MTH5130 2021-2022

Shu Sasaki

9th December 2022

Contents

1 Introduction

Number theory can be thought of as having its roots in the study of *Diophantine equations*. Diophantine equations are polynomial equations with rational coefficients which we seek to solve while we insist the solutions should again be rational numbers.

Number theory has a long history. The Greeks knew, by about 400BC, that $X^2-2=0$ has no solution in rational numbers (the solutions defines a parabola).

Babylonians before 1600BC were already interested in rational solutions to the equation $X^2+Y^2=1$ (e.g. $(X,Y)=(1,0),(\frac{3}{5},\frac{4}{5}),(\frac{5}{13},\frac{12}{13}),\dots$ and the Greeks knew that there are infinitely many of them!), with a stone tablet to prove it.

An Arab manuscript around 972 AD, more or less, asks, for which integer N, is there a right angled triangle with area N whose sides have rational length? Algebraically, this amounts to solving the simultaneous equations

$$X^2+Y^2=Z^2$$

and

$$XY=2N$$

in positive rational numbers. In fact, the problem boils down to solving the equation

$$Y^2=X^3-N^2X$$

in non-zero rational numbers, and the equation defines an example of what we call *elliptic curves*.

Number theory is full of surprises and one does not have to be an expert in number theory to witness them. Let me give you another example. Let E denote the equation

$$Y^2+Y=X^3-X^2$$

and N_p denote the number of solutions to

$$Y^2+Y\equiv X^3-X^2 \bmod p.$$

Here is a table

p	2	3	5	7	11	13	17	19	$\cdots$
$p-N_p$	-2	-1	1	-2	1	4	-2	0	$\cdots$

On the other hand, consider the following infinite product in q

$$f = q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2$$

and it is an exercise in binomial expansions to find it equals

$$q-2q^2-q^3+2q^4+q^5+2q^6-2q^7-2q^9-2q^{10}+q^{11}-2q^{12}+4q^{13}+4q^{14}-q^{15}-4q^{16}-2q^{17}+4q^{18}+\cdots.$$

Can you see a pattern? The number $p - N_p$ is just the coefficient of the p-th power of q. Is this a coincidence, or is there any explanation for it? The Shimura-Taniyama conjecture (it is a theorem of C. Breuil, B. Conrad, F. Diamond, R. Taylor and A. Wiles) asserts that any equation of the form

$$Y^2 + aY = X^3 + bX^2 + cX + d,$$

where $a, b, c, d \in \mathbb{Q}$, corresponds to a power series like f in a similar manner.

Number Theory has become an extremely technical subject drawing on techniques from all over mathematics. Nonetheless, it retains, at the heart of the subject, a particular beauty and elegance in its simplicity of messages it inspires in people (Gauss called number theory the 'queen of mathematics'): you might have heard the following:

- **Fermat's Last Theorem** (P. Fermat, L. Euler, ..., A. Wiles): the equation $X^n + Y^n = Z^n$ has no solutions in integers when $n \geq 3$.
- **the Twin prime conjecture** (..., Y. Zhang, J. Maynard, T. Tao, B. Green, ... not completely proved yet): there exist infinitely many pairs of primes that differ by 2 (e. g. $\{3, 5\}$ and $\{17, 19\}$).
- **the Goldbach conjecture** (open): every integer (> 2) is the sum of two primes.
- **the Riemann hypothesis** (open): the "non-trivial" zeros of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$ in $s \in \mathbb{C}$ all have real part $\frac{1}{2}$ (this is one of the seven Millennium problems; if you solve one of these, you will receive one million dollars from the Clay Maths Institute).

They are extremely hard to prove. For example, it took about 350 years for the FLT to be proved completely and this is the only one in the list that has been proved! David Hilbert, one of the greatest mathematicians in the 19th and early 20th centuries, famously said "If I were to awaken after having slept for a thousand years, my first question would be: has the Riemann Hypothesis been proven?".

Topics such as **the Langlands program**, **the Birch-Swinnerton-Dyer conjecture**, **the Fontaine-Mazur conjecture** and **the abc conjecture** are right at the centre of very active research in number theory that continues to inspire many researchers in mathematics.

There are a lot of number theorists in

https://www.bbc.co.uk/programmes/b00srz5b/episodes/player

In my youth, I spent a lot of time reading entries in

https://mathshistory.st-andrews.ac.uk

I occasionally stumble upon articles from

https://www.quantamagazine.org/tag/number-theory/

2 Samplers

So what exactly are we going to learn? Inevitably, they are all about numbers (by which I normally mean integers/natural numbers). Let p be an odd prime number. We will answer questions such as

- (quadratic reciprocity) Given a number a, is it congruent to the square of an integer mod p (e.g. $-1 \equiv 5^2 \bmod 13$ but no square is congruent to $-1 \bmod 19$)? Is there an easy way (i.e., easier than checking all p mod p residues) to figure this out?
- (Continued fraction and Diophantine approximation) How closely can $\sqrt{p}$ be approximated by a rational number (e.g $\sqrt{2}$ is approximately $\frac{141421}{100000}$ but $\frac{1393}{985}$ is an even better approximation with much smaller numerator and denominator)? What do we mean exactly by 'good' approximation?
- (Pell equations) Does the equation $x^2 - py^2 = 1$ have a solution? What about $x^2 - py^2 = -1$ (e.g. $18^2 - 13 \cdot 5^2 = -1$ but there is no solution to $x^2 - 19y^2 = -1$)?
- (Representations of primes as sums of squares) Can we express p in the form $x^2 + y^2$ for some natural numbers x and y (e.g. $13 = 3^2 + 2^2$ but 19 cannot be)?

3 Revision

Let $\mathbb{N} = \{1, 2, \dots\}$ be the set of natural numbers. Let $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \dots\}$ be the set of integers.

3.1 Euclid's algorithm

We all know by experience that if a pair of non-negative integers a and b with $b > 0$, there exist integers $q \geq 0$ and $r \geq 0$ such that $0 \leq r < b$ and

$$a = bq + r.$$

The q (resp. r) is often referred to as the *quotient* (resp. *remainder*) when we divide a by b.

In fact, something more general is true (See Appendix): for integers a and b with $b > 0$ (i.e. a can be negative!), there exist unique integers q and r such that $0 \leq r < b$ and

$$a = bq + r.$$

When the residue $r = 0$, we say b divides a, we often write

$$b|a.$$

Though b has been assumed to be positive, this definition of 'division' holds more generally for $b = 0$, i.e, 0 divides a if there exists an integer q such that $a = 0q$. And this forces $a = 0$. In other words, the only integer which 0 divides is 0! Note that we are only considering '0 divides 0', and not considering the fraction $\frac{a}{b} = \frac{0}{0}$ (which makes sense if $b > 0$ divides a). Indeed, it also works for negative b!: for integers a and b, we say that b divides a (and write $b|a$) if there exists an integer q such that $a = bq$.

Remark. Note that $a|b$ means different from $b|a$. Also do not confuse this with b/a which is a rational number with numerator b and denominator a (I will typically write $\frac{b}{a}$ though).

Definition. The highest common factor, or the greatest common divisor in this course, $d = \gcd(a, b)$ of two *integers* (not necessarily positive!) a and b is a non-negative integer d characterised by the following properties:

- $d|a$ and $d|b$,
- if e is a natural number satisfying $e|a$ and $e|b$, then $e|d$.

When $\gcd(a, b) = 1$, we often say that a and b are relatively prime, or coprime.

Example. $\gcd(4, 6) = 2$. The only integers that divide 4 and 6 are ± 1 and ± 2. They all divide 2. The point is that gcd is *defined to be* non-negative.

Remark. Note that, by definition, gcd is a non-negative integer. It is certainly possible to *define* it to be merely an integer satisfying the properties above (in which case 2 and -2 are both $\gcd(4, 6)$) but it would be more convenient for everyone to talk about *the* gcd.

Example. $\gcd(0, 0) = 0$. Indeed, for any integer $n \geq 0$, $\gcd(n, 0) = n$.

To prove the latter (which specialises to $n = 0$), we use the definition of gcd. Let $g = \gcd(n, 0)$ for brevity. Firstly, g divides n by definition. Therefore there exists an integer r such that $n = rg$. On the other hand, since $n|n$ and $n|0$ (because $0 = n0$), it follows from the second property of gcd that $n|g$, i.e., there exists an integer s such that $g = sn$. Combining, $n = rg = rsn$. It follows that either n is zero or n is non-zero with $rs = 1$. If n is zero, g is zero (since $g = s0 = 0$) and we are done. If n is non-zero, $rs = 1$, hence $(r, s) = (1, 1)$ or $(-1, -1)$. However, since both g and n are non-negative, the only possibility is $(r, s) = (1, 1)$, i.e., $n = g$.

Remark. Do you know $\gcd(a, b) = \gcd(-a, b) = \gcd(a, -b) = \gcd(-a, -b)$?

Let us prove the first equality (NON-EXAMINABLE). Observe that if d is an integer, then $d|a$ and $d|b$ is equivalent to $d| - a$ and $d|b$. If $\gcd(a, b) = 0$, then $a = 0$ and $b = 0$, and both gcd's are 0. If $\gcd(a, b) > 0$, then $\gcd(-a, b)$ is also > 0. For if it were zero, it would imply $a = 0$ and $b = 0$ and $\gcd(a, b) = 0$ which is a contradiction. The aforementioned equivalence shows that $\gcd(a, b)|\gcd(-a, b)$ and $\gcd(-a, b)|\gcd(a, b)$. Since both gcd's are positive, $\gcd(a, b) = \gcd(-a, b)$. The other equalities can be proved similarly.

Euclid's algorithm is a procedure to find the gcd systematically, given a pair of integers a, b with $b > 0$. The algorithm is based on the observation that

$$\gcd(a, b) = \gcd(b, r)$$

where r is the unique integer satisfying $a = bq + r$ with $0 \leq r < b$. See Appendix 2.

Note that Euclid'a algorithm computes $\gcd(a, b)$ when at least only one of them, often labelled as 'b', is positive, but Remark above shows that gcd of two negative integers, a and b say, can be computed by $\gcd(-a, -b)$ for example. The latter can be computed via Euclid's algorithm.

Exercise. Find $\gcd(225, 157)$.

$$\begin{aligned} 225 &= 157 \cdot 1 + 68 \\ 157 &= 68 \cdot 2 + 21 \\ 68 &= 21 \cdot 3 + 5 \\ 21 &= 5 \cdot 4 + 1 \\ 5 &= 1 \cdot 5 + 0 \end{aligned}$$

Repeatedly applying the 'key observation', one sees $\gcd(225, 157) = \gcd(157, 68) = \gcd(68, 21) = \gcd(21, 5) = \gcd(5, 1) = 1$.

Exercise. What is $\gcd(123, 456)$? What is $\gcd(123, -456)$?

$$\begin{aligned} 456 &= 123 \cdot 3 + 87 \\ 123 &= 87 \cdot 1 + 36 \\ 87 &= 36 \cdot 2 + 15 \\ 36 &= 15 \cdot 2 + 6 \\ 15 &= 6 \cdot 2 + 3 \\ 6 &= 3 \cdot 2 + 0 \end{aligned}$$

On the other hand,

$$\begin{aligned} -456 &= 123 \cdot (-4) + 36 \\ 123 &= 36 \cdot 3 + 15 \\ 36 &= 15 \cdot 2 + 6 \\ 15 &= 6 \cdot 2 + 3 \\ 6 &= 3 \cdot 2 + 0 \end{aligned}$$

Euclid's algorithm also finds a pair of *integers* r and s such that

$$ar + bs = \gcd(a, b).$$

In the first example above, we work back up the chain:

$$\begin{aligned} 1 &= 21 - 5 \cdot 4 \\ &= 21 - (68 - 21 \cdot 3) \cdot 4 = 21 \cdot 13 - 68 \cdot 4 \\ &= (157 - 68 \cdot 2) - 68 \cdot 4 = 157 \cdot 13 - 68 \cdot 30 \\ &= 157 \cdot 13 - (225 - 157) \cdot 30 = 157 \cdot 43 - 225 \cdot 30 \end{aligned}$$

So $r = -30$ and $s = 43$ work. In fact:

Proposition 1. *For any $d \in \mathbb{N}$ and $a, b \in \mathbb{Z}$, the following are equivalent:*

- *the equation $ax + by = d$ is soluble (in integers) in x and y.*
- $\gcd(a, b)$ *divides d.*

Proof. For brevity, let g denote $\gcd(a, b)$. Firstly, suppose that there exist r, s in $\mathbb{Z}$ such that $ar + bs = d$. Since any common divisor of a and b divides the LHS, it divides d on the RHS. In particular, g, the *greatest* common divisor of a and b, divides d.

Conversely, suppose that g divides d. Let $d = zg$ for some $z \in \mathbb{Z}$. By Euclid'a algorithm, one can find r, s in $\mathbb{Z}$ such that $ar + bs = g$. Multiplying the equation by z, we have

$$zg = z(ar + bs) = a(zr) + b(zs).$$

Since $d = zg$, the pair $(x, y) = (zr, zs)$ defines a solution for the equation $ax + by = d$.□

Example. We know $\gcd(225, 157) = 1$, so $225x+157y = d$ is soluble in integers for any integer $d > 0$. Indeed, the proof explains how to find a solution: using Euclid's algorithm, find integers r, s such that $225r+157s = 1$. Then $(x, y) = (rd, sd)$, as $225rd+157sd = (225r+157s)d = 1d = d$.

Example. We know $\gcd(123, -456) = 3$. It follows from the proposition that $123x + (-456)y = 2$ is *not* soluble. On the other hand, $123x+(-456)y = 6$ is soluble, because Euclid's algorithm find a pair of integers r and s such that $123r+(-456)s = 3$ and therefore $(x, y) = (2r, 2s)$ does the job.

One can extend these concepts to more than two numbers: if $a_1, \ldots, a_N \in \mathbb{N}$, then we have a gcd $d = \gcd(a_1, \ldots, a_N)$ such that $a_1x_1 + \cdots a_Nx_N = d$.

3.2 Primes and factorisation

A natural number $p \in \mathbb{N} \cup \{0\}$ is said to be *prime* if

- $p > 1$,
- if $p = ab$ holds for some $a, b \in \mathbb{Z}$, then we either have $(a, b) = (p, 1), (-p, -1), (1, p)$ or $(-1, -p)$.

We are going to show that every positive integer greater than 1 can be factored into primes, and the factorisation is unique up to the possibility of writing the factors in a different order (e.g. $12 = 2 \cdot 2 \cdot 3 = 2 \cdot 3 \cdot 2 = 3 \cdot 2 \cdot 2$). This innocuous 'fact' is in fact known as the Fundamental Theorem of Arithmetic. We will prove this rather carefully.

Proposition 2. *(Bezout's identity) Given $a, b \in \mathbb{Z}$, there exist integers r, s such that*

$$ar + bs = \gcd(a, b).$$

Proof. See Example sheet 1. □

Lemma 3. *Let a and b be integers and p be a prime. If $p|ab$, then either $p|a$ or $p|b$ holds.*

Proof. Suppose that p does not divide a. It suffices to prove, assuming p divides ab, that p divides b. Since p does not divide a, $\gcd(a,p) = 1$ [this is where we use the assumption that p is prime; since p is a prime, it follows from the definition that a divisor of p is either ± 1 or $\pm p$ and it is clear that only ± 1 commonly divides a as p does not divides a by assumtion]. It therefore follows that there exists r, s in $\mathbb{Z}$ such that $ar + ps = 1$ by Bezout. Multiplying both sides by b, we obtain

$$b = b(ar + ps) = abr + pbs.$$

Since p divides ab, it divides the term abr. Also p certainly divides pbs. It therefore follows that p divides b. □

Remark. The lemma shows that p is prime (in the sense defined above) if and only if the assertion $\boxed{\text{if } p|ab \text{ for some integers } a \text{ and } b \text{ then either } p|a \text{ or } p|b \text{ holds}}$ holds. So it is possible to use $\boxed{\text{if...}}$ as the definition of a prime number!– in fact the latter definition is more amenable to generalisations and is used to define prime elements/prime ideals in number fields (algebraic number theory).

To see the equivalence advertised above, we argue as follows:

Suppose firstly that $\boxed{\text{if...}}$ holds. And suppose that $p = \alpha\beta$ for some integer α and β. By assumption, $p|\alpha\beta$. It then follows from $\boxed{\text{if...}}$ that either $p|\alpha$ or $p|\beta$. If it is the former, then α must be either p or $-p$ (hence β is either 1 or -1), while if it is the latter, β must be either p or $-p$ (in which case α is either 1 or -1). It follows that p is prime.

On the other hand, suppose that p is a prime number. The lemma proves that $\boxed{\text{if...}}$ holds.

Lemma 4. *Let $a_1, \ldots, a_N$ be integers and p be a prime. If $p|a_1 \cdots a_N$, then $p|a_n$ for some $1 \leq n \leq N$.*

Proof. By the lemma, p divides a_1 or the product $a_2 \cdots a_N$. If p divides a_1, we are done. Otherwise p divides $a_2 \cdots a_N$. Using the lemma again, p therefore divides either a_2 or $a_3 \cdots a_N$. Repeat the argument. □

Theorem 5. *(The Fundamental Theorem of Arithmetic) Any natural number greater than 1 can be written as a product of prime numbers, and this product expression is unique apart from re-ordering of the factors.*

Proof. We prove the existence of prime factorisation by induction. Let N be a natural number. Suppose that the statement of the theorem holds for any natural number $\leq N - 1$. If N itself is a prime, then there is nothing to prove. If N is not a prime, it is a product of two integers each of which is $< N$. For these integers, we know from the inductive hypothesis that these two numbers are indeed product of prime numbers. Putting them together, N is a product of prime numbers.

To prove the uniqueness of prime factorisation, let $N = p_1 \cdots p_r = q_1 \cdots q_s$ be prime factorisations of N. Since p_1 divides $q_1 \ldots q_s$, it follows from the lemma above that p_1 divides q_n for some $1 \leq n \leq s$. By re-ordering q's if necessary, we may assume that p_1 divides q_1. Since they are both positive integers, $p_1 = q_1$. Repeat the argument, starting with $p_2 \cdots p_r = q_2 \cdots q_s$. □

3.3 Appendix: Euclidian algorithm (PROOFS NON-EXAMINABLE)

Proposition. Let $a, b \in \mathbb{Z}$ and suppose $b > 0$. There exist unique $q, r \in \mathbb{Z}$ such that $0 \leq r < b$.

Proof. Existence Let $S = \{a + zb \mid z \in \mathbb{Z}, a + zb \geq 0\}$. Since $a \in S$, it follows that S is non-empty, and let r be the smallest element of S. Necessarily, r is of the form

$$r = a + (-q)b \geq 0$$

for some $q \in \mathbb{Z}$.

$r < b$ If $r \geq b$, then

$$0 \leq r - b = a - (q+1)b < a - qb = r$$

contradicting the minimality r. Hence $r < b$.

Uniqueness Suppose $a = qb + r$ with $0 \leq r < b$; and $a = q'b + r'$ with $0 \leq r' < b$. Observe that $r' = a - bq' \in S$, hence $r' \geq r$ (by the minimality of r). It follows from

$$r' = a - bq' \geq a - bq = r$$

that $q \geq q'$. We may let $q' = q - s$ for some $s \geq 0$. It follows that

$$r' = a = bq' = a - b(q - s) = a - bq + bs = r + bs.$$

It then follows that $s = 0$, therefore $q = q'$ and $r = r'$. □

Corollary. If $a = qb + r$ as above, then

$$\gcd(a, b) = \gcd(b, r).$$

Proof. Let $g = \gcd(a, b)$ and $h = \gcd(b, r)$.

Firstly suppose that $g = 0$. Since g divides a and b, and 0 is the only integer 0 can divide, $a = 0$ and $b = 0$. Therefore $g = 0$. It also follows that $r = 0$, hence $h = 0$.

On the other hand, if $g > 0$, then so is h. If h was 0, an argument similar to the one above would show that $g = 0$ which contradicts the assumption.

$g|h$ Since $g|a$ and $g|b$, $g|(a - bq)$, i.e., $g|r$. Since $g|b$ by definition, it follows from the second property of gcd that $g|h$. By the minimality of h, we then conclude that $g \leq h$.

Similarly,

$h|g$ Since $h|b$ and $h|r$, $h|(qb + r)$, i.e., $h|a$. Since $h|b$ by definition, it follows from the second property of gcd that $h|g$.

Combining, $g = h$ as both g and h are positive integers. □

3.4 Congruences and modular arithmetic

Let n be a positive natural number. We say that $a, b \in \mathbb{Z}$ are congruent mod n if $n|(a - b)$ and write

$$a \equiv b \pmod{n}$$

or even

$$a \equiv b$$

if 'mod n' is clear from the context.

Fact. Congruence mod n is an equivalence relation; the equivalence classes (there are n of them, corresponding to the n possible remainders, $0, 1, \ldots, n-1$, when we divide a number by n) are called congruence classes modulo n. We denote by $[a]_n$ the congruence class modulo n that is *represented* by a. As a set

$$[a]_n = \{z \in \mathbb{Z} | \, z \equiv a \bmod n\} = \{\ldots, a-2n, a-n, a, a+n, a+2n, \ldots\}.$$

By definition, any member of $[a]_n$ can represent the class. To put it simply,

$$\boxed{a \equiv b \bmod n \text{ if and only if } [a]_n = [b]_n}$$

We let $\mathbb{Z}/n\mathbb{Z}$ denote the set of all congruence classes mod n (I would write it $\mathbb{F}_p$ if n is a prime number p); it is a ring (see Supplementary notes 1) with addition

$$[a]_n + [b]_n = [a+b]_n$$

and multiplication:

$$[a]_n \cdot [b]_n = [ab]_n.$$

Remark. Strictly speaking, one needs to check that these operations are independent of representatives one chooses, i.e., if $a \equiv a' \bmod n'$, is it true that $[a]_n + [b]_n = [a']_n + [b]_n$?

Example. The set $\mathbb{Z}/4\mathbb{Z}$ has 4 classes:

$$\begin{array}{rcl} [0]_4 & = & \{\ldots, -12, -8, -4, 0, 4, 8, 12, \ldots\} \\ [1]_4 & = & \{\ldots, -11, -7, -3, 1, 5, 9, 13, \ldots\} \\ [2]_4 & = & \{\ldots, -10, -6, -2, 2, 6, 10, 14, \ldots\} \\ [3]_4 & = & \{\ldots, -9, -5, -1, 3, 7, 11, 15, \ldots\} \end{array}$$

and they add:

$+$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[0]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[1]_4$	$[1]_4$	$[2]_4$	$[3]_4$	$[0]_4$
$[2]_4$	$[2]_4$	$[3]_4$	$[0]_4$	$[1]_4$
$[3]_4$	$[3]_4$	$[0]_4$	$[1]_4$	$[2]_4$

and multiply:

$\times$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[0]_4$	$[0]_4$	$[0]_4$	$[0]_4$	$[0]_4$
$[1]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[2]_4$	$[0]_4$	$[2]_4$	$[0]_4$	$[2]_4$
$[3]_4$	$[0]_4$	$[3]_4$	$[2]_4$	$[1]_4$

Proposition 6. If p is a prime, then $\mathbb{F}_p$ is a field.

Proof. The hardest part is to show that any non-zero element of $\mathbb{F}_p$ has multiplicative inverse. Let $[a]$ be a non-zero element (the zero element being $[0]$); this amounts to assuming that p does not divide a. Since $\gcd(a,p) = 1$, there exists r, s in $\mathbb{Z}$ such that $ar + ps = 1$ (Bezout). This means that $ar \equiv 1 \bmod p$. Phrased in terms of the corresponding congruence classes, it follows that $[a][r] = [1]$, i.e. $[a]$ has an inverse $[r]$. □

Remark. Observe that this proof is constructive, and the key input is Bezout's identity/Euclid's algorithm.

Remark. Note that if n is not a prime, $\mathbb{Z}/n\mathbb{Z}$ is not necessarily a field. For example, when $n = 4$, the class $[2]_4$ in $\mathbb{Z}/4\mathbb{Z}$ does not have multiplicative inverse, i.e. there is no integer z such that $[2]_4[z]_4 = [1]_4$. This can be check from the multiplication table above and see the row/column of $[2]_4$. Whether z is $0, 1, 2, 3$. $[2]_4[r]_4$ is never $[1]_4$.

Note that 'division' makes sense only over a field– in a ring R (e.g. $\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}$,...), 'division' by an element r of R is nothing other than 'multiplication by the (multiplicative) inverse of r', and this operation inherently assumes that the inverse exists in R in the first place (i.e. an element s of R such that $rs = sr = 1$ in R). It is only when we know it exists in R that one can write s as $\frac{1}{r}$. For example, in $\mathbb{Q}$, it is not possible to divide a rational number by 0, because 0 does not have multiplicative inverse; any non-zero rational number $\frac{r}{s}$, where r and s are both non-zero integers, do have multiplicative inverse inverse (which we know is as $\frac{s}{r}$). On the other hand, the ring $\mathbb{Z}$ is not closed under 'division', because almost all elements do not have their multiplicative inverses in $\mathbb{Z}$. For example, 2 does not have multiplicative inverse– indeed $\frac{1}{2}$ is the inverse, but it is not an element of $\mathbb{Z}$.

What Proposition 6 ascertains is that one can perform division in $\mathbb{F}_p$, or equivalently 'mod p'– in fact, the proof demonstrates how to find the multiplicative inverse of $[a]_p$ or, equivalently the inverse of a mod p (an integer z such that $ar \equiv 1 \bmod p$). On the other hand, it is not possible to do so in $\mathbb{Z}/4\mathbb{Z}$.

Example. If $\mathbb{F}_p$ is a field, any non-zero element has an inverse. Let $p = 157$. Then the congruence class $[225]_{157}$ defines a non-zero element in $\mathbb{Z}/157\mathbb{Z}$. What is the inverse? In other words, what is $x \in \mathbb{Z}$ such that $[225]_{157}[x]_{157} = [1]_{157}$? Recall from earlier that Euclid's algorithm gave us $157 \cdot 43 - 225 \cdot 30 = 1$. Reducing mod 157, we have $[-30]_{157}[225]_{157} = [1]_{157}$, so $x = -30$ works.

One of my favourite theorems in number theory:

Theorem 7 (Fermat's Little Theorem). Let p be a prime number. Then $z^p \equiv z \bmod p$ for any $z \in \mathbb{N}$.

Remark. Do not confuse this with Fermat's Last Theorem! In theory, both can be abbreviated as 'FLT'.

Proof. If p divides z, or equivalently z is congruent to 0 mod p, the assertion clearly holds. We

therefore suppose that z is not congruent to 0 mod p. Consider the set

$$\{z, 2z, \ldots, (p-1)z\}$$

and the set

$$\{r_1, \ldots, r_{p-1}\}$$

of 'residues' where $1 \leq r_j \leq p-1$ is defined to be the residue (in the range $[1, p-1]$) of jz when divided by p.

$\boxed{\{r_1, \ldots, r_{p-1}\} = \{1, \ldots, p-1\}}$ By definition, we only know that $\{r_1, \ldots, r_{p-1}\}$ is a subset of $\{1, \ldots, (p-1)\}$ but they are indeed equal. To see this, it suffices to establish if $1 \leq i, j \leq p-1$ are distinct, then r_i and r_j are distinct. This is equivalent to proving that if $r_i = r_j$ (i.e. $iz \equiv jz$ mod p), then $i \equiv j$ mod p (so that $i = j$ since $1 \leq i, j \leq p-1$).

Suppose that $r_i = r_j$. By definition, we have $iz = pq_j + r_i$ and $jz = pq_j + r_j$ for some integers q_i and q_j. Subtracting one from the other, we have

$$(i-j)z = p(q_i - q_j) + (r_i - r_j) = p(q_i - q_j) \equiv 0$$

mod p. Since z has inverse mod p (for z is coprime to p), it is possible to divide the congruence by z and we have

$$(i-j) \equiv 0$$

mod p as desired.

The upshot of this observation is that, for every $1 \leq j \leq p-1$, there exists a unique $1 \leq i \leq p-1$ such that $jz \equiv r_i$.

In particular, the product of all elements in $\{z, 2z, \ldots, (p-1)z\}$ and $\{r_1, \ldots, r_{p-1}\} = \{1, \ldots, p-1\}$ must coincide mod p. This therefore results in

$$\prod_{j=1}^{p-1} j \equiv \prod_{j=1}^{p-1} zj = z^{p-1} \prod_{j=1}^{p-1} j.$$

Every j $(1 \leq j \leq p-1)$ has inverse mod p; therefore, by multiplying both sides of the congruence identity by these inverses, we obtain $1 \equiv z^{p-1}$ mod p. Multiplying z on both sides, $z \equiv z^p$ mod p. □

Here's another proof.

Proof. We may suppose that z is not congruent to 0 mod p. The congruence class $[z]$ is a non-zero element of $\mathbb{F}_p$. It then follows from the Lagrange's theorem (see the Introduction to Algebra notes) that the order of $[z]$ divides the order, $p-1$, of the multiplicative group of $\mathbb{F}_p$. It therefore follows that $[1] = [z]^{p-1} = [z^{p-1}]$. In other words, $z^{p-1} \equiv 1$ mod p. Multiplying both sides by z, we have $z^p \equiv z$ mod p. □

Example. Let $p = 7$. If $a \equiv b$ mod p, i.e., a and b belong to the same congruence class, then $a^p \equiv b^p$. So it suffices to check on representatives $0 \leq r \leq p-1$.

z	0	1	2	3	4	5	6
z^7	0	1	128	2187	16384	78125	279936
$z^7 - z$	$7 \cdot 0$	$7 \cdot 0$	$7 \cdot 18$	$7 \cdot 312$	$7 \cdot 234$	$7 \cdot 11160$	$7 \cdot 39990$

One possibly surprising application of Fermat's Little Theorem is that, given a number N, there is a chance that we will know N is composite (i.e. not a prime). All one has to do is spot a natural number z such that z^N is *not* congruent to z mod N. For if N were a prime, then z^N would be congruent to z mod N for *any* z.

Be careful that it is certainly possible that $z^N \equiv z \bmod N$ holds for *some* z even if N is not a prime number. For example, $3^6 = 729 \equiv 3 \bmod 6$ (see Example Sheet 1). What if $z^N \equiv z \bmod N$ holds for *any* z? Does it mean that N is a prime number?

Example. $3^{2047} \equiv 992 \pmod{2047}$, hence 2047 is *not* a prime number.

3.5 Congruence equations

Proposition 8. Let a, n be natural numbers and let b be an integer. The congruence equation $ax \equiv d \bmod n$ is soluble if and only if $\gcd(a, n)$ divides d.

Proof. Suppose that $ax \equiv d \bmod n$ is soluble in x, i.e., there exists an integer r such that $ar \equiv d \bmod n$. In other words, $ar + zn = d$ for some $z \in \mathbb{Z}$. A common divisor of a and n, in particular, $\gcd(a, n)$, divides the LHS and therefore the RHS.

Conversely, suppose that $g = \gcd(a, n)$ divides d. We may then let $a = ga', d = gd'$ and $n = gn'$. It suffices to see that the congruence equation $a'x \equiv d' \bmod n'$ is soluble.

[Why does this suffice? Suppose that r is an integer such that $a'r \equiv d' \bmod n'$. Evidently, $r + sn'$ for any integer s is also a solution for the congruence equation, because $a'(r + sn') - d' = a'r - d' + a'sn'$ and both $a'r - d$ and $a'sn'$ are divisible by n'. With this in mind, let t be an integer such that $a'(r + sn') - d' = tn'$. It then follows that

$$a(r + sn') = ga'(r + sn') = g(d' + tn') = d + tn \equiv d$$

mod n. This shows that any integer congruent to r mod n' is a solution to the congruence equation $ax \equiv d \bmod n$.]

Since $\gcd(a', n') = 1$, it follows from Bezout that there exists $r, s \in \mathbb{Z}$ such that $a'r + n's = 1$. Multiplying the both sides by d', we find $a'(d'r) + n'(d's) = d'$ and the congruence equation $a'x \equiv b' \bmod n'$ has a solution $x = d'r \bmod n'$. □

Remark. In the lecture, I did not talk about solving $a'x \equiv d' \bmod n'$. I simply made appeal to $\gcd(a', n') = 1$ and Bezout to find r and s such that $a'r + n's = 1$ and multiply $1 = a'r + n's$ by $d = d'g$ to get $d = d(a'r + n's) = bdg(a'r + n's) = a(d'r) + n(d's)$. This establishes, rather quickly, that $ax \equiv d \bmod n$ has a solution $x = d'r \bmod n$. This is correct in terms of proving the proposition, but the proof above actually shows something stronger(!), namely that $d'r$ is a solution mod n', not just modulo n.

Example. Let $a = 2, n = 3, b = 5$. Since $\gcd(2, 3) = 1$ and this divides 5, the theorem asserts that the congruence equation $2x \equiv 5 \bmod 3$ is soluble (of course, it is possible to rewrite the congruence equation as $2x \equiv 2 \bmod 3$ but it is more instructive to keep 'b' in its original form).

A dogged approach (effective when n is very small):

$x \bmod 3$	0	1	2
$2x \bmod 3$	0	2	4
$2x - 5 \bmod 3$	1	0	2

Hence $x \equiv 1 \bmod 3$ is a solution.

A slick approach (effective when n is large). While the theorem itself does not say 'how to solve the congruence equation', the proof is constructive and one can follow it to find a solution. Since $\gcd(2,3) = 1$, $a' = a = 2$, $n' = n = 3$ and $b' = b = 5$. Apply Euclid's algorithm to find $1 = (-1) \cdot 2 + 1 \cdot 3$. Hence $(-1) \cdot 5 \bmod 3$, i.e., $1 \bmod 3$ is the solution for $2x \equiv 5 \bmod 3$.

Example. Let $a = 2, n = 4, b = 5$. Since $\gcd(2,4) = 2$, the theorem says the congruence equation $2x \equiv 5 \bmod 4$ is not soluble. For any $x \bmod 4$, $2x$ is always even mod 4 and cannot possibly be congruent to 5.

3.6 The Chinese Remainder Theorem

The CRT is about solving simultaneous congruences to different moduli. We say that m and n are coprime if $\gcd(m,n) = 1$.

Theorem 9. Let m, n be coprime natural numbers. Then there is a solution to the simultaneous congruence equations:

$$x \equiv a \pmod{m}$$
$$x \equiv b \pmod{n}.$$

Indeed the solution is unique modulo mn in the sense that if x and y are both solutions, then $x \equiv y \bmod mn$ holds.

Proof. We firstly show the *existence*. Since $\gcd(m,n) = 1$, there are integers r, s such that $mr + ns = 1$. We therefore have

$$mr \equiv 0 \bmod m,$$
$$mr \equiv 1 \bmod n,$$
$$ns \equiv 1 \bmod m,$$
$$ns \equiv 0 \bmod n.$$

Let $x = mrb + nsa$. This is what we are looking for. Indeed, $x \equiv nsa \equiv a \bmod m$ while $x \equiv mrb \equiv b \bmod n$.

To prove the *uniqueness*, suppose that x and y are solutions. On one hand, it follows from $x \equiv a \equiv y \bmod m$ that $m|(x-y)$. On the other hand, $x \equiv b \equiv y \bmod n$ implies that $n|(x-y)$. Since m and n are coprime, we may then conclude $mn|(x-y)$, in other words, $x \equiv y \bmod mn$. □

More generally,

Theorem 10. Let $n_1, \ldots, n_r$ be pairwise coprime natural numbers. Then there is a solution, unique modulo $n_1 \cdots n_r$, to the congruences

$$x \equiv a_i \pmod{n_i}.$$

Example. Find x satisfying the following simultaneous equations:

$$x \equiv 2 \pmod 3,$$

$$x \equiv 1 \pmod 4,$$

$$x \equiv 3 \pmod 5.$$

The CRT theorem says that there is a unique solution mod 60 which can be found either by trial-and-error, or following the systematic argument in the proof.

Firstly, we solve the fist two equations– we apply the argument in the proof of CRT with $m = 3, n = 4, a = 2$ and $b = 1$. By Euclid's algorithm, we find $3 \cdot (-1) + 4 \cdot 1 = 1 = \gcd(3, 4)$ and

$$x = 3 \cdot (-1) \cdot 1 + 4 \cdot 1 \cdot 2 = 5$$

define a solution mod 12. We need to solve the following simultaneous equations

$$y \equiv 5 \pmod{12},$$

$$y \equiv 3 \pmod 5.$$

Since $\gcd(12, 5) = 1$, we may apply CRT with $m = 12, n = 5, a = 5$ and $b = 3$. By Euclid's algorithm, we find $12 \cdot (-2) + 5 \cdot 5 = 1 = \gcd(12, 5)$ and

$$y = 12 \cdot (-12) \cdot 3 + 5 \cdot 5 \cdot 5 = 53$$

defines a solution mod 60.

3.7 Prime numbers

Not that, apart from 2, every prime is congruent to either 1 or $-1 \equiv 3 \bmod 4$. Indeed,

Theorem 11. There are infinitely many primes congruent to $-1 \bmod 4$.

Proof. Suppose that there are only finitely many such primes, say $q_1, \ldots, q_r$. Consider $N = 4q_1 \ldots q_r - 1$. It is congruent to $-1 \bmod 4$.

$\boxed{N \text{ is not a prime}}$ Indeed if it were a prime, it would be one of the q's but N is clearly bigger than any one of them.

If N is not a prime, then it is composite. However,

$\boxed{2 \text{ is not a factor of } N}$ If it were, N would be even, but it is not (since N is congruent to $-1 \bmod 4$, it is congruent to $-1 \bmod 2$).

Similarly,

$\boxed{q \in \{q_1, \ldots, q_r\} \text{ is not a factor of } N \text{ either}}$ If it were, $N \equiv 0 \bmod q$, but by definition $N = 4q_1 \ldots q_r - 1 \equiv -1 \bmod q$.

We may then conclude

any prime factor of N is congruent to 1 mod 4 We have established that any prime factor of N is *not* congruent to -1 mod 4. On the other hand, a prime factor cannot be congruent to 0 nor 2 mod 4 (if it were, 2 would be a prime factor of N which, we know, is not true).

However, the product of integers (we only need to observe it for primes numbers though) that are congruent to 1 mod 4 again is congruent to 1 mod 4 and this contradicts $N \equiv -1$ mod 4. Therefore, there are infinitely many primes congruent to -1 mod 4. □

Remark. It is true that there are infinitely many primes congruent to 1 mod 4, but what goes wrong with the argument if we run it for primes congruent to 1 mod 4?

4 Euler's totient function and primitive roots

4.1 Euler's totient function

Definiton. Euler's totient function, or Euler's ϕ-function, is the function $\phi : \mathbb{N} \to \mathbb{N}$ that sends n in $\mathbb{N}$ to the number of natural numbers $1 \leq z \leq n$ coprime to n (i.e. $\gcd(z, n) = 1$).

Example. If p is a prime, $\phi(p) = p - 1$, since $1, 2, \ldots, p - 1$ are all coprime to p.

Example. $\phi(8) = 4$; the odd numbers $1, 3, 5, 7$ are coprime to 8, while the even numbers $2, 4, 6, 8$ are not.

Definition. If R is a commutative ring with identity, then an element r in R is said to be a unit if there exists s in R such that $rs = 1$. The units in R form a group under multiplication.

Proposition 14. The number $|(\mathbb{Z}/n\mathbb{Z})^\times|$ of elements in the group $(\mathbb{Z}/n\mathbb{Z})^\times$ of units in $\mathbb{Z}/n\mathbb{Z}$ is $\phi(n)$.

Proof. It suffices to establish that $[z]$ is a unit in $\mathbb{Z}/n\mathbb{Z}$ if and only if z is relatively prime to n.

Suppose that z is relatively prime to n, i.e., $\gcd(z, n) = 1$. It then follows that there exists r, s in $\mathbb{Z}$ such that $zr + ns = 1$. Hence $zr \equiv 1$ mod N and this is nothing other than saying that $[z][r] = [1]$ and $[z]$ is a unit.

Conversely, suppose that $[z]$ is a unit. Then there exists a congruence class $[r] \in \mathbb{Z}$ such that $[z][r] = [zr] = [1]$. It follows that $zr \equiv 1$ mod n and we may write $zr + ns = 1$ for some r in $\mathbb{Z}$. Let $d = \gcd(z, n)$. By definition, d divides z and divides n, and therefore it divides $zr + ns$. Therefore $d = 1$. □

From this, we can deduce

Theorem 15. Let n be a positive integer and z be an integer such that $\gcd(z, n) = 1$. Then $z^{\phi(n)} \equiv 1$ mod n.

Proof. Compare this proof with the proof of Fermat's Little Theorem. For brevity, let $N = \phi(n)$ and let $z_1, \ldots, z_N$ be the integers in $\{0, 1, \ldots, n\}$ that are relatively prime to n. Consider the set $\{r_1, \ldots, r_N\}$ where r_j is defined to be the residue $1 \leq r_j \leq n-1$ of zz_j when divided by n. Indeed, $\{r_1, \ldots, r_N\} = \{z_1, \ldots, z_N\}$. To see this, it suffices to establish that if $r_i = r_j \bmod n$ for $1 \leq i,j \leq N$, i.e., $zz_i \equiv zz_j \bmod n$, then $i \equiv j \bmod n$ (hence $i = j$). But this follows by multiplying both sides by the inverse of z (it exists because z is coprime to n). Since $\{z_1 z, \ldots, z_N z\} \equiv \{r_1, \ldots, r_N\} = \{z_1, \ldots, z_N\} \bmod n$, we have

$$z^N \prod_{j=1}^{N} z_j = \prod_{j=1}^{N} z_j z \equiv \prod_{j=1}^{N} z_j$$

Since z_j for every $1 \leq j \leq N$ is invertible, so is $\prod_{j=1}^{N} z_j$, and it follows that $1 \equiv z^N \bmod n$. □

Corollary 16. Let p be a prime. Then $z^p \equiv z \bmod p$ for any integer z.

[This *is* Fermat's Little Theorem. In other words, Theorem 15 generalises the FLT.]

Proof. Let $n = p$ in the theorem. Then $z^{p-1} \equiv 1 \bmod p$ for z not divisible by p. Multiplying the both sides by z, we have $z^p \equiv z \bmod p$. On the other hand, if p divides z, then $z \equiv 0 \bmod p$ and $z^p \equiv 0 \equiv z \bmod p$. □

Theorem 17.

1. If p is a prime and $r > 0$, then $\phi(p^r) = p^{r-1}(p-1)$.
2. If $\gcd(k, \ell) = 1$, then $\phi(k\ell) = \phi(k)\phi(\ell)$.
3. If $n = p_1^{r_1} \cdots p_s^{r_s} = \prod_{j=1}^{s} p_j^{r_j}$, where $p_1, \ldots, p_s$ are distinct primes and $r_1, \ldots, r_s > 0$, then

$$\phi(n) = \prod_{j=1}^{s} p_j^{r_j - 1}(p_j - 1) = n \prod_{j=1}^{s} (1 - 1/p_j)$$

Proof. Non-examinable. □

Example. $720 = 2^4 \cdot 3^2 \cdot 5$

$$\phi(720) = 2^3(2-1)3^1(3-1)5^0(5-1) = 8 \cdot 6 \cdot 4 = 192.$$

Proposition 18. Let d be a divisor of n. Then the number of integers z with $1 \leq z \leq n$ and $\gcd(z, n) = d$ is $\phi\left(\frac{n}{d}\right)$.

[Note that $\gcd(z, n) = d$ forces z to be greater than, or equal to, d!]

Proof. Let $n = d\ell$. The multiplication by d define a map

$$S_\ell = \{1 \leq z' \leq \ell \,|\, \gcd(z', \ell) = 1\} \to \{d \leq z \leq n \,|\, \gcd(z, n) = d\} = S_n.$$

It suffices to establish that this map is bijective (since $|S_\ell| = \phi(\ell)$). To prove the surjectivity, let z be an element on the RHS. Since $\gcd(z, n) = d$, we may divide z by d. Call it z'. By definition, $1 \leq z' \leq \ell$ and $\gcd(z', \ell) = 1$, hence z' is an element of S_ℓ. The injectivity follows immediately because $dz' = dz''$ for $z', z'' \in S_\ell$ immediately implies $z' = z''$. $\square$

Example. Let $n = 60$. According to the proposition, the number of integers $1 \leq z \leq 60$ such that $\gcd(z, 60) = 4$ is $\phi(\frac{60}{4}) = \phi(15) = \phi(3 \cdot 5) = (3-1)(5-1) = 8$. They are

$$\{4, 8, 16, 28, 32, 44, 52, 56\}.$$

The number of integers $1 \leq z \leq 60$ such that $\gcd(z, 60) = 6$ is $\phi(\frac{60}{6}) = \phi(10) = \phi(2 \cdot 5) = (2-1)(5-1) = 4$. They are

$$\{6, 18, 42, 54\}.$$

Definition. Let $n \in \mathbb{N}$. If there exists a positive integer d such that $z^d \equiv 1 \pmod{n}$, then the order of z mod n is the smallest of all such integer d. Alternatively, the order of z may be defined as the smallest integer d such that $[z]_n^d = [1]_n$ in the set $\mathbb{Z}/n\mathbb{Z}$ of congruence classes mod n.

Example. Let $n = 7$.

z mod 7	order mod 7
1	1
2	3
3	6
4	3
5	6
6	2

For example, the following table shows that the order of 3 is indeed 6:

3^n	$3^n \pmod 7$
3^0	1
3^1	3
3^2	2
3^3	6
3^4	4
3^5	5
3^6	1

Lemma 19. Suppose z has order d. If $z^e \equiv 1 \bmod n$, then $d|e$.

Proof. Write $e = dq + r$ with $0 \leq r \leq d-1$. It suffices to show that $r = 0$. It then follows that

$$1 \equiv z^e = z^{dq+r} = z^{dq} z^r \equiv z^r$$

mod n. But by definition, d is the smallest power for which $z^d \equiv 1$. Since $r \leq d-1$, the only possibility is that $r = 0$. $\square$

The following proposition explains 'when' it makes sense for us to talk about the order of an integer mod n:

Proposition 20. For an integer z, there exists $d \in \mathbb{N}$ such that $z^d \equiv 1 (\text{mod } n)$ if and only if $\gcd(z, n) = 1$. If so, the order of z divides $\phi(n)$.

Proof. If $z^d \equiv 1 \bmod n$ holds, then $\gcd(z^d, n) = 1$ (if not, i.e., if $\gcd(z^d, n) > 1$, then it would have to divide 1, which is absurd). Evidently, $\gcd(z, n) = 1$ (because $\gcd(r, n) | \gcd(r^d, n)$. To see this, obverse that if $r | \gcd(z, n)$, then $r | z$ and $r | n$, hence $r | z^d$ and $r | n$. As a result, it follows that $r | \gcd(z^d, n)$).

Conversely, if $\gcd(z, n) = 1$, then $z^{\phi(n)} \equiv 1 \bmod n$ (Theorem 15), hence there do exist such integers. The order d is the smallest among them.

The last assertion follows from the lemma. □

Example. Let $n = 12$. In this case, $\phi(12) = 4$ with the integers between 1 and 12 coprime to 12 are $1, 5, 7, 11$. We have $1^1 \equiv 1$, $5^2 \equiv 1$, $7^2 \equiv 1$ and $11^2 \equiv 1$ modulo 12. They have orders $1, 2, 2, 2$ respectively and they divide 4. Note that *not every divisor of $\phi(n)$ necessarily occurs as the order of an element*.

4.2 Primitive roots

While we are still on the subject of mod n orders of integers, we specialise n to be a prime number p and spotlight a class of integers of order $p - 1 \bmod p$.

Definition. Let p be a prime number. An integer z is said to be a primitive root mod p if z has order $p-1 (\text{mod } p)$. Note that, since $\phi(p) = p-1$, a primitive root has the maximum possible order.

In terms of congruence classes, this is paraphrased as follows: an integer z is a primitive root mod p if its mod p congruence class $[z]_p$ has order $p - 1$ in the multiplicative group $\mathbb{F}_p^\times$, i.e., the smallest positive integer N such that $[z]^N = [z^N] = [1]$ holds is $N = p - 1$; a slick way of saying this is that $[z]$ generates the multiplicative group $\mathbb{F}_p^\times$, i.e. $\{[z], [z]^2, \ldots, [z]^{p-1}\} = \mathbb{F}_p^\times$.

Since $\mathbb{F}_p^\times = \{[1], \ldots, [p - 1]\}$ and if $z' \equiv z \bmod p$ then $[z'] = [z]$, it is only necessary to understand the orders of $1, \ldots, p - 1 \bmod p$ to spot *all* primitive roots.

Example. What are the primitive roots mod $p = 7$? Looking at the table above, every integer that is congruent to 3 or 5 mod 7 is a primitive root mod 7.

Example. Is it possible to find a primitive root mod $p = 17$? Since $2^8 \equiv 1 \bmod 17$, 2 is not a primitive root. In fact 3 is a primitive root mod 17. It seems rather laborious to check all 3^r for $1 \leq r \leq 15$ is not congruent to 1 mod 17 and only 3^{16} is. However, Lemma 19 and Theorem 15 show that if d is the order of 3 mod p, then d has to divide $\phi(17) = 16$. Since $1, 2, 4, 8, 16$ are the divisors of 16, the order d has to be one of them. To determine d exactly, we need do try-and-error:

$3^1 = 3$, $3^2 = 9$, $3^4 = 81 \equiv 13 \equiv (-4)$, $3^8 \equiv (-4)^2 = 16 \equiv (-1)$, $3^{16} \equiv (-1)^2 = 1$; hence 16 is the order of 3 mod 17 and 3 is a primitive root mod 17. Here we are using the trick that if $a \equiv b \bmod n$, then $a^r \equiv b^r \bmod n$ for any integer $r \geq 1$.

Lemma 21. Let p be a prime and d be a divisor of $p - 1$. Then the number of elements in

$\{1, \ldots, p-1\}$ of order d mod p is either 0 or $\phi(d)$.

Proof. Suppose that the number of such elements is non-zero. So there is at least one element z of order d mod p.

the numbers $1, z, z^2, \ldots, z^{d-1}$ are all distinct mod p For if $z^i = z^j \bmod p$ with $0 \leq i < j \leq d-1$, then $z^{j-i} \equiv 1 \bmod p$. But $j - i < d$ and this contradicts the minimality of d.

these $1, z, \ldots, z^{d-1}$ all have order at most d mod p For every $1 \leq j \leq p-1$,

$$(z^j)^d = (z^d)^j \equiv 1^j = 1$$

mod p, the order of z^j is at most d.

for $0 \leq j \leq d-1$, that z^j has order d if and only if $\gcd(j, d) = 1$

Firstly, suppose $\gcd(j, d) > 1$. The goal is to show that z^j does not order d; since we know that z^j has order at most d, this is equivalent to asserting that the order of z^j is $< d$.

In this case, there exists $g > 1$ that divides $\gcd(n, d)$. Let $j = gi$ and $d = ge$. In particular, since $g > 1$, we have $e < d$. Observe $z^{je} = z^{di} \equiv 1$, hence z^j has order at most $e < d$.

Conversely, suppose that $\gcd(j, d) = 1$. The goal is to show that z^j has order exactly d. Let r be the order of z^j mod p; in particular, $z^{jr} \equiv 1 \bmod p$. On the other hand, since $z^d \equiv 1 \bmod p$, it follows from Lemma 19 that $d | jn$. Since $\gcd(j, d) = 1$, it follows that $d | r$; in particular, $d \leq r$. We already know that the order r of z^j is at most d. Hence $d = r$. □

The proof of the lemma actually explains how to find *all* elements in $\{1, \ldots, p-1\}$ of order d mod p, as soon as we find an element 'z' of order d to go on with. Let us work out examples.

Example. Let $p = 17$. To find the elements of order $d = 16$, i.e., the primitive roots in $\{1, \ldots, p-1\}$, firstly we find 'z'. For example, 3 is a primitive root mod 17. According to the proof, the elements of order $d = 16$ in $\{1, \ldots, 16\}$ therefore are

$$\{3^j \mid 1 \leq j \leq 16 \text{ and } \gcd(j, 16) = 1\} = \{3, 3^3, 3^5, 3^7, 3^9, 3^{11}, 3^{13}, 3^{15}\} \equiv \{3, 10, 5, 11, 14, 7, 12, 6\}.$$

The right-most is worked out by finding r_j which is congruent to 3^j mod 17 satisfying $0 \leq r_j \leq 16$.

To find the elements of order $d = 8$, we use 2 which has order 8 mod 17. Then the elements of order 8 in $\{1, \ldots, 16\}$ are

$$\{2^j \mid 1 \leq j \leq 7 \text{ and } \gcd(j, 8) = 1\} = \{2^1, 2^3, 2^5, 2^7\} \equiv \{2, 8, 15, 9\}.$$

As is more or less clear from the proof of the lemma and the examples that we do not know yet if there is indeed an element in $\{1, \ldots, p-1\}$ of order d mod p exists at all or not (to start the process). The following proves the 'existence'.

Theorem 22. Let p be a prime. For every number d dividing $p-1$, let S^d_{p-1} denote the set of elements in $\{1, \ldots, p-1\}$ of order d mod p. Then

$$|S^d_{p-1}| = \phi(d).$$

In particular, there are $\phi(p-1)$ primitive roots mod p.

Proof. Let $\varphi(d)$ denote the number $|S_{p-1}^d|$ of elements in $\{1,\dots,p-1\}$ of order d mod p. The goal is to show that $\varphi(d) = \phi(d)$. We show

(a) $\displaystyle\sum_{d|(p-1)} \phi(d) = p-1,$

(b) $\displaystyle\sum_{d|(p-1)} \varphi(d) = p-1,$

(c) For any d, we have $\varphi(d) \le \phi(d)$.

It follows immediately from these that $\varphi(d) = \phi(d)$ for $d|(p-1)$.

(a) We have

$$\{1,\dots,p-1\} = \bigcup_{d|(p-1)} \{1 \le z \le p-1 \,|\, \gcd(z,p-1) = (p-1)/d\,\}$$

since every $1 \le z \le p-1$ satisfies that $\gcd(z,p-1) = (p-1)/d$ for some d.

By Proposition 18, $|\{1 \le z \le p-1 \,|\, \gcd(z,p-1) = (p-1)/d\,\}|$ is indeed $\phi\left((p-1)/\dfrac{(p-1)}{d}\right) = \phi(d)$. Hence (a) follows.

(b) We have

$$\{1,\dots,p-1\} = \bigcup_{d|(p-1)} \{1 \le z \le p-1 \,|\, z \text{ has order } d \bmod p\} = \bigcup_{d|(p-1)} S_{p-1}^d$$

since, by Fermat's Little Theorem, every integer $1 \le z \le p-1$ has some mod-p order d that divides $p-1$. Hence (b) follows.

(c) Lemma 21 shows the stronger result that, for every $d|(p-1)$, we have either $\varphi(d) = 0$ or $\varphi(d) = \phi(d)$. □

Example. $p = 7$.

d	1	2	3	6
S_d	$\{1\}$	$\{6\}$	$\{2,4\}$	$\{3,5\}$
$\varphi(d)$	1	1	2	2
$\phi(d)$	1	1	2	2

Theorem 23 Let z be a primitive root mod p. Then the order mod p of z^n is equal to $(p-1)/\gcd(n,p-1)$.

Proof (NON-EXAMINABLE). It is possible to tinker the proof of Lemma 21 to prove this, but we will provide a direct proof. Let $\gcd(n,p-1) = r$ and write $n = rk$ and $(p-1) = r\ell$. We let d

denote the order of z^n mod p. The goal is to show that $\boxed{d = \ell}$.

$\boxed{d|\ell}$ By definition, $\gcd(k, \ell) = 1$. We have

$$(z^n)^\ell = z^{n\ell} = z^{rk\ell} = z^{(p-1)k} \equiv 1^k = 1$$

mod p (the last congruence follows because the order of z is $p - 1$). Hence it follows from Lemma 19 that $d|\ell$.

$\boxed{\ell|d}$ Since z has order $p - 1$, it follows from Lemma 19 that $p - 1$ divides nd (as we know that the order of z^n is d, hence $z^{nd} \equiv 1 \bmod p$). In other words, $r\ell = (p - 1)$ divides $rkd = nd$, hence ℓ divides kd. On the other hand, ℓ is coprime to k, hence ℓ divides d. Combining $d|\ell$ and $\ell|d$, we obtain $d = \ell$ as desired. □

5 Quadratic residues and non-residues, Gauss reciprocity law

The goal of this section is to decide, when p is an odd prime, whether the congruence equation

$$x^2 \equiv a \pmod{p}$$

has integer solutions or not, for any integer a not divisible by p.

Definition. Let a be an integer not divisible by p. It is a quadratic residue $(\bmod p)$ if there exists an integer z with $z^2 \equiv a (\bmod p)$; and a is a quadratic non-residue if no such z exists.

Remark. It makes sense to define an integer a, divisible p, to be a quadratic residue mod p. As $a \equiv 0$ mod p by assumption, for any z divisible by p (e.g. $z = 0$), we have $z^2 \equiv 0 \equiv a$ mod p. On the other hand, this 'exceptional' case breaks 'symmetry' and we will not lose much by excluding it from the mix.

Remark. If $a \equiv b \pmod{p}$, then a is a quadratic residue if and only if b is. Therefore, it suffices to consider $a \in \{1, \ldots, p - 1\}$.

To work out which integers $1 \leq a \leq p - 1$ are quadratic residue mod p or not, one way of doing it is to list all square integers mod p.

Example Let $p = 3$.

a mod 3	a^2 mod 3
1	1
2	1

So, the quadratic residues are the integers congruent mod 3 to

1,

while the quadratic non-residues are the integers congruent mod 3 to

$$2$$

Example. Let $p = 7$.

$a \bmod 7$	$a^2 \bmod 7$
1	1
2	4
3	2
4	2
5	4
6	1

So, the quadratic residues are (the integers congruent mod 7 to)

$$1, 2, 4,$$

while the quadratic non-residues are (the integers congruent mod 7 to)

$$3, 5, 6.$$

Example. Let $p = 11$.

$a \bmod 11$	$a^2 \bmod 11$
1	1
2	4
3	9
4	5
5	3
6	3
7	5
8	9
9	4
10	1

So, the quadratic residues are (the integers congruent mod 11 to)

$$1, 3, 4, 5, 9$$

while the quadratic non-residues are (the integers congruent mod 11 to)

$$2, 6, 7, 8, 10$$

Remark. Have you noticed that $a^2 \equiv (p-a)^2 \equiv (-a)^2 \bmod p$? Because of this, one only has to check up to $a \leq \frac{p}{2}$ in general.

Let p be an odd prime number. By Theorem 22, there exists a primitive root $z \bmod p$ (in fact there are $\phi(p-1)$ primitive roots mod p exist). Consider the the set

$$\{1, z, z^2, \ldots, z^{p-2}\}.$$

Firstly, none of the elements is congruent to 0 mod p. If it were, say $z^j \equiv 0 \bmod p$ for some $1 \leq j \leq p-2$ (if $j = 0$, then $z^j = 1$ and this is clearly not congruent to 0 mod p), then $z \equiv 0 \bmod p$. This follows since z has (multiplicative) inverse (more precisely z^{p-2}).

Secondly, the set $\{1, z, z^2, \ldots, z^{p-2}\}$ is is in bijection with

$$\{1, \ldots, p-1\}.$$

An alternative way of phrasing this is that $\{[1], [z], \ldots, [z^{p-2}]\} = \{[1]_p, \ldots, [p-1]\}$. To see the bijection, it suffices to show that the z^j's are all distinct mod p, as this implies that the residues of z^j are all distinct and $\{1, 2, \ldots, p-1\}$ is the set of all possible residues (for integers not divisible by p). To show that the z^j are distinct mod p, suppose that they are not distinct and that there exist integers $0 \leq i < j \leq p-2$ such that $z^i \equiv z^j \bmod p$. Then, since z has multiplicative inverse mod p, we see that $z^{j-i} \equiv 1 \bmod p$. However, $1 < j - i < p-1$ and this contradicts the minimality of the order $p-1$ of z mod p.

It follows from this discussion that

$$\boxed{\text{if } a \text{ is an integer not divisible by } p, \text{ then } a \equiv z^j \text{ for some } 0 \leq j \leq p-2}$$

because the residue of a when divided by p defines an element of $\{1, \ldots, p-1\}$. With this in mind:

Proposition 24. *a is a quadratic residue mod p if and only if a is an even power of z; and is a quadratic non-residue if and only if it is an odd power of z.*

Proof. We show that $\boxed{z^j \text{ is a quadratic residue if and only if } j \text{ is even}}$.

Suppose, firstly, that j is even and let $j = 2i$. Then $z^j = (z^i)^2$, hence z^j is clearly a quadratic residue. Conversely, suppose that $a \equiv z^j$ is a quadratic residue, hence there exists an integer b such that $a \equiv b^2 \bmod p$. Replace b by its residue if necessary, we may assume that $1 \leq b \leq p-1$. As observed earlier, there must exist $0 \leq i \leq p-2$ such that $b \equiv z^i \bmod p$. Substituting, we have $z^j \equiv a \equiv b^2 \equiv z^{2i}$. Since z is primitive and has inverse, we deuce that $z^{2i-j} \equiv 1 \bmod p$. It then follows from Lemma 19 that $p-1$ divides $2i - j$. Since $2i$ and $p-1$ are both even [this is where the assumption that p is odd is used!], we then conclude that j is even, as desired. □

Example. Let $p = 7$. We know that $z = 3$ is a primitive root mod 7.

3^j	$3^j \bmod 7$
3^0	1
3^1	3
3^2	2
3^3	6
3^4	4
3^5	5
3^6	1

So any integer congruent to 1, 2 or 4 mod 7 is a quadratic residue, while any integer congruent to 3, 5 or 6 is a quadratic non-residue. This is consistent with the example earlier.

5.1 the Legendre symbol

Definition. The Legendre symbol is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p|a \\ +1 & \text{if } p \text{ does not divide } a \text{ and } a \text{ is quadratic residue mod } p \\ -1 & \text{if } p \text{ does not divide } a \text{ and } a \text{ is quadratic non-residue mod } p \end{cases}$$

Remark. By definition, for any integer a not divisible by p, we have $\left(\frac{a}{p}\right)\left(\frac{a}{p}\right) = 1$.

Theorem 25.

(Rule 0) If $a \equiv b \bmod p$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(Rule 1) If p is an odd prime and $a, b \in \mathbb{Z}$, then

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$

(Rule 2) If p is an odd prime, then

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} +1 & \text{if } p \equiv 1 \bmod 4 \\ -1 & \text{if } p \equiv 3 \bmod 4 \end{cases}$$

(Rule 3) If p is an odd prime, then

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} +1 & \text{if } p \equiv 1 \text{ or } 7 \bmod 8 \\ -1 & \text{if } p \equiv 3 \text{ or } 5 \bmod 8 \end{cases}$$

(Rule 4) (Quadratic Reciprocity) For any pair of distinct odd primes p and q,

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4} = \begin{cases} -1 & \text{if } p \equiv q \equiv 3 \bmod 4 \\ +1 & \text{otherwise} \end{cases}$$

Before proving these assertion,

Example. $\left(\frac{13}{17}\right) = 1$. Firstly, Rule 4 asserts that $\left(\frac{13}{17}\right)\left(\frac{17}{13}\right) = (-1)^{\frac{13-1}{2}\frac{17-1}{2}} = 1$, hence $\left(\frac{13}{17}\right)$ is computed by $\left(\frac{17}{13}\right)$. We then make appeal to Rule 0 to deduce that $\left(\frac{17}{13}\right) = \left(\frac{4}{13}\right)$ since $17 \equiv 4 \bmod 13$. On the other hand, Rule 1 says $\left(\frac{4}{13}\right) = \left(\frac{2}{13}\right)\left(\frac{2}{13}\right) = 1$.

Example. 38 is a quadratic residue mod 43. One way of checking this, of course, is to solve the congruence equation $x^2 \equiv 38 \bmod 43$. We will use Theorem 25 to prove $\left(\frac{38}{43}\right) = 1$:

$$\begin{aligned}
\left(\frac{38}{43}\right) &= \left(\frac{2}{43}\right)\left(\frac{19}{43}\right) && \text{(Rule 1)}\\
&= -\left(\frac{19}{43}\right) && \text{(Rule 3)}\\
&= \left(\frac{43}{19}\right) && \text{(Rule 4)}\\
&= \left(\frac{5}{19}\right) && \text{(Rule 0)}\\
&= \left(\frac{19}{5}\right) && \text{(Rule 4)}\\
&= \left(\frac{4}{5}\right) && \text{(Rule 0)}\\
&= +1 && (4 \equiv 2^2 \bmod 5)
\end{aligned}$$

Of course, this not the only way to get to $\left(\frac{38}{43}\right) = 1$. For example,

$$\begin{aligned}
-\left(\frac{19}{43}\right) &= -\left(\frac{-24}{43}\right) && \text{(Rule 0)}\\
&= -\left(\frac{-1}{43}\right)\left(\frac{2}{43}\right)^2\left(\frac{6}{43}\right) && \text{(Rule 1)}\\
&= \left(\frac{6}{43}\right) && \text{(Rule 2)}\\
&= \left(\frac{49}{43}\right) && \text{(Rule 0)}\\
&= \left(\frac{7}{43}\right)^2 && \text{(Rule 1)}\\
&= +1
\end{aligned}$$

Remark. Even if we know that 38 is a quadratic residue mod 43 in terms of the Legendre symbol, we still do not know the solutions to the congruence equations $x^2 \equiv 38 \bmod 43$. We will come back to this issue shortly.

Corollary 26. If p is an odd prime and not equal to 3, then

$$\left(\frac{3}{p}\right) = \begin{cases} +1 & \text{if } p \equiv 1 \text{ or } 11 \bmod 12 \\ -1 & \text{if } p \equiv 5 \text{ or } 7 \bmod 12 \end{cases}$$

Proof. By Rule 4, it follows that $\left(\frac{3}{p}\right)\left(\frac{p}{3}\right) = (-1)^{\frac{p-1}{2}\frac{3-1}{2}} = (-1)^{\frac{p-1}{2}}$, hence

$$\left(\frac{3}{p}\right) = \begin{cases} +\left(\frac{p}{3}\right) & \text{if } p \equiv 1 \bmod 4, \\ -\left(\frac{p}{3}\right) & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

Also

$$\left(\frac{p}{3}\right) = \begin{cases} +1 & \text{if } p \equiv 1 \bmod 3 \\ -1 & \text{if } p \equiv 2 \bmod 3 \end{cases}$$

since 1 is a quadratic residue mod 3 while 2 is not (see Example above). Combining these:

$$\left(\frac{3}{p}\right) = \begin{cases} +\left(\frac{p}{3}\right) & \text{if } p \equiv 1 \bmod 4, \text{ which yields } \begin{cases} +1 & \text{if } p \equiv 1 \bmod 3, \\ -1 & \text{if } p \equiv 2 \bmod 3, \end{cases} \\ -\left(\frac{p}{3}\right) & \text{if } p \equiv 3 \bmod 4, \text{ which yields } \begin{cases} -1 & \text{if } p \equiv 1 \bmod 3, \\ +1 & \text{if } p \equiv 2 \bmod 3, \end{cases} \end{cases}$$

hence

$$\left(\frac{3}{p}\right) = \begin{cases} +1 & \text{if (1) } p \equiv 1 \bmod 4 \ \& \ p \equiv 1 \bmod 3 \text{ or (2) } p \equiv 3 \bmod 4 \ \& \ p \equiv 2 \bmod 3, \\ -1 & \text{if (3) } p \equiv 1 \bmod 4 \ \& \ p \equiv 2 \bmod 3 \text{ or (4) } p \equiv 3 \bmod 4 \ \& \ p \equiv 1 \bmod 3. \end{cases}$$

For example, (2) amounts to finding the prime numbers in the solutions of the system of congruence equations $x \equiv 3 \bmod 4$ & $x \equiv 2 \bmod 3$. By the Chinese Reminder Theorem, its unique solution is $x \equiv (-1) \equiv 11 \bmod 12$. Hence (2) is equivalent to $p \equiv 11 \bmod 12$. Do similar calculations to this for (1), (3) and (4). □

Let us prove Rule 1 and Rule 2.

Proof of Rule 1. If either a or b is divisible by p, the assertion follows immediately. We therefore assume that both a and b are not divisible by p.

Let z be a primitive root mod p (it exists by Theorem 22). As we saw in the proof of Proposition 24, a (being not divisible by p) is congruent to $z^j \bmod p$ for some $0 \leq j \leq p-2$ and it follows that $\left(\frac{a}{p}\right) = (-1)^j$. On the other hand, we may also let $b \equiv z^i$ for some $0 \leq i \leq p-2$, and therefore $\left(\frac{b}{p}\right) = (-1)^i$. Then $ab \equiv z^{i+j}$ and therefore $\left(\frac{ab}{p}\right) = (-1)^{i+j} = (-1)^j(-1)^i = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$. □

Proof of Rule 2. Let z be a primitive root mod p and let $\zeta = z^{(p-1)/2}$. We then have $\zeta^2 = z^{p-1} \equiv 1 \bmod p$, but ζ is *not* congruent to $1 \bmod p$ (if it were, z would have order $(p-1)/2 < (p-1)$, contradicting the minimality of the order $p-1$), so it has to be congruent to $-1 \bmod p$. It follows that -1 is a quadratic residue (resp. non-residue) if $(p-1)/2$ is even (resp. odd), i.e., if $p \equiv 1$ (resp. $p \equiv 3$) mod 4. □

Proposition 27 (Euler's Criterion). Let a be an integer not divisible by p. Then

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}.$$

Proof. Let z be a primitive root mod p and let $a \equiv z^j \bmod p$ for some $0 \leq j \leq p-2$ (as seen in the proof of Proposition 24). Since $z^{p-1} \equiv 1 \bmod p$, we have

$$a^{(p-1)/2} \equiv z^{j(p-1)/2} \equiv \begin{cases} 1 & \text{if } j \text{ is even} \\ z^{(p-1)/2} & \text{if } j \text{ is odd} \end{cases}$$

On the other hand, we know from the proof of Rule 2 that $z^{(p-1)/2} \equiv -1 \bmod p$. Combining this into the mix, we have

$$a^{(p-1)/2} \equiv \begin{cases} 1 & \text{if } j \text{ is even} \\ -1 & \text{if } j \text{ is odd} \end{cases}$$

By Proposition 24, the RHS is exactly the Legendre symbol $\left(\frac{a}{p}\right)$. □

Remark. One can indeed use this proposition to prove Rule 1.

5.2 Solving the equation $x^2 \equiv a \bmod p$

Using Legendre symbol and the quadratic reciprocity, it is possible to quickly determine whether or not the equation

$$x^2 \equiv a \pmod{p}$$

has a solution, when p is an odd prime and a is not divisible by p. It only tells us the existence, or non-existence of a solution, and does not tell us how to find it.

Before we delve into the subject, let us ask ourselves: how many solutions are we expected to find (mod p)? It is the quadratic equation, so there should be max two solutions (mod p). Can we have all the solutions? Suppose $x = z$ is a solution for the equation above. Then $-z$ will automatically be the other solution. To see this, firstly observe that $(-z)^2 = z^2 \equiv a \bmod p$. Note that $-z$ is distinct from z mod p; for if it were, then $2z \equiv 0 \bmod p$ and therefore $z \equiv 0$ (since p and 2 are coprime); and $0 \equiv z^2 \equiv a$ would be a contradiction to the assumption that p does not divide a.

Proposition 28. Let p be a prime congruent to 3 mod 4 (hence $(p+1)/4$ is an integer). Suppose that $\left(\frac{a}{p}\right) = 1$. Then $z = a^{(p+1)/4}$ is a solution to the equation $x^2 \equiv a \bmod p$.

Proof. By Euler's criterion,

$$z^2 = a^{(p+1)/2} = a^{(p-1)/2+1} \equiv \left(\frac{a}{p}\right) a = a.$$

□

Example. Find all solutions z to each of the following equations with $1 \leq z \leq p = 131$:

1. $x^2 \equiv 2 \bmod 131$,
2. $x^2 \equiv 3 \bmod 131$.

Proof. 1) Firstly, we need to know if there are solutions. To this end, we compute the Legendre symbol:

$$\left(\frac{2}{131}\right) = -1$$

since $131 \equiv 3 \bmod 8$ (Rule 3). Hence there are no solutions.

2) Since

$$\left(\frac{3}{131}\right) \overset{R4}{=} -1 \left(\frac{131}{3}\right) \overset{R0}{=} -1 \left(\frac{2}{3}\right) = 1,$$

the equation does have (two) solutions. Using Proposition 28, the solutions are

$$x \equiv \pm 3^{(131+1)/4} \equiv \pm 3^{33}$$

mod 131. We compute 3^{33} mod 131. To see this, $3^4 = 81$, hence $3^8 = 81^2 \equiv 11$ mod 131. As $3^{16} \equiv 11^2 \equiv -10$, $3^{32} \equiv (-10)^2 = 100$ mod 131. Finally,

$$3^{33} \equiv 100 \cdot 3 \equiv 38$$

mod 131. On the other hand,

$$-33^{33} \equiv -300 \equiv -38 \equiv 93$$

mod 131. The solutions for $x^2 \equiv 3$ mod 131 are 38 and 93 mod 131.

Proposition 29. Let p be a prime congruent to 1 mod 4. Suppose that $\left(\frac{a}{p}\right) = -1$. Then $z = a^{(p-1)/4}$ is a solution to the equation $x^2 \equiv -1$ mod p.

Proof. By Euler's criterion,

$$z^2 = a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) = -1.$$

□

Example. Find all solutions z to the equation

$$x^2 \equiv -1 \pmod{229}$$

with $1 \leq z \leq 229$. The first step is to find a quadratic non-residue 'a'. This is done by trial and error. Note that 1 is always a quadratic residue. So let's try 2:

$$\left(\frac{2}{229}\right) \overset{R3}{=} -1$$

since $229 \equiv -3$ mod 8. Hence, using Proposition 29,

$$z = 2^{(229-1)/4} = 2^{57}$$

mod 229 is a solution to the equation. To compute 2^{57}, we observe

$$2^8 \equiv 27,$$

$$2^{16} \equiv 27^2 \equiv 42$$

and

$$2^{32} \equiv 42^2 \equiv 161.$$

It therefore follows that

$$2^{57} = 2^{32+16+8+1} \equiv 161 \cdot 42 \cdot 27 \cdot 2 \equiv 122$$

mod 229. So 122 mod 229 is a solution. Since $-122 \equiv 107$ mod 229, 107 is also a solution.

5.3 Hensel's lemma

If $P(x)$ is a polynomial in $\mathbb{Z}[x]$, then let $P'(x) \in \mathbb{Z}[x]$ denote its first (formal) derivative with respect to x.

Theorem 30. Let p be a prime and $N \geq 1$ be an integer. Suppose that there exists $z \in \mathbb{Z}$ such that $P(z) \equiv 0 \bmod p^N$. If $P'(z)$ is not congruent to $0 \bmod p$, then there exists an integer r, unique mod p, such that $z' = z + rp^N$ satisfies that $P(z') \equiv 0 \bmod p^{N+1}$.

Remark. If z satisfies $P(z) \equiv 0 \bmod p^{N+1}$, it certainly satisfies $P(z) \equiv 0 \bmod p^N$. The theorem proves, under certain conditions, that one can prove the converse, i.e., one can 'lift' a mod p^N solution of the polynomial P to a mod p^{N+1} solution.

Proof. Suppose that P has degree d. Let z be an integer such that $P(z) \equiv 0 \bmod p^N$.

Firstly, the Taylor expansion with respect to z finds $d+1$ integers $c_0, \ldots, c_d$ (some of them could be zero) such that

$$P(x) = \sum_{j=0}^{d} c_j (x-z)^j.$$

For $0 \leq j \leq d$, it is easy to check $P^{(j)}(z) = c_j j!$ and $c_j = P^{(j)}(z)/j!$ is an integer. Substituting back into the expansion, we therefore get

$$P(x) = \sum_{j=0}^{d} \frac{P^{(j)}(z)}{j!} (x-z)^j.$$

Substituting $x = z + rp^N$, we then get

$$P(z + rp^N) = \sum_{j=0}^{d} \frac{P^{(j)}(z)}{j!} (rp^N)^j.$$

It therefore follows that

$$P(z + rp^N) \equiv P(z) + P'(z) rp^N$$

mod p^{2N}. It follows that

$$P(z + rp^N) \equiv 0 \bmod p^{N+1} \text{ if and only if } P(z) \equiv -rp^N P'(z) \bmod p^{N+1}$$

(because terms divisible by p^{2N} are certainly divisible by p^{N+1}). Since $P(z) \equiv 0 \bmod p^N$ by assumption, we may cancel a factor of p^N from this equation so that

$$P(z + rp^N) \equiv 0 \bmod p^{N+1} \text{ if and only if } \frac{P(z)}{p^N} \equiv -rP'(z) \underline{\bmod p}.$$

By assumption, $\gcd(P'(z), p) = 1$ and $P'(z)$ therefore has an inverse $\underline{\bmod p}$; we shall call it $Q'(z)$. It follows that

$$P(z + rp^N) \equiv 0 \bmod p^{N+1} \text{ if and only if } r \equiv -\frac{P(z)}{p^N} Q'(z) \underline{\bmod p}.$$

This proves the existence and uniqueness of $r \bmod p$. $\square$

Remark. The proof explicitly constructs a lift of the mod p^N solution $P(z_N) \equiv 0$ to a mod p^{N+1} solution z_{N+1} by defining it to be

$$z_{N+1} \equiv z_N - P(z_N)Q'(z_N)$$

mod p^{N+1} (where $Q'(z_N)$, as in the proof, is the inverse of $P'(z_N) \bmod p$). Since p^N divides $P(z_N)$ by assumption,

$$z_{N+1} \equiv z_N$$

mod p^N. It is in this sense that z_{N+1} is a lift of z_N.

Example. Find all solutions to $x^2 + 1 \equiv 0 \bmod 5^3$.

(Step 1) Find a solution mod 5. By trial and error, $\pm 2 \bmod 5$ are the solutions to $x^2 + 1 \bmod 5$.

(Step 2) Let $z = 2$ and see if it is possible to lift this mod 5 solution to a mod 5^2 solution, using the 'algorithm' in the proof. Firstly, since $P'(x) = 2x$, we have $P'(z) = 4$ which is manifestly not congruent to 0 mod 5. As a result, $P'(z)$ has an inverse $Q'(z)$; indeed $Q'(z) \equiv 4 \bmod 5$. We know

$$z_1 \equiv z - P(z)Q'(z) \equiv 2 - 5 \cdot 4 = -18 \equiv 7 \bmod 5^2$$

is a mod 5^2 solution.

(Step 3) See if we can lift the mod 5^2 solution z_1 to a mod 5^3 solution. We observe $P(z_1) = P(7) = 7^2 + 1 = 50$ and the inverse $Q'(z_1)$ of $P'(z_1) = P'(7) = 2 \cdot 7 \equiv 4$ is, for example, 4 mod 5. We then know that

$$z_2 \equiv z_1 - P(z_1)Q'(z_1) \equiv 7 - 50 \cdot 4 \equiv 57 \bmod 5^3$$

is a mod 5^3 solution.

To find the other solution, we start with $z = -2$ and we would get $z_2 \equiv -57 \equiv 68 \bmod 5^3$.(Exercise to fill in the argument).

To sum up, 57 and 68 mod 5^3 are the (two) solutions to $x^2 + 1 \equiv 0 \bmod 5^3$, since the Hensel lift is unique and any solution to $x^2 + 1 \equiv 0 \bmod 5^3$ is a solution to $x^2 + 1 \equiv 0 \bmod 5$.

Remark. This process can be iterated to find roots of $P(x) \bmod 5^r$ for any r.

Example. Find all solutions to $x^3 + 10x^2 + x + 3 \equiv 0 \bmod 3^3$.

(Step 1) Find a solution mod 3. Since $x^3 + 10x^2 + x + 3 \equiv x^3 + x^2 + x \bmod 3$, it is easy to see 0 and 1 mod 3 are the solutions (mod 3).

(Step 2) Let $z = 0$ and see if it is possible to lift the mod 3 solution z to a 3^2 solution z_1. As $P'(z) \equiv 1 \bmod 3$, we can lift the mod 3 solution $z = 0$ to a mod 3^2 solution

$$z_1 = z - P(z)Q'(z) \equiv 0 - 3 \cdot 1 \equiv 6 \bmod 3^2.$$

(Step 3) See if mod 3^2 solution $z_1 = 6$ lifts to a 3^3 solution. As $P(z_1) = 6^3 + 10 \cdot 6^2 + 6 + 3 = 585 \equiv 18 \bmod 3^3$ and $P'(z_1) = 3 \cdot 6^2 + 20 \cdot 6 + 1 \equiv 1 \bmod 3$,

$$z_2 \equiv z_1 - P(z_1)Q'(z_1) \equiv 6 - 18 \cdot 1 \equiv 15 \bmod 3^3$$

does the trick.

We would be tempted to carry out the same process starting with $z \equiv 1 \bmod 3$, but $P'(z) \equiv 0$ in this case, and we need to argue differently. If z_1 were a mod 3^2 lift of $z \equiv 1 \bmod 3$, then z_1 would have to be 1 mod 3. Therefore z_1 would be either $1, 4, 7 \bmod 3^2$. However, none of these is a solution to P mod 3^2:

$$P(1) \equiv P(4) \equiv P(7) \equiv 6 \bmod 3^2$$

We therefore conclude that there is no mod 3^2 lift. There is no mod 3^3 lift either, for if there were, it would define a mod 3^2 lift, which, we know, does not exist. In summary, the equation $x^3 + 10x^2 + x + 3 \equiv 0 \bmod 3^3$ has only one solution $15 \bmod 3^3$.

Remark. There is no general behaviour to determine when $P'(x) \equiv 0 \bmod p$.

6 Continued fractions

6.1 Finite continued fractions

Recall the calculation of $\gcd(225, 157) = 1$ in terms of Euclid's algorithm:

$$\begin{aligned} 225 &= 157 \cdot 1 + 68 \\ 157 &= 68 \cdot 2 + 21 \\ 68 &= 21 \cdot 3 + 5 \\ 21 &= 5 \cdot 4 + 1 \\ 5 &= 1 \cdot 5 + 0 \end{aligned}$$

We may interpret these steps into the following:

$$\begin{aligned} \tfrac{21}{5} &= 4 + \tfrac{1}{5} \\ \tfrac{68}{21} &= 3 + \cfrac{1}{4+\frac{1}{5}} \\ \tfrac{157}{68} &= 2 + \cfrac{1}{3+\cfrac{1}{4+\frac{1}{5}}} \\ \tfrac{225}{157} &= 1 + \cfrac{1}{2+\cfrac{1}{3+\cfrac{1}{4+\frac{1}{5}}}} \end{aligned}$$

These expressions are called *continued fractions*.

Definition. For $N \geq 1$, $a, a_1, \ldots, a_{N-1} \in \mathbb{Z}$ and $a_N \in \mathbb{R}$, we will write

$$[a; a_1, \ldots, a_{N-1}, a_N]$$

to mean

$$a + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{N-1} + \cfrac{1}{a_N}}}}.$$

When '$N = 0$', we only allow $[a;]$ to mean a for $a \in \mathbb{Z}$.

By definition,

$$[a; a_1, \ldots, a_N] = [a; a_1, \ldots, a_{N-2}, a_{N-1} + \frac{1}{a_N}] = [a; a_1, \ldots, a_{N-3}, a_{N-2} + \frac{1}{a_{N-1} + \frac{1}{a_N}}] = \cdots .$$

For example,

$$\begin{aligned} \frac{21}{5} &= [4;5] \\ \frac{68}{21} &= [3;4,5] \\ \frac{157}{68} &= [2;3,4,5] \\ \frac{225}{157} &= [1;2,3,4,5] \end{aligned}$$

Remark. Note you cannot 'add' continued fractions: for example, $[1;1] = 1 + \frac{1}{1} = 2$, so $[1:1] + [1:1]$ is 4 (or $[4;]$). On the other hand $[1+1;1+1] = [2;2] = 2 + \frac{1}{2} = \frac{5}{2}$.

Proposition 31. Let $r = s/t$ be a rational number $r > 1$ in its lowest terms, in the sense that $\gcd(s,t) = 1$. Then r can be written as a continued fraction $[a; a_1, a_2, \ldots, a_N]$ for some $a, a_1, \ldots, a_N \in \mathbb{N}$ with $a_N > 1$. If $t = 1$ (i.e. $r = s$ is an integer), then the continued faction is just $[s;]$.

Remark. Conversely, any sequence $a, a_1, \ldots, a_N$ of positive integers with $a_N > 1$, defines a unique rational number > 1: when $N = 0$,

$$[a;] = a = a_N > 1$$

and, when $N > 0$,

$$[a; a_1, \ldots, a_N]$$

may be defined inductively

$$[a_j; a_{j+1}, \ldots, a_N] = a_j + \frac{1}{[a_{j+1}; a_{j+2}, \ldots, a_N]}$$

as j assumes $N-1, N-2, \ldots, 0$, starting with $[a_{N-1}; a_N] = a_{N-1} + \dfrac{1}{a_N}$. By induction, for every $0 \leq j \leq N-1$, $[a_j; a_{j+1}a_{j+1}, \ldots, a_N]$ is a rational number > 1. Indeed, since $a_{N-1} \geq 1$ and $\dfrac{1}{a_N} > 0$, it follows that $[a_{N-1}; a_N] = a_{N-1} + \dfrac{1}{a_N} > 1$ and it is evidently a rational number. Suppose that $[a_{j+1}; a_{j+1}, \ldots, a_N]$ is a rational number > 1. Then since $a_j \geq 1$ and $\dfrac{1}{[a_{j+1}; a_{j+1}, \ldots, a_N]}$, it follows that $[a_j; a_{j+1}, \ldots, a_N] = a_j + \dfrac{1}{[a_{j+1}; a_{j+1}, \ldots, a_N]} > 1$.

Proof. We run the Euclid's algorithm:

$$\begin{aligned}
s = at + t_1 &\quad\rightsquigarrow\quad \frac{s}{t} = a + \frac{t_1}{t} \\
t = a_1 t_1 + t_2 &\quad\rightsquigarrow\quad \frac{t}{t_1} = a_1 + \frac{t_2}{t_1} \\
t_1 = a_2 t_2 + t_3 &\quad\rightsquigarrow\quad \frac{t_1}{t_2} = a_2 + \frac{t_3}{t_2} \\
&\quad\cdots \\
t_{N-2} = a_{N-1} t_{N-1} + 1 &\quad\rightsquigarrow\quad \frac{t_{N-2}}{t_{N-1}} = a_{N-1} + \frac{1}{t_{N-1}} \\
t_{N-1} = a_N \cdot 1 &\quad\rightsquigarrow\quad t_{N-1} = a_N
\end{aligned}$$

We have

$$1 < t_{N-1} < \cdots < t_1 < t < s,$$

and substituting these all, we deduce

$$r = \frac{s}{t} = a + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_{N-1} + \cfrac{1}{a_N}}}},$$

in other words, $r = [a; a_1, \ldots, a_N]$. □

The following algorithm may be more useful: following the notation from the proof of Proposition 31, we firstly let

$$\rho = \frac{s}{t} (= r), \quad \rho_1 = \frac{t}{t_1}, \ldots, \quad \rho_j = \frac{t_{j-1}}{t_j}, \ldots, \rho_{N-1} = \frac{t_{N-2}}{t_{N-1}} \quad , \rho_N = t_{N-1}$$

and strive to relate ρ_j to ρ_{j+1}; this leads to an algorithm computing a_j's and ρ_j's without involving t_j's (this somehow knocks off repeated/redundant t_j's in the process).

Firstly, note that, by definition, apart from $j = N$, ρ_j is NOT an integer; it is a rational number strictly greater than 1 (because $\rho_j = \dfrac{t_{j-1}}{t_j}$ and $t_j < t_{j-1}$).

Secondly, as we have

$$\frac{t_{j-1}}{t_j} = a_j + \frac{t_{j+1}}{t_j},$$

i.e.,

$$\rho_j = a_j + \frac{1}{a_{j+1}}$$

for $0 \le j \le N-1$ (where we let $t_{-1} = s, t_0 = t, a_0 = a$). It therefore follows that

$$\rho_{j+1} = \frac{1}{\rho_j - a_j}$$

for $0 \le j \le N-1$.

The following algorithm bypasses the Euclid's algorithm and only keeps track of the ρ_j's (with the goal of computing a_j's); and the process stops when it reaches an integer ρ_N.

Definition. For ρ in $\mathbb{R}$, we let $\lfloor \rho \rfloor$ denote the largest integer N satisfying $N \le \rho$.

Example. For $0 \le j \le N-1$,

$$a_j = \lfloor \rho_j \rfloor.$$

This follows simply from $\rho_j = a_j + \dfrac{1}{\rho_{j+1}}$ and $\rho_{j+1} > 1$. Furthermore,

$$a_N = \lfloor \rho_N \rfloor = \rho_N.$$

Then

$$\begin{array}{rcl}
a = \lfloor r \rfloor = \lfloor \rho \rfloor & \Rightarrow & \rho_1 = \dfrac{1}{\rho - a} \\
\swarrow & & \\
a_1 = \lfloor \rho_1 \rfloor & \Rightarrow & \rho_2 = \dfrac{1}{\rho_1 - a_1} \\
\swarrow & & \\
\vdots & & \\
\swarrow & & \\
a_{N-1} = \lfloor \rho_{N-1} \rfloor & \Rightarrow & \rho_N = \dfrac{1}{\rho_{N-1} - a_{N-1}} \in \mathbb{N} \\
\swarrow & & \\
a_N = \lfloor \rho_N \rfloor = \rho_N & &
\end{array}$$

Example. $r = \dfrac{87}{38}$.

$$\begin{array}{rcl}
a = \lfloor \frac{87}{38} \rfloor = 2 & \Rightarrow & \rho_1 = \dfrac{1}{\frac{87}{38} - 2} = \dfrac{38}{11} \\
\swarrow & & \\
a_1 = \lfloor \frac{38}{11} \rfloor = 3 & \Rightarrow & \rho_2 = \dfrac{1}{\frac{38}{11} - 2} = \dfrac{11}{5} \\
\swarrow & & \\
a_2 = \lfloor \frac{11}{5} \rfloor = 2 & \Rightarrow & \rho_3 = \dfrac{1}{\frac{11}{5} - 2} = \dfrac{5}{1} \in \mathbb{N} \\
\swarrow & & \\
a_3 = \lfloor \frac{5}{1} \rfloor = 5 = \rho_3 & &
\end{array}$$

Hence $r = [a; a_1, a_2, a_3] = [2; 3, 2, 5]$.

Remark. *This algorithm allows us to compute the continued fraction for a non-positive rational number*– the only difference from ones for positive rational number is that we need to allow non-positive a's as a result): let us compute $r = -\frac{3}{5}$ following the algorithm:

$$\begin{array}{rcl} a = \lfloor -\frac{3}{5} \rfloor = -1 & \Rightarrow & \rho_1 = \frac{1}{-\frac{3}{5} - (-1)} = \frac{5}{2} \\ & \swarrow & \\ a_1 = \lfloor \frac{5}{2} \rfloor = 2 & \Rightarrow & \rho_2 = \frac{1}{\frac{5}{2} - 2} = \frac{2}{1} \in \mathbb{N} \\ & \swarrow & \\ a_2 = \lfloor \frac{2}{1} \rfloor = 2 = \rho_2 & & \end{array}$$

Hence $-\frac{3}{5} = [-1; 2, 2]$. On the other hand, $r = \frac{3}{5}$ is computed by

$$\begin{array}{rcl} a = \lfloor \frac{3}{5} \rfloor = 0 & \Rightarrow & \rho_1 = \frac{1}{\frac{3}{5} - 0} = \frac{5}{3} \\ & \swarrow & \\ a_1 = \lfloor \frac{5}{3} \rfloor = 1 & \Rightarrow & \rho_2 = \frac{1}{\frac{5}{3} - 1} = \frac{3}{2} \\ & \swarrow & \\ a_2 = \lfloor \frac{3}{2} \rfloor = 1 & \Rightarrow & \rho_3 = \frac{1}{\frac{3}{2} - 1} = \frac{2}{1} \in \mathbb{N} \\ & \swarrow & \\ a_3 = \lfloor \frac{2}{1} \rfloor = 2 = \rho_3 & & \end{array}$$

Hence $\frac{3}{5} = [0; 1, 1, 2]$.

When the last term of the continued fraction expression satisfies $a_N > 1$, it is possible to prove the uniqueness of the expression:

Theorem 32 If r is a rational number with

$$r = [a; a_1, \ldots, a_k] = [\beta; \beta_1, \ldots, \beta_\ell]$$

where $a, a_1, \ldots, a_k, \beta, \beta_1, \ldots, \beta_\ell$ are non-negative integers such that $a_k > 1$ and $\beta_\ell > 1$, then $k = \ell$ and $a_j = \beta_j$ for every j.

Proof. We prove this by induction on k.

Suppose $k = 0$. Then $r = a$ is an integer. If $\ell > 0$, then $r = \beta + \frac{1}{[\beta_1; \beta_2, \ldots, \beta_\ell]}$ with its fraction $0 < \frac{1}{[\beta_1; \beta_2, \ldots, \beta_\ell]} < 1$ (this is where $\beta_\ell > 1$ is used); this is impossible. It therefore follows that $k = \ell$ and $r = a = \beta$.

Suppose that the assertion holds, with $k-1$ in place of k. By assumption, we know

$$a + \frac{1}{[a_1; a_2, \ldots, a_k]} = \beta + \frac{1}{[\beta_1; \beta_2, \ldots, \beta_\ell]}.$$

Since $a_k > 1$ and $\beta_\ell > 1$, the fractions are both less than 1 (this follows from Proposition 31) and we may deduce $a = \lfloor r \rfloor = \beta$. It then follows that

$$[a_1; a_2, \ldots, a_k] = [\beta_1; \beta_2, \ldots, \beta_\ell]$$

and it follows from the inductive hypothesis that $k-1 = \ell - 1$ and $a_j = \beta_j$ for every $1 \leq j \leq k$. □

Following Proposition 31, we define:

Definition. Let a be an integer and $a_1, \ldots, a_N$ be integers > 1. For $0 \leq n \leq N$,

$$r_n = [a; a_1, \ldots, a_n].$$

is a rational number and $\{r_0, \ldots, r_N\}$ are called the convergents of the continued fraction $r = [a; a_1, \ldots, a_N]$.

Remark. This seems like a misnomer to call them 'convergents', but we will see in the next section that the convergents do converge when r is an irrational number.

From the definition, you might be tempted to think that the r_n's are increasing sequence. They are NOT! On the other hand, it is very hard to keep track of how the r_n behave in terms of the definition we have just seen. To this end, we introduce the following:

Definition. Given an integer a and integers $a_1 > 1, \ldots, a_N > 1$, we define: $s_{-2} = 0, s_{-1} = 1, s_0 = a$ and, for $n = 1, \ldots, N$,

$$s_n = a_n s_{n-1} + s_{n-2}$$

and define: $t_{-2} = 1, t_{-1} = 0, t_0 = 1$ and, for $n = 1, \ldots, N$,

$$t_n = a_n t_{n-1} + t_{n-2}.$$

Remark. If $a > 0$, then the s_n and t_n are both strictly increasing sequences of positive integers. One can see this by induction on n. Suppose that every s_j, for $j \leq n-1$, is a positive integer. It then follows from $a_n > 1$ that

$$s_n = a_n s_{n-1} + s_{n-2} > a_n s_{n-1} > s_{n-1}.$$

Similarly for the t_n.

Proposition 33. $r_n = \dfrac{s_n}{t_n}$ for every $0 \leq n \leq N$.

Remark Even if both the s_n's and the t_n's are strictly increasing, it does not mean that their ratios r_n's are!

Proof. We prove this by induction on n. By definition, $r_0 = \dfrac{s_0}{t_0} = a$. Suppose that $r_n = \dfrac{s_n}{t_n}$ holds (we would like to establish that $r_{n+1} = \dfrