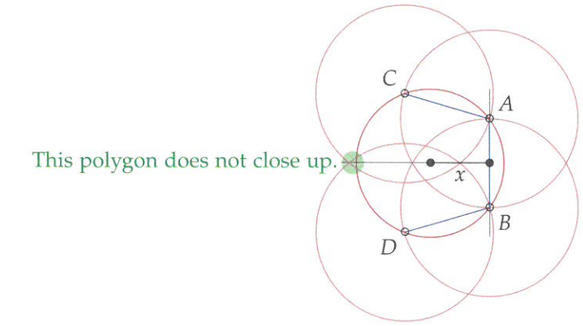


Figure 3.8: Begin with a unit circle, and draw a chord at distance x from the circle's center; this chord will intersect the circle at two points, A and B. Now draw circles with centers A and B, sharing the same radius AB. These circles will intersect the unit circle at new points C and D. Connect AC and BD. Carry on in this manner, inscribing a polygon in the unit circle. If you get lucky with your choice of initial length x, the polygon will close up perfectly.

If you begin with $x = \cos(\pi/n)$, then this construction "closes up" to give a regular[22] polygon with n sides. It follows that

Proposition 3.10 *If n is an integer, $n \geq 3$, and $\cos(\pi/n)$ is a constructible number, then one can inscribe a regular n-gon in a unit circle with only straightedge and compass.*

Not only did Euclid construct a regular triangle in his *Elements*. In Proposition IV.6, Euclid inscribes a square in a given circle, and by Proposition IV.11, Euclid inscribes a regular pentagon. In light of the proposition above, the inscriptions of regular triangle, quadrilateral, and pentagon are possible because $\cos(\pi/3)$ and $\cos(\pi/4)$ and $\cos(\pi/5)$ are constructible numbers:

$$\cos\left(\frac{\pi}{3}\right) = \frac{1}{2}, \quad \cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}, \quad \cos\left(\frac{\pi}{5}\right) = \frac{1+\sqrt{5}}{4}.$$

By March 30, 1796, the almost-19-year-old Gauss wrote in his journal of a construction of the 17-gon. The relevant trigonometric identity appears in his *Disquisitiones*. The 17-gon is constructible, because $\cos(\pi/17)$ equals an irrational but constructible number:

$$\frac{1}{8}\sqrt{30+2\sqrt{17}+2\sqrt{2}\left(\sqrt{34+6\sqrt{17}+\sqrt{2}(\sqrt{17}-1)\sqrt{17-\sqrt{17}}-8\sqrt{2}\sqrt{17+\sqrt{17}}}+\sqrt{17-\sqrt{17}}\right)}.$$

The reader may be happy to return immediately to rational numbers.

[22] Here **regular** means equilateral (all sides have the same length) and equiangular (all angles have the same measure). For triangles, equilateral is equivalent to equiangular. But for quadrilaterals, note that all rectangles are equiangular, and all rhombi are equilateral. Only the squares are regular quadrilaterals.

A **Fermat prime** is a prime number p of the form $2^{2^n}+1$. The first few Fermat primes are

$$2^{2^0}+1 = 3, \quad 2^{2^1}+1 = 5, \quad 2^{2^2}+1 = 17.$$

Gauss's method generalizes to prove constructibility of a p-gon, whenever p is a Fermat prime. Unfortunately, there are only five known Fermat primes:

$$3, 5, 17, 257, 65537.$$

THE WRONG WAY to add fractions is to add the numerators and add the denominators. So we use a new symbol for this operation:

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

Critical is the fact that this is an operation on *fractions* and not on *rational numbers*; the effect of this operation depends on whether the fractions are reduced or not. For example,

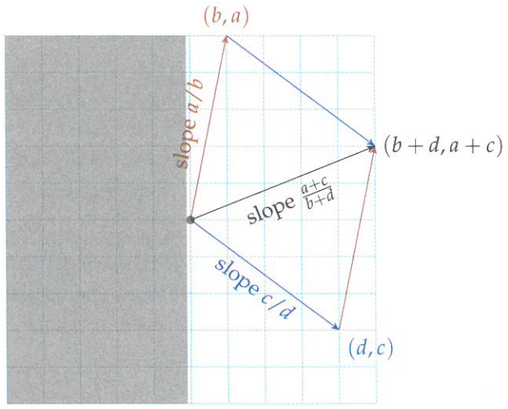
$$\frac{2}{4} \vee \frac{1}{3} = \frac{3}{7} \text{ and } \frac{1}{2} \vee \frac{1}{3} = \frac{2}{5}.$$

Though $2/4 = 1/2$, the results $3/7$ and $2/5$ are different rational numbers! This is just one sign that $\vee$ cannot be addition.

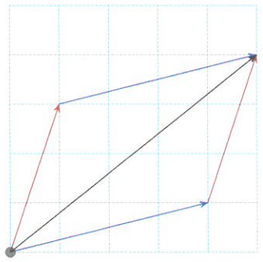
This operation is called the **mediant**. We say that $2/5$ is the mediant fraction of $1/2$ and $1/3$. A geometric interpretation explains the name.

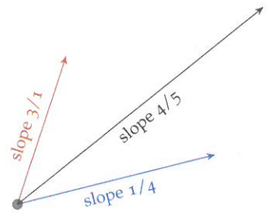
Proposition 3.11 (Mediant fractions lie between) *Let a/b and c/d be fractions with $b > 0$ and $d > 0$. Then the mediant $(a/b) \vee (c/d)$ represents a rational number that lies between a/b and c/d.*

PROOF: Since $b > 0$ and $d > 0$, $b + d > 0$. It follows that $(a/b) \vee (c/d) = (a+c)/(b+d)$ has positive denominator; it represents a rational number. To see that the mediant lies between a/b and c/d, we view the rational numbers as slopes in the right half[23] of the plane.

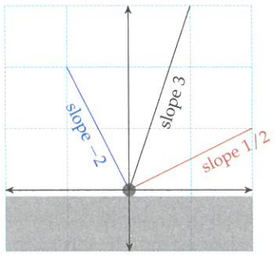


The diagonal black line lies between the blue and red edges in the right half-plane; hence the slope of the black line is between the slopes of the blue and red edges. Thus the rational number $(a/b) \vee (c/d)$ is between (a/b) and (c/d). ■





[23] It is crucial that we work in the right half of the plane for this geometric reasoning. "Betweenness" of vectors in the right half of the plane reflects betweenness of slopes. If, instead, we worked in the top half of the plane, this would no longer be true. Consider the figure below:



Certainly, the black line is between the red and blue. But 3 is not between $1/2$ and -2. The slope of a line does not vary continuously, as a line rotates past vertical.

Problem 3.12 Which is greater, $2/5$ or $3/7$?

SOLUTION: Observe that

$$\frac{3}{7} = \frac{2}{5} \vee \frac{1}{2}.$$

Hence $3/7$ is between $2/5$ and $1/2$. Since $2/5 < 1/2$, we find that

$$\frac{2}{5} < \frac{3}{7} < \frac{1}{2}.$$ ✓

This method is quite powerful, even with larger fractions.

Problem 3.13 Which is greater, $11/17$ or $17/30$?

SOLUTION: Observe the following mediant identity:

$$\frac{17}{30} = \frac{11}{17} \vee \frac{6}{13}.$$

Since $6/13$ is less than a half, and $11/17$ is more than a half, we find that

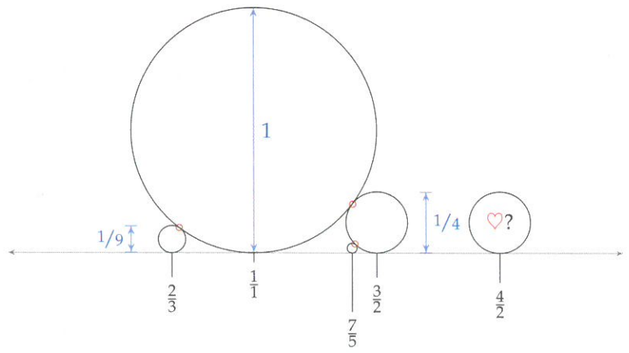
$$\frac{6}{13} < \frac{17}{30} < \frac{11}{17}.$$ ✓

The mediant fraction lies between two given reduced fractions. But *where* between them? The mediant is rarely the mean; it is not simply the average of two rational numbers.[24] To locate the mediant, we introduce **kissing fractions**.

Definition 3.14 Let a/b and c/d be fractions. We say a/b kisses c/d if $ad - bc = \pm 1$. For notation, we denote kissing[25] with a heart:

$$\frac{a}{b} \heartsuit \frac{c}{d} \text{ means that } ad - bc = 1 \text{ or } ad - bc = -1.$$

Each fraction has its place on the number line. To see the fractions kiss, you must draw their Ford circles. Lying atop the number line at a/b, the **Ford circle** is the circle of diameter $1/b^2$.



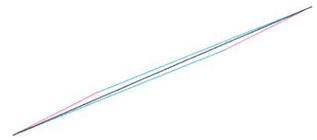


Figure 3.9: The deviations between slopes of $1/2$, $2/5$, and $3/7$ are almost indiscernable; but still, $3/7$ is between $1/2$ and $2/5$.

[24] It could not be an honest operation on rationals like the mean, since it is not an operation on rational numbers at all!

[25] Note that if $\frac{a}{b} \heartsuit \frac{c}{d}$ then $\frac{c}{d} \heartsuit \frac{a}{b}$. The equality $ad - bc = \pm 1$ is equivalent to the equality $bc - ad = \mp 1$. The heart symbol is symmetric, just like the kissing relation.

Figure 3.10: The Ford circles on top of $1/1$, $2/3$, $7/5$, $3/2$, and $7/3$. Their diameters are 1, $1/9$, $1/25$, $1/4$, and $1/9$, respectively. The circles *osculate*, or *kiss*, at their point of tangency. Observe three kisses:

$$\frac{2}{3} \heartsuit \frac{1}{1}.$$
$$\frac{1}{1} \heartsuit \frac{3}{2}.$$
$$\frac{3}{2} \heartsuit \frac{7}{5}.$$

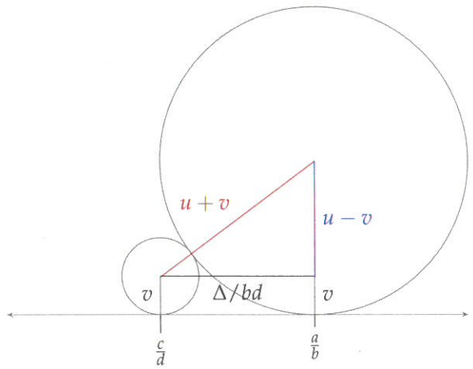
FRACTIONS KISS precisely when their Ford circles are tangent.

Theorem 3.15 (Kissing fractions have tangent Ford circles) *Let a/b and c/d be fractions with nonzero denominators. Then a/b kisses c/d if and only if the Ford circle atop a/b is tangent to the Ford circle atop c/d.*

PROOF: Let $\Delta = ad - bc$; thus $\Delta = \pm 1$ if and only if a/b kisses c/d. The number Δ occurs naturally when subtracting fractions:

$$\frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd} = \frac{\Delta}{bd}.$$

Draw the Ford circles atop a/b and c/d.



Let $u = 1/2b^2$ and $v = 1/2d^2$, the radii of the two circles. The distance between their centers, by the Pythagorean theorem, equals

$$\sqrt{\left(\frac{\Delta}{bd}\right)^2 + (u - v)^2}.$$

The circles are tangent if and only if the distance between their centers is the sum of the radii $(u + v)$, as pictured. Hence the circles are tangent if and only if

$$\left(\frac{\Delta}{bd}\right)^2 + (u - v)^2 = (u + v)^2.$$

This is equivalent to

$$\frac{\Delta^2}{b^2d^2} = (u + v)^2 - (u - v)^2 = 4uv.$$

But notice that $4uv = 1/b^2d^2$. So the circles are tangent if and only if

$$\frac{\Delta^2}{b^2d^2} = \frac{1}{b^2d^2}.$$

Hence the circles are tangent if and only if $\Delta^2 = 1$, if and only if $\Delta = \pm 1$, if and only if the fractions kiss. ∎

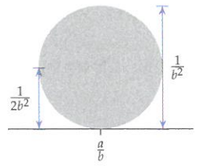


Figure 3.11: The Ford circle atop a/b has diameter $1/b^2$ and radius $1/2b^2$.

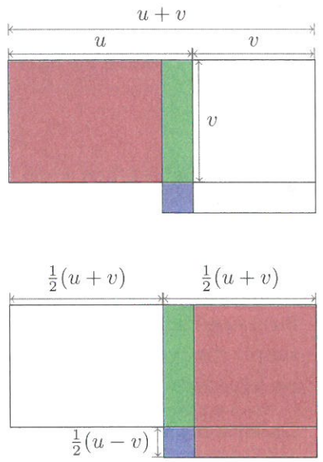


Figure 3.12: Based on Euclid, Proposition II.5, and related figures (unpublished) by W. Casselman. This figure demonstrates the identity

$$\left(\frac{u+v}{2}\right)^2 - \left(\frac{u-v}{2}\right)^2 = uv.$$

The left side corresponds to all colored regions except for the blue, in the bottom figure. The right side comes from arranging the red and green regions as in the top figure. Multiplying through by 4 gives the identity

$$(u + v)^2 - (u - v)^2 = 4uv$$

used in this proof.

KISSING FRACTIONS lead inevitably to more kissing. The following result demonstrates that when two fractions kiss, the mediant fraction gets in on the action.

Proposition 3.16 *Let a/b and c/d be two kissing fractions. Then the mediant $(a/b) \vee (c/d)$ kisses both a/b and c/d.*

PROOF: The proof is algebraic, beginning with the kissing assumption: $ad - bc = \pm 1$. A fortuitous cancellation yields

$$(a + c)b - (b + d)a = ab + cb - ba - da = cb - da = -(ad - bc) = \mp 1.$$

The mediant is $(a/b) \vee (c/d) = (a + c)/(b + d)$, and the above computation implies

$$\frac{a+c}{b+d} \heartsuit \frac{a}{b}.$$

The same computation, reversing the roles of a/b and c/d, implies that

$$\frac{a+c}{b+d} \heartsuit \frac{c}{d}.$$

∎

The geometric consequence is the following: if one begins with two kissing fractions, and continues production of fractions by mediants, then the resulting Ford circles are squeezed tight.

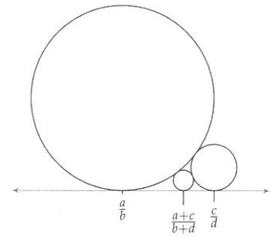


Figure 3.13: Given $(a/b) \heartsuit (c/d)$, we find that

$$\frac{a}{b} \heartsuit \frac{a+c}{b+d} \heartsuit \frac{c}{d}.$$

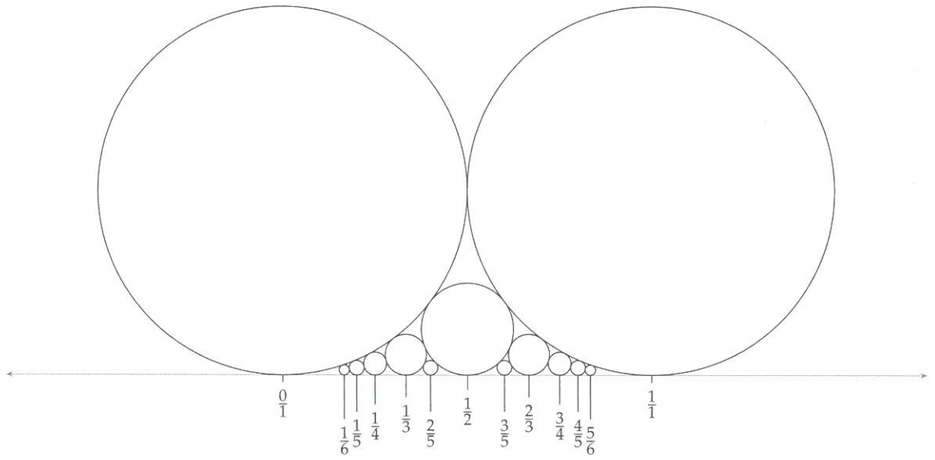


Figure 3.14: Mediants, kissing fractions and tangent Ford circles.

Begin with the kissing fractions $0/1$ and $1/1$. Their mediant, $1/2$ kisses both, and therefore the Ford circle on $1/2$ squeezes perfectly between the two large circles. The mediant of $0/1$ and $1/2$ is $1/3$. The Ford circle on $1/3$ squeezes perfectly between the circle at $1/2$ and the circle at $0/1$.

Which fractions kiss? The kissing equation $ad - bc = \pm 1$ requires $\mathrm{GCD}(a,b) = 1$ and $\mathrm{GCD}(c,d) = 1$. Indeed, if $g \mid a$ and $g \mid b$, then $g \mid ad - bc$, so $g \mid \pm 1$, so $g = \pm 1$. Hence $\mathrm{GCD}(a,b) = 1$, and the same argument yields $\mathrm{GCD}(c,d) = 1$. Nonreduced fractions cannot kiss.

Proposition 3.17 *If $\mathrm{GCD}(a,b) \neq 1$, then a/b kisses no other fractions.*

In contrast, reduced fractions find many mates.

Proposition 3.18 *Let a/b be a reduced fraction. Then a/b kisses infinitely many fractions. If $b > 1$ then among the reduced fractions kissing a/b, there are exactly two with denominator smaller than b. These two fractions kiss each other and have mediant a/b.*

Proof: Since $\mathrm{GCD}(a,b) = 1$, there exists a solution (x_0, y_0) to the linear Diophantine equation

$$ay - bx = 1.$$

Moreover, by Theorem 1.21, there are infinitely many solutions,[26] one for every integer n:

$$y = y_0 + bn, \quad x = x_0 + an.$$

Each such solution (y, x) gives a fraction x/y which kisses a/b.

If $b > 1$, then among the numbers $y_0 + bn$ (as n varies over all integers), exactly two are smaller than b in absolute value.

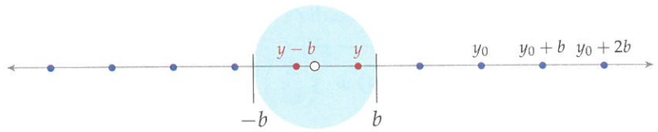


Figure 3.15: The numbers $y_0 + bn$ are highlighted with dots on the number line, spaced at intervals of length b. Zero is not among the highlighted numbers since $y = 0$ and $ay - bx = 1$ implies $bx = 1$, but $b > 1$, a contradiction. Thus exactly two dots can fit into the circle of radius b.

Among the two, call the positive number y, so its negative neighbor is $y - b$. These give two reduced fractions

$$\frac{x}{y} \heartsuit \frac{a}{b}, \quad \frac{a-x}{b-y} \heartsuit \frac{a}{b}.$$

The fractions (x/y) and $(a-x)/(b-y)$ kiss (a direct algebraic computation suffices), and their mediant is a/b. ∎

This Theorem has a geometric interpretation using Ford circles.

Corollary 3.19 *Let a/b be a reduced fraction. Then the Ford circle atop a/b is tangent to infinitely many other Ford circles; if $b > 1$, then two of these have larger diameters and kiss each other.*

As the proof of the Theorem uses linear Diophantine equations, we can find kissing fractions by using the Euclidean algorithm.

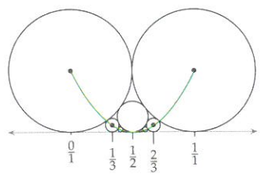


Figure 3.16: Infinitely many fractions kiss the reduced fraction 1/2, among which two are bigger. Centers of their Ford circles lie on the parabola

$$y = 2 \cdot (x - 1/2)^2.$$

[26] The signs here are slightly different from Theorem 1.21, but the result follows directly if one keeps track of signs carefully.

Problem 3.20 Find a fraction which kisses 71/83.

Solution: We look for integers x, y such that $(x/y) \heartsuit (71/83)$; this is equivalent to the statement $71y - 83x = \pm 1$. To solve this Diophantine equation, we use the Euclidean algorithm.

$$\begin{aligned} 83 &= 1(71) + 12 \\ 71 &= 5(12) + 11 \\ 12 &= 1(11) + 1. \end{aligned}$$

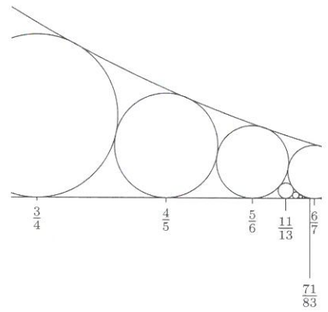
Solving for 1 (review Problem 1.15) yields

$$\begin{aligned} 1 &= 12 - 1(11) \\ &= 12 - 1(71 - 5(12)) = 6(12) - 1(71) \\ &= 6(83 - 1(71)) - 1(71) \\ &= 6(83) - 7(71). \end{aligned}$$

We have found that $71(-7) - 83(-6) = 1$, so $71(7) - 83(6) = -1$, and

$$\frac{71}{83} \heartsuit \frac{6}{7}. \quad \checkmark$$

By finding kissing fractions, of smaller and smaller denominators, we can work backwards until the denominator equals 1.



Why does 71/83 occur? First, note that it is a mediant:

$$\frac{71}{83} = \frac{6}{7} \vee \frac{65}{76}.$$

These are both mediants of fractions with smaller denominators:

$$\frac{6}{7} = \frac{1}{1} \vee \frac{5}{6}, \quad \frac{65}{76} = \frac{6}{7} \vee \frac{59}{69}.$$
$$\frac{5}{6} = \frac{1}{1} \vee \frac{4}{5}, \quad \frac{59}{69} = \frac{6}{7} \vee \frac{53}{62}.$$
$$\frac{4}{5} = \frac{1}{1} \vee \frac{3}{4}, \quad \frac{53}{62} = \frac{6}{7} \vee \frac{47}{55}.$$
$$\frac{3}{4} = \frac{1}{1} \vee \frac{2}{3}, \quad \frac{47}{55} = \frac{6}{7} \vee \frac{41}{48}.$$
$$\frac{2}{3} = \frac{1}{1} \vee \frac{1}{2}, \quad \frac{41}{48} = \frac{6}{7} \vee \frac{35}{41}.$$
$$\frac{1}{2} = \frac{1}{1} \vee \frac{0}{1}, \quad \frac{35}{45} = \frac{6}{7} \vee \frac{29}{34}.$$
$$\frac{29}{34} = \frac{6}{7} \vee \frac{23}{27}, \quad \frac{23}{27} = \frac{6}{7} \vee \frac{17}{20}.$$
$$\frac{17}{20} = \frac{6}{7} \vee \frac{11}{13}, \quad \frac{11}{13} = \frac{6}{7} \vee \frac{5}{6}.$$

These facts can be used as directions, to begin with Ford circles on 0/1 and 1/1, and nest down to the Ford circle on top of 71/83.

Theorem 3.21 (Integers generate all reduced fractions via mediants) *Begin with the integer fractions $n/1$ for every integer n. Whenever fractions kiss, take their mediants, and continue this process indefinitely. Only reduced fractions will occur, and all reduced fractions will eventually occur.*

Proof: Note first that only fractions with positive denominators will occur; the mediant of two fractions with positive denominator is another fraction with positive denominator. Moreover, since mediant fractions get in on the action, and kissing fractions (with positive denominators) are always reduced, this process *only* yields reduced fractions.

To see that *all* reduced fractions eventually occur, we begin with a reduced fraction a/b. If $b = 1$, then $a/b = a/1$ occurred at the beginning of the process. Otherwise, the previous Theorem implies that a/b is the mediant of two reduced fractions x/y and $(a-x)/(b-y)$ with smaller denominator. Repeat the process on x/y and $(a-x)/(b-y)$ and so on, until all denominators equal 1. This traces the ancestry of a/b back to fractions of denominator 1. ∎

In this way, Ford circles and kissing fractions resolve the problem of locating rational numbers on the number line. Each rational number lands in its place when its Ford circle is carried on its back.

We approximate real numbers by rational numbers every time we write a decimal expansion. When we write $\pi \approx 3.14$, we are saying that π is approximately equal to the rational number $314/100$. But this rational approximation is not particularly good; the approximation $22/7 = 3.\overline{142857}$ is better, and for a much smaller *price*.

The **price**[27] of a fraction is the size of its denominator. So the fraction $314/100$ costs 100 dollars; even reducing it to $107/50$, it still costs 50 dollars. For 50 dollars, we can approximate π within an error of about 0.0016. But for only 7 dollars, we may purchase the fraction $22/7$, which approximates π within an error of about 0.0013 – a better approximation for lower price!

Diophantine approximation studies the question: given a real number x, what is the relationship between the *accuracy* of a rational approximation and its *cost*.

The first Diophantine approximation theorem follows from the geometry of the number line.

Proposition 3.22 *Let x be a real number. Let b be a positive integer. Then there exists a rational number a/b such that*

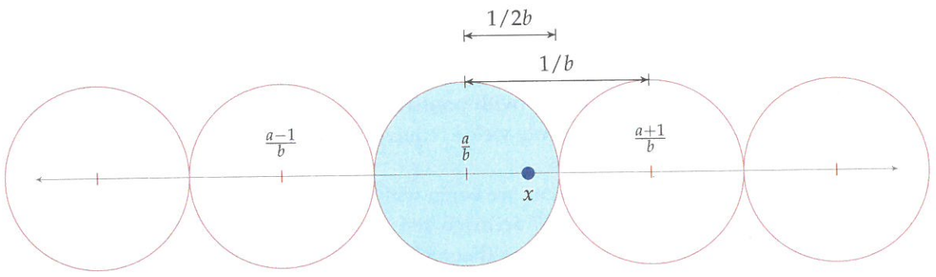
$$\left|x - \frac{a}{b}\right| \leq \frac{1}{2b}.$$

In other words, given b dollars, one can approximate every real number x within a distance of $1/2b$.

Proof: Plot x on the number line, and mark every fraction with denominator b. Surround each fraction with denominator b by a circle[28] of diameter $1/b$ to completely cover the number line by circles.

[27] Edward Burger taught me to think of the denominator of a fraction as a price or cost, in the context of Diophantine approximation. See "Making Transcendence Transparent", by E. Burger and R. Tubbs, Springer (2004) for much more.

[28] Intervals suffice but circles look better.



If a/b is the fraction whose circle contains x, then the distance between x and a/b is bounded by the radius of the circle.

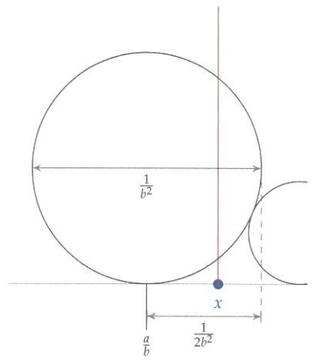
$$\left|x - \frac{a}{b}\right| \leq \frac{1}{2b}.$$

■

The previous theorem guarantees that b dollars buys an approximation within $1/2b$. Ford circles can give closer approximations, often for a better price.

Suppose that x is a real number which lies in the shadow of a Ford circle. If a/b is the rational number touching the Ford circle, then as the diameter of the Ford circle is $1/b^2$, we observe that

$$\left|x - \frac{a}{b}\right| < \frac{1}{2b^2}.$$

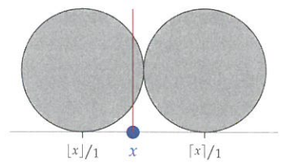


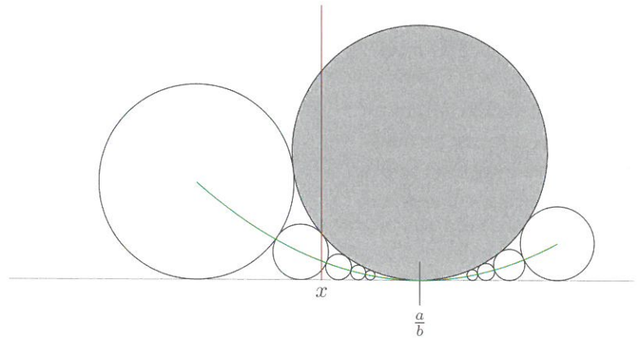
Theorem 3.23 (Dirichlet approximation theorem) *Let x be an* irrational *real number. Then there exist infinitely many reduced fractions a/b such that*

$$\left|x - \frac{a}{b}\right| < \frac{1}{2b^2}.$$

Proof: To demonstrate the theorem, it suffices to prove that x lies in the shadows of infinitely many Ford circles. To begin, x lies in the shadow of one Ford circle: either the Ford circle atop $\lfloor x \rfloor/1$ or the Ford circle atop $\lceil x \rceil/1$ overshadows x.

Now consider a Ford circle atop a/b whose shadow contains x, and all of the Ford circles tangent to it.





These Ford circles kissing a/b form an unbroken[29] chain, covering all points in the shadow, except a/b itself. As x is irrational, $x \neq a/b$, and the vertical line above x must pass through the chain.

Hence when x lies in the shadow of one Ford circle, it lies in the shadow of a smaller Ford circle; this continues indefinitely, to demonstrate that x lies in the shadows of infinitely many Ford circles. ■

[29] The chain is unbroken since each Ford circle is obtained from an adjacent Ford circle by taking its mediant with a/b. The x-coordinates of the centers along the chain approach a/b, since

$$\lim_{n \to \pm\infty} \frac{x_0 + an}{y_0 + bn} = \frac{a}{b}.$$

Hence all points, arbitrarily close to a/b, are covered by the chain.

Historical notes

This chapter owes much to *La Geometrie* (published in 1637) of René Descartes (1596 – 1650CE), which brought algebraic techniques to bear results on classical Greek geometry.[30] Our perspective differs from Descartes, as we place arithmetic in the center and geometry in its service. In Book 1 of his *Geometrie*, Descartes writes[31]

> And if it can be solved by ordinary geometry, that is, by the use of straight lines and circles traced on a plane surface, when the last equation shall have been entirely solved there will remain at most only the square of an unknown quantity, equal to the product of its root by some known quantity, increased or decreased by some other quantity also known.

Descartes states that, whatever Euclidean constructions are carried out, the resulting equations can be solved with nothing more the quadratic formula. In effect, our Theorem 3.3 is due to Descartes.

[30] See "The Geometry of René Descartes," translated from the French and Latin by David Eugene Smith and Marcia L. Latham, Dover Publications, Inc., New York, NY 1954.

[31] The French from 1637 reads "Et que si elle peut estre resolue par la Geometrie ordinaire, c'est a dire, en ne se seruant que de lignes droites & circulaires tracées fur une superficie plate, lorsque la derniere Equation aura esté entierement démeslée il n'y restera tout au plus qu'un quarré inconnu, esgal a ce qui se produist de l'Addition, ou soustraction de sa racine multipliée par quelque quantité connue, & de quelque autre quantité aussy connue."

PYTHAGOREAN TRIPLES occur whenever right triangles arise, e.g. in the construction of altars in Vedic India. Many authors[32] have suggested that a Babylonian tablet, Plimpton 322, is a table of Pythagorean triples. But Eleanor Robson has offered an alternative hypothesis,[33] based on contemporary mathematics of Mesopotamia such as reciprocal pairs and completing the square.

In Lemma 1 before Proposition X.29 of the *Elements*, Euclid gives a recipe to produce Pythagorean triples. Euclid's recipe can be modernized into an algebraic recipe: Begin with positive integers x and y, both even or both odd, with $x > y$; thus the difference $x - y$ and sum $x + y$ are both even. There is an equality of integers

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

If xy is a square number, then this gives a Pythagorean triple.[34]

Diophantus, in Book II, Problem 8 of his *Arithmetica*, demonstrates a similar recipe. Pierre de Fermat (c. 1601–1665) owned Bachet's Latin edition of the *Arithmetica*, and that is where Fermat wrote his famous marginal note:

> Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadatos & generaliter nullam in infinitum ultra quadratum potestatem in duos eius dem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.[35]

Here Fermat states that, to express a cube as two cubes or a fourth-power as two fourth-powers, and so on, is impossible. Fermat claims a wonderful proof of this impossibility, too small to fit in the margin.

[32] See "Sherlock Holmes in Babylon," by R. C. Buck, in the *American Mathematical Monthly* **87** (1980) pp. 335–345, which refers back to "The Exact Sciences in Antiquity," by O. Neugebauer (original, 1951) for two examples.

[33] See "Neither Sherlock Holmes nor Babylon: a reassessment of Plimpton 322," by E. Robson, in *Historia Mathematica* **28**, pp. 167–202 (2001).

[34] For example, choose $x = 9$ and $y = 1$, so that $xy = 3^2$ is a square as required. Then substitution yields $(x-y)/2 = 4$ and $(x+y)/2 = 5$, and the Pythagorean triple $3^2 + 4^2 = 5^2$. If one chooses $x = 32$ and $y = 2$, then $xy = 64 = 8^2$; substitution yields $(x-y)/2 = 15$ and $(x+y)/2 = 17$, and the Pythagorean triple $8^2 + 15^2 = 17^2$.

[35] Pierre de Fermat's marginal note was included in a 1670 edition of the *Arithmetica*, published by Fermat's son. Read "Fermat's Enigma," by Simon Singh, Anchor Books ed., New York (1998) for a good popular account of the history of the Last Theorem.

Fermat's Last Theorem refers to this marginal suggestion of Fermat. Here it is, stated in modern terms.

Theorem 3.24 (Fermat's Last Theorem) *If n is a positive integer, and[36] $n > 2$, then the equation $x^n + y^n = z^n$ has no solutions for which x, y, and z are all positive integers.*

Over 300 years, efforts to prove Fermat's conjecture led to outstanding discoveries in number theory, even as progress on the conjecture itself was incremental.[37]

Following suggestions of G. Frey (1986), and proving the ϵ-conjecture of J.P. Serre, K. Ribet (1990) demonstrated that a 1957 conjecture of Shimura and Taniyama implied Fermat's Last Theorem. By proving enough cases of Shimura and Taniyama's conjecture, Andew Wiles with Richard Taylor proved Fermat's Last Theorem in 1995.[38]

FORD CIRCLES are named for L.R. Ford, who first revealed[39] the connections between the circles and Diophantine approximation.

We have given only two results in Diophantine approximation of an irrational number x; the more powerful result guarantees the existence of infinitely many rational numbers a/b for which $|x - a/b| \le 1/2b^2$.

There are two ways one might try to improve this: by increasing the constant 2 or by increasing the exponent 2. Ford improved the constant from 2 to the best possible $\sqrt{5}$ (note that $\sqrt{5} \approx 2.236 > 2$):[40]

$$\left|x - \frac{a}{b}\right| \le \frac{1}{\sqrt{5}b^2}, \text{ for infinitely many rational numbers } \frac{a}{b}.$$

Improving the exponent in this estimate is more difficult. A series of results, beginning with Liouville (1844), through results of Thue and Siegel, culminated with Roth's proof[41] that the exponent cannot be improved for *algebraic* numbers:

Theorem 3.25 (Thue-Siegel-Roth Theorem) *Let x be an* algebraic *irrational number. Then for any positive number ϵ and constant C, the inequality*

$$\left|x - \frac{a}{b}\right| \le \frac{C}{b^{2+\epsilon}}$$

is only satisfied for finitely many fractions a/b.

In other words, one cannot improve the exponent, even slightly, from 2 to $2 + \epsilon$, at least not for *algebraic* numbers. This gives an effective way of proving transcendence of some numbers – if it is possible to approximate an irrational number extremely well by infinitely many rational numbers, then it must be transcendental!

[36] If $n = 2$, then the equation $x^2 + y^2 = z^2$ has infinitely many solutions in positive integers. They are the Pythagorean triples.

[37] Fermat proved the Theorem when $n = 4$. Euler proved the Theorem when $n = 3$, or at least his arguments can be fixed to make a proof in this case. Legendre and Dirichlet proved the Theorem when $n = 5$. Lamé and Lebesgue (*not* of Lebesgue measure fame) proved the Theorem when $n = 7$. The Theorem was divided into two cases later, and Germain proved the "first case" for odd primes up to 100.

[38] "Modular elliptic curves and Fermat's Last Theorem," by A. Wiles and "Ring theoretic properties of certain Hecke algebras," by R. Taylor and A. Wiles, in *Annals of Mathematics*, vol. 141 **3** (1995).

[39] "Fractions," by L. R. Ford, in *The American Mathematical Monthly*, vol. 45, **9** (Nov.,1938), pp. 586–601.

[40] The best-possible constant $\sqrt{5}$ was found earlier, by different methods, by A. Hurwitz, in "Über die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche," in *Mathematische Annalen*, vol. 39 (1891) pp. 279–284.

[41] "Rational approximations to algebraic numbers," by K. F. Roth, in *Mathematika*, vol. 2, **1** (1955), pp. 1–20. The story does not completely end with Roth. Instead of increasing the exponent from b^2 to $b^{2+\epsilon}$, on can try a slighter change from b^2 to $b^2 \log(b)^{1+\epsilon}$. This gives the conjecture of S. Lang (1927–2005)

Exercises

1. Beginning with a line segment of length 1, construct line segments of length 3, 1/3, 4/3, and $2\sqrt{3}/3$.

2. Prove that there are infinitely many positive integer triples (x, y, z) such that $x^2 + 2y^2 = 3z^2$. Hint: Follow the proof for Pythagorean triples, replacing the point $(0, 1)$ by the point $(1, 1)$.

3. Demonstrate that $\sqrt{5/7}$, $\sqrt{3} + \sqrt{5}$, and $\sqrt{8 - \sqrt{7}}$ are irrational.

4. (Challenge) Prove that if p and q are distinct prime numbers, then $\sqrt{p} + \sqrt{q}$ is irrational.

5. The number e is defined as the sum of the reciprocals of the factorials,[42] If e were rational, let n be its denominator when represented as a fraction, let x be the sum of the terms up to $1/n!$ and let y be the sum of the rest of the terms. Demonstrate in this case that $n! \cdot x$ is an integer and $n! \cdot e$ is an integer, and that $0 < n! \cdot y < 1$. Use this to achieve a contradiction, and fill in the steps to prove that e is irrational.

6.(a) Use the half-angle identity to prove that if an n-gon is constructible, then a $2n$-gon is constructible.

(b) Prove that if $\mathrm{GCD}(a, b) = 1$, and an a-gon is constructible and a b-gon is constructible, then an (ab)-gon is constructible. Hint: the equation $ax + by = 1$ implies $x \cdot \pi/a + y \cdot \pi/b = \pi/ab$. Use sum-difference identities for the cosine..

(c) Let e_2 be a natural number, and let $p_1, \ldots, p_s$ be distinct Fermat primes. Let $N = 2^{e_2} \cdot p_1 \cdots p_n$. Use the fact that p-gons are constructible for every Fermat prime p, to demonstrate that the N-gon is constructible.[43]

7. Use mediants to compare 13/21 and 17/27.

8. Consider the sequence of Fibonacci numbers, defined recursively by $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. Let r_n be the sequence of fractions obtained from consecutive Fibonacci numbers: $r_n = F_n/F_{n-1}$ for all $n \geq 1$. Thus r_n is the sequence of fractions:

$$\frac{1}{0}, \frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \ldots.$$

(a) Prove that each fraction in the sequence r_n is the mediant of the previous two fractions, and kisses the previous two fractions.

(b) Prove that every fraction in the sequence r_n is reduced.

(c) Draw a diagram of the Ford circles above the fractions in the sequence r_n. Prove any observations you can make.

[42] The convergent series defining e is

$$e = \frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots.$$

Hint for proving $n! \cdot y < 1$: the best way to bound a series is to compare it with a geometric series.

[43] More difficult is the following result: if p is a prime number, and $p \neq 2$ and p is *not* a Fermat prime, then the p-gon is *not* constructible. For example, the 7-gon is not constructible. The proof is a famous application of Galois theory.

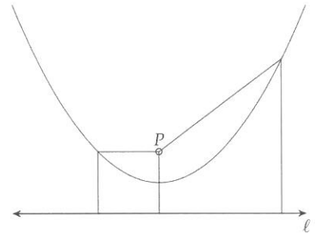
(d) Prove that the sequence r_n converges to some real number.

(e) Let $\phi = 1/2 \cdot (1 + \sqrt{5})$. This is called the **golden ratio**. Prove that $\lim_{n\to\infty} r_n = \phi$. Hint: squeeze ϕ between consecutive terms in the sequence.

9. Find fractions which kiss 7/13, 77/133, and 101/37.

10. Write a formula which describes all fractions which kiss 7/11.

11. If P is a point in the plane, and ℓ is a line that does not pass through P, then the set of points equidistant from P and ℓ forms a parabola. Prove that, given a Ford circle C, all centers of Ford circles adjacent to C lie on a parabola.



12. Prove that if Fermat's Last Theorem is true for all prime exponents p, and for the exponent 4, then Fermat's Last Theorem is true for all exponents $n \geq 3$.

13. Let x be a real number, and let n be a positive integer.[44]

(a) For each integer b between 1 and n, let $r_b = bx - \lfloor bx \rfloor$. The set $S = \{0, r_1, r_2, \ldots, r_n, 1\}$ contains $n + 2$ real numbers between 0 and 1. Prove that there are elements $s, t \in S$ such that

$$0 < t - s \leq \frac{1}{n+1}.$$

(b) Prove that there exist integers a, b such that $1 \leq b \leq n$ and

$$|bx - a| \leq \frac{1}{n+1}.$$

Hint: consider the previous part. Three cases are needed: $s = 0$, $t = r_b$, or $s = r_b$, $t = 1$, or $s = r_b$, $t = r_c$.

(c) Prove that if $|bx - a| \leq \frac{1}{n+1}$ and $b \leq n$, then

$$\left| x - \frac{a}{b} \right| < \frac{1}{b^2}.$$

Conclude that if x is a real number, there exist infinitely many fractions a/b satifying the above inequality.

[44] This exercise leads the reader through a proof of Dirichlet's Approximation Theorem (proven by P.G.L. Dirichlet, c.1840CE), slightly weaker than our result from Ford circles, but stronger than the naïve bound.

14. Use the Thue-Siegel-Roth theorem (Theorem 3.25) to prove that there are only finitely many integer solutions (x, y) to the equation

$$x^3 - 5y^3 = 100.$$

Hint: Prove[45] that $|\sqrt[3]{5 + x} - \sqrt[3]{5}| < 1/3 \cdot 5^{-2/3} x$ when $x > 0$. Use this to show that if (a, b) is any solution to the equation, then a/b is very close to $\sqrt[3]{5}$.

15. Use the Thue-Siegel-Roth theorem to give an example of a decimal expansion which represents a transcendental number.

[45] The graph of $y = \sqrt[3]{x}$ near $x = 5$ lies beneath its tangent lines; use the derivative at $x = 5$ to prove the estimate.