Continued fractions, Fermat, Euler, Lagrange

Introduction. Continued fractions are a natural way of expressing irrational numbers. Unlike decimal fractions, which depend on the choice of base ten, continued fractions are free of artificial choices. Thus patterns in the continued fraction expansions have a universal and deeper meaning. Beginning with exploratory calculations of continued fractions of roots of integers, we derive a celebrated theorem of Fermat that characterizes the prime numbers that can be written as sum of two squares. Generally, roots of quadratic polynomials play a particular role in the theory of continued fractions, as they are the ones that produce eventually periodic continued fractions. This is a pair of theorems by Euler and by Lagrange that we also explore.

Continued fractions. A continued fraction is an infinitely nested fraction

$$x = m_1 + \cfrac{1}{m_2 + \frac{1}{m_3 + \frac{1}{m_4 + \dots}}} \tag{1}$$

where

$$m_1, m_2, m_3, m_4, \dots \tag{2}$$

is a sequence of natural numbers. By natural number in this chapter we mean an integer greater than zero. The infinite nesting is somewhat intimidating. Even worse, the natural starting point for calculating a nested fraction is the deepest level of nesting, where at the displayed continued fraction we find nothing but dots. The difficulty persists, even if we have a very clear understanding of a particular sequence (2).

For example, consider the sequence

$$2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1, 12, \dots, \tag{3}$$

which except for the first miscellaneous terms is the sequence of even integers interlaced with double ones. To solve the impasse of having no place to start the calculation, one may calculate the stopped fractions

$$m_1,\ m_1 + \frac{1}{m_2},\ m_1 + \frac{1}{m_2 + \frac{1}{m_3}}, \dots.$$

Problem 1. *Calculate the stopped fractions for the sequence* (3)*, both as standard fractions and as approximate decimal fractions, up to level ten of nesting.*

If you have solved this problem, you have seen that the stopped fractions quickly come close to each other and appear to produce more and more definite decimal places of a certain presumably infinite decimal fraction. One can indeed show for general continued fractions that one obtains in this way an irrational number. In this particular example, one can show that one obtains Euler's number e.

In this chapter we will be more concerned with the inverse problem, that is calculating the sequence (2) from a given real number x. Curiously, this process is much more straight forward than the forward calculation of the continued fraction. Assume we start with a positive number x. We shall assume that x is an irrational number, that is it is not the fraction of two natural numbers. We shall also assume is is at least one, since in the continued fraction (1) the term m_1 is already at least one, while the remaining nested fraction is nonneagitve since it is built from positive numbers using only addition and division, operations which do not leave the realm of nonnegative numbers.

The key observation is that if we write

$$x = m_1 + \frac{1}{x_1} \tag{4}$$

and compare with the continued fraction (1), then the number x_1 that we shall call residue is again a continued fraction

$$x_1 = m_2 + \frac{1}{m_3 + \frac{1}{m_4 + \frac{1}{m_5 + \cdots}}} ,$$

with a "layer stripped" sequence where the first term of the original sequence has been removed. Hence x_1 is at least one again, and thus $1/x_1$ is at most one. Note that $x - m$ cannot be an integer since x is not a fraction of integers and therefore not an integer neither. Hence whatever $1/x_1$ is, it is a well defined reciprocal since it is nonzero, and it is strictly less than one. Then the only possible choice for m_1 is the integer part of x, the largest integer not exceeding x. This determines m_1 and also x_1 as the reciprocal of the fractional part of x.

To see that we are equally safe from division by zero if we study the continued fraction of x_1, we observe that x_1 is again not a fraction of natural numbers. If was, say it was p/q, then x itself was a fraction:

$$x = m_1 + \frac{q}{p} = \frac{m_1 p + q}{q} ,$$

but we assumed it was not.

Now we can simply proceed to peel off the sequence (2) by taking away integer parts and iterating with the reciprocal of the fractional parts. An example will be discussed shortly.

Roots of integers. The number $\sqrt{2}$ is an example of a real number that cannot be written as a fraction. To see that, we take any fraction, call it p/q, and we convince ourselves that its square is not 2. For this it suffices to convince ourselves that

$$p^2 \neq 2q^2, \tag{5}$$

as one sees by dividing by q^2. We can assume the fraction p/q is reduced, or else we reduce it before continuing the discussion. If p is odd, then we conclude (5) since the left hand side is odd and the right hand side is even. If p is even, then q is odd since the fraction p/q is reduced. Then we conclude (5) again since the left hand side is divisible by four, while the right hand side is not divisible by four.

Problem 2. *Show that a natural number which is not a square of a natural number is not the square of a fraction of natural numbers neither.*

This problem gives us the maybe most basic source of real numbers which are not fractions. Ofcourse we are eager to test our continued fraction expansion on these numbers. We start with $x = \sqrt{2}$. Its integer part m_1 is 1, since

$$1^2 < 2 < 2^2 \ .$$

Note we did not yet algebraically use that $x^2 = 2$, we just used inequalities and the fact that squaring positive real numbers preserves the relative size. Thus m_1 has to be the largest natural number whose square does not exceed 2. Hence the first fractional part is $\sqrt{2} - 1$ and we obtain for its reciprocal:

$$x_1 = \frac{1}{\sqrt{2}-1} = \frac{\sqrt{2}+1}{(\sqrt{2}-1)(\sqrt{2}+1)} = \frac{\sqrt{2}+1}{2-1} = \sqrt{2}+1.$$

The key identity in the last display was the third one using the binomial formula

$$(a+b)(a-b) = a^2 - b^2 \ .$$

after suitable expansion of the fraction. Application of the bionomial formula causes the denominator of the fraction to become an integer again. This "turning of the table" helps to unravel the nesting of fractions in the continued fraction. In the present calculation, the integer in the denominator accidentally turns out to be one and thus disappears entirely, a luxury that may not always happen in future similar calculations.

We already know that $\sqrt{2}$ lies between one and two, hence x_1 lies between two and three, and we conclude $m_2 = 2$.

Now a great simplification occurs, in that we realize that the fractional part of x_2 is again $\sqrt{2}-1$, a fractional part we have just encountered. It is thus easy to anticipate the following calculations, they lead all to the same integer parts two. Thus (2) becomes

$$1, 2, 2, 2, 2, 2, \ldots .$$

If we compare with the difficulty of calculating the decimal expansion of $\sqrt{2}$, the relative ease of calculating the continued fraction

$$\sqrt{2} = 1 + \cfrac{1}{2 + \frac{1}{2 + \frac{1}{2 + \ldots}}} .$$

is quite remarkable.

Problem 3. *Calculate the continued fractions of square roots of small integers that are not integers themselves:*

$$\sqrt{3}, \sqrt{5}, \sqrt{6}, \ldots, \sqrt{n}, \ldots,$$

until you see certain patterns emerge. For each such square root, record the sequence (2) *of integer parts as well as the sequence* $x_1, x_2, x_3, x_4 \ldots$ *of residues, each written in the form*

$$\frac{\sqrt{n} + a}{b}$$

with integers a *and* b*. Watch out for repetitions in the residues.*

To get you started we show another example, say for $\sqrt{19}$. The calculations give for the sequence (2) of integer parts

$$4, 2, 1, 3, 1, 2, 8, 2, 1, 3, 1, 2, 8, \ldots,$$

while the first few terms of the sequence of residues are

$$\frac{\sqrt{19}+4}{3}, \frac{\sqrt{19}+2}{5}, \frac{\sqrt{19}+3}{2}, \frac{\sqrt{19}+3}{5}, \frac{\sqrt{19}+2}{3}, \frac{\sqrt{19}+4}{1},$$

at which point the next residue happens to be the same as the first residue shown and the sequence of residues repeats periodically.

Seeing the pattern emerge. We assume you and your team have done many calculations for Problem 2. What are the patterns that emerge? The following questions help understand the patterns.

Problem 4. *Why can we express all residues in the form*

$$\frac{\sqrt{n}+a}{b} \tag{6}$$

with an integer b dividing the integer $n-a^2$?

Solution: Initially we have $a=0$ and $b=1$, and it is clear that b divides $n-a^2$ in this case. Assume we have arrived at an expression

$$\frac{\sqrt{n}+a}{b},$$

either as initial term or as a calculated residue, such that b divides $n-a^2$. Subtracting the integer part m and using the binomial formula gives for the fractional part

$$\frac{\sqrt{n}+a-mb}{b} = \frac{n-(a-mb)^2}{b(\sqrt{n}-(a-mb))}.$$

Since b divides $n-a^2$, it also divides

$$n-(a-mb)^2 = (n^2-a^2)+2amb-b^2,$$

since it divides each summand on the right-hand-side. Hence we can write the next residue, that is the reciprocal of the penultimate display, as

$$\frac{\sqrt{n}-(a-mb)}{\frac{n-(a-mb)^2}{b}} = \frac{\sqrt{n}+a'}{b'}$$

where we have extracted the new integers

$$a' = mb-a \tag{7}$$

$$b' = \frac{n-(a-mb)^2}{b} \tag{8}$$

and clearly b' divides $n-(a')^2$. □

Note that not only each remainder can be written in the form (6), but this representation is unique because if

$$\frac{\sqrt{n}+a}{b} = \frac{\sqrt{n}+\tilde{a}}{\tilde{b}},$$

then trying to solve for $\sqrt{n}$ we obtain

$$(b-\tilde{b})\sqrt{n} = a-\tilde{a}.$$

Using that $\sqrt{n}$ is not a fraction we can conclude that $b=\tilde{b}$ and $a=\tilde{a}$.

Problem 5. *Why do we always have* $0 \le a \le \sqrt{n}$ *and* $0 < b$ *?*

Solution: The original x has $a = 0$ and $b = 1$ which satisfy the inequalities. Assume we have arrived at an expression

$$\frac{\sqrt{n} + a}{b}$$

where $0 \le a < \sqrt{n}$ and $0 < b$. The integer part m of that expression is the largest integer such that

$$\sqrt{n} + a - mb > 0,$$

hence with the formula (7) for a' we have

$$\sqrt{n} > mb - a = a'.$$

Also, since $m \ge 1$ is maximal,

$$\sqrt{n} + a - 2mb < 0$$

and since $a < \sqrt{n}$

$$2a - 2mb < 0$$

and hence $a' > 0$ by (7). That $b > 0$ then follows from

$$bb' = n - (a')^2.$$

□

Problem 6. *Why are we guaranteed that eventually we obtain a residue that we have calculated before?*

Solution: There are only finitely many choices of a, namely the integers between 0 and $\sqrt{n}$, and there are finitely many choices of b, namely the natural numbers less than n. Hence there eventually needs to be a residue that has occurred before. □

Problem 7. *Why do we always have* $\sqrt{n} \le a + b$*, except for the initial points* $a = 1$ *and* $b = 0$*?*

Solution: With the notation as in (7) and (8) and with $a < \sqrt{n}$ and $m \ge 1$, we conclude

$$b < mb + \sqrt{n} - a = \sqrt{n} + a' \ .$$

Since

$$bb' = n^2 - (a')^2 = (\sqrt{n} - a')(\sqrt{n} + a'),$$

we conclude

$$b' > \sqrt{n} - a'.$$

This shows the desired inequality $\sqrt{n} < a' + b'$ for each calculated residue. □

To assure that the property $a+b<\sqrt{n}$ is also satisfied for the initial x, we could slightly modify the task to calculate the continued fraction of

$$\frac{\sqrt{n}+[\sqrt{n}]}{1}$$

where $[\sqrt{n}]$ denotes the largest integer smaller than $\sqrt{n}$. Note that this doubles the first integer m_1 but gives the same first residue x_1 and thus the same sequence of residues. In the examples calculated this modification always lets the periodic pattern of the sequence (2) start with the first term.

Problem 8. *Given a' and b' as in* (7) *and* (8)*, how does one calculate the previous residue a and b (or the initial real number of the modified problem)?*

Solution: We calculate b using (8) in the form

$$b=(n^2-(a')^2)/b'.$$

Note that by Problems 5 and 7 we have

$$a=mb-a'\leq\sqrt{n},$$

$$a+b=(m+1)b-a'>\sqrt{n},$$

and hence m is determined as the largest integer such that $mb-a'<\sqrt{n}$. With that m, we may calculate a from (7). □

Problem 9. *Can you describe and prove the palindromic symmetry of the period of the residues?*

Solution: The modified problem is the continued fraction of

$$x=\frac{\sqrt{n}+[\sqrt{n}]}{1}.$$

We claim that the first residue that that is not new, call it x_N, is equal to x. Otherwise it was equal to some x_M with $0<M<N$. But then, by Problem 8, we would be able to calculate x_{N-1} and x_{M-1} (or the initial x if $M=1$) and observe they are equal, so x_N would not be the first residue repeated.

To discuss the palindromic symmetry, we look at the residues from x_1 to x_N. Now we claim that if for $1\leq K\leq N/2$ the residue x_K has the form (6) then x_{N+1-K} has the form

$$\frac{\sqrt{n}+a}{(n-a^2)/b}.$$

This claim describes the palindromic symmetry. The claim is immediately verified for $K = 1$ with $a = [\sqrt{n}]$ and $b = n - a$. If the claim is true for certain x_K, then x_{K+1} has the form

$$\frac{\sqrt{n} + a'}{b'}$$

with notation as in (7) and (8), and with Problem 8 the residue x_{N-K} has the form

$$\frac{\sqrt{n} + a'}{b} = \frac{\sqrt{n} + a'}{(n - (a')^2)/b'}.$$

This verifies the claim for x_{K+1} and proves the palindromic symmetry.
□

Problem 10. *Which neat equation for a and b can you prove about the residue in the symmetry point of this symmetry in the case when the length of the period is odd?*

Solution: If the period is odd (for example in the case of $\sqrt{2}$), then among the residues x_K there is a fixed point under the reflection symmetry and the fixed point has to satisfy

$$\frac{\sqrt{n} + a}{b} = \frac{\sqrt{n} + a}{(n - a^2)/b}$$

by the calculations in the solution to Problem 9. Simplifying, we conclude

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

□

Problem 11. *Now assume n is an odd prime number. Can you determine the last residue before the symmetry point if the period of the residues is even?*

Solution: An example for an even period is the case of $\sqrt{19}$. If the last residue before the symmetry point is written as in (6), then the next residue satisfies by the calculations in the solution to Problem 9 the equation

$$\frac{\sqrt{n} + a'}{b'} = \frac{\sqrt{n} + a}{(n - a^2)/b}.$$

We conclude $a = a'$ and by (7) we see that b divides $2a$. Assume first that b is odd. Then b divides a. Since b divides $n - a^2$ by Problem 4, it then also divides n. Since n is prime and $b < n$, we see $b = 1$. By Problem (7), we need a to be the largest integer less than $\sqrt{n}$. This pair of a and b is attained at the last point of the period, so it cannot be the last residue before the symmetry point. Hence b cannot be odd, and thus b has to be even.

If b is even, we conclude similarly to above that $b/2$ divides a and thus divides n and thus $b/2 = 1$. By Problem 7 we conclude that a is $[\sqrt{n}]$ or $[\sqrt{n}] - 1$. But since b is even and divides $n^2 - a$, a has to be odd und is uniquely determined. $\square$

Problem 12. *Now assume n is a prime number of the form $4k+1$ for some integer k. Can you prove that n can be written as the sum of two squares?*

Solution:

Using Problem (10) we need to show the period is odd. Assume to get a contradiction that the period is even. By Problem 11 the last residue before the symmetry point satisfies $b = 2$. Let a' and b' be as in (7) and (8) be the parameters of the next residue, the first one after the symmetry point. Since every square has remainder 0 or 1 modulo four, the difference $n - (a')^2$ has remainder 0 or 1 modula four. Since $n - (a')^2$ has to be divisible by 2, it is divisible by four and thus b' is even as well by (8). We claim that then all residues after the symmetry point written in the form (6) have even denominator. Namely, assume that any consecutive residues in the form (6) have even denominators and assume that the second residue is

$$\frac{\sqrt{n}+a}{b}.$$

Since the previous residue has even denominator, by (8) we have that $(n-a^2)/b$ is even. But then repeating the calculations for Problem (4)

$$(n-(mb-a)^2)/b$$

turns out also even, and thus the next residue also has even denominator. On the other hand, we know that the last residue in the period has denominator 1 by the considerations in Problem 9, a contradiction. $\square$

We have shown the hard part of the following Theorem, the rest of it is left as an easy problem:

Problem 13. *[Theorem of Fermat] A prime number is the sum of two squares if and only if dividing it by 4 gives a remainder of one or two.*

The theorem was stated by Pierre de Fermat, but since he had the habit of not writing down proofs, the first known proof is by Leonhard Euler.

Continued fractions of roots of general quadratic polynomials. We generalize the considerations to roots of more general quadratic polynomials. Let x be a number that solves a quadratic equation

$$rx^2 + sx + t = 0$$

with integers r, s, t. Note that $r = -1$, $s = 0$ and t a natural number that is not a square is the case discussed so far.

Problem 14. *Show that x is not a fraction of integers, if the discriminant*

$$s^2 - 4rt$$

is not a square integer.

Solution: We can write

$$x = \frac{s \pm \sqrt{s^2 - 4rt}}{2r}$$

with a certain choice of sign in place of $\pm$. If x was a fraction of integers, then so was $\sqrt{s^2 - 4rt}$, which would imply by Problem 2 that $s^2 - 4rt$ was a square of a natural number. $\square$

We may thus expand such roots of quadratic polynomials into continued fractions.

Problem 15. *Let x be a number greater than one that solves a quadratic equation*

$$rx^2 + sx + t = 0$$

for some integers r, s, t and assume x is irrational. Show that all residues x_K also satisfy such quadratic equations, say

$$r_K x_K^2 + s_K x_K + t_K = 0.$$

We may normalize so that

$$t_{K+1} = r_K, \tag{9}$$

and all discriminants $s_K^2 - 4r_K t_K$ have the same value.

Solution: First we notice that the equation

$$r(x - m_1)^2 + 2rm_1(x - m_1) + rm_1^2 + s(x - m_1) + sm_1 + t = 0$$

is true since it reduces to the assumed quadratic equation for x. Hence $x - m_1$ satisfies the quadratic equation

$$t_1(x - m_1)^2 + s_1(x - m_1) + r_1 = 0$$

with $t_1 = r$ and $s_1 = s + 2rm_1$ and $r_1 = rm_1^2 + sm_1 + t$. Then the reciprocal $x_1 = 1/(x - m_1)$ satisfies the equation

$$r_1 x_1^2 + s_1 x_1 + t_1 = 0.$$

The discriminant $s_1^2 - 4r_1t_1$ is easily calculated to be the same as the original one, $s^2 - 4rt$. Iterating this argument solves the problem. □

Note that since x_K is not a fraction, we have $r_K \neq 0$ for all K.

Problem 16. *Let x be an irrational number that solves a quadratic equation*

$$rx^2 + sx + t = 0$$

for some integers r, s, t and let r_K, s_K, t_K be as in Problem 15. Show that there are infinitely many indices K so that r_K and r_{K+1} have different signs.

Solution: Assume to obtain a contradiction that there are finitely many sign changes. Then there is an N such that all r_K have the same sign for $K \geq N$. By the normalization (9), the coefficients t_K for $K > N$ also have this same sign. We claim that there is a s_K with large L such that s_{N+L} also has the same sign. But we have $s_{K+1} = s_K + 2m_{K+1}r_K$ for all K and and since m_{K+1} is always positive and at least 1, we see that since the r_K all have the same sign past N, eventually s_{K+1} will have the same sign as well. This however gives a contradiction, as x_{N+L} is positive and cannot solve a quadratic equation with all coefficients of the same sign. □

For the next problem we ask to prove the following theorem of Joseph-Louis Lagrange.

Problem 17. *[Theorem of Lagrange] Let x be an irrational number that solves a quadratic equation*

$$rx^2 + sx + t = 0$$

Show that x has an eventually periodic continued fraction expansion.

Solution: We need to show that at some point we observe a residue that has occured before. Each time there is a sign change between r_K and r_{K+1}, we have that the discriminant

$$s_{K+1}^2 - 4r_{K+1}t_{K+1}$$

is a sum of two positive terms. Hence each of r_{K+1}, s_{K+1}, t_{K+1} is bounded by the discriminant, here we use that r_K and r_{k+1} are nonzero integers. There are only finitely many choices for such coefficients, and since there are infinitely many sign changes, eventually one has to encounter repetition in the coefficients. Hence the residues will eventually have a repetition. □

The converse of this theorem is slightly simpler. We first look at a specific continued fraction:

Problem 18. *Which number has the continued fraction expansion with sequence* (2) *consisting just of*

$$1, 1, 1, 1 \ldots ?$$

For the solution, observe $x_1 = x$ in (4) and calculate.

Each time a sequence (2) has a large entry m_K, the stopped fraction just prior to this entry is a relatively good approximation to the desired number since the error at the deepest level is the small number $1/m_K$. For example, the continued fraction expansion of π starts has a sequence (2) beginning with

$$3, 7, 15, 1, 292, 1, 1, \ldots .$$

The relatively large 15 causes the stopped fraction $3\frac{1}{7}$ to be a good approximation to π given the size of the denominator, and the large 292 causes the stopped fraction $3\frac{16}{113}$ to be an execllent approximation with six correct digits after the decimal point. Unfortunately the sequence (2) for π is not easy to understand.

The solution number to Problem 18 is for the discussed reason sometimes called the worst approximable irrational of all. It is also known as the golden ratio and has a colorful history.

The converse to Lagrange's theorem was observed by Euler, its proof is a modification of the solution to the last problem.

Problem 19. *[Theorem of Euler] Let x be a number that has an eventually periodic continued fraction expansion. Show that it solves a quadratic equation*

$$rx^2 + sx + t = 0$$

for some integers r, s, t.

We end this discussion with a note that one can also try to peel off a sequence (2) from a fraction $x = p/q$ of natural numbers with $p > q$.

Problem 20. *Show that the continued fraction algorithm for a rational x stops with some remainder $r_k = 0$.*

In case of a rational x the continued fraction algorithm is called Euclid's algorithm. If one refrains from reducing the fraction p/q and all subsequent residues, then one can read off the greatest common denominator of p and q from the last residue.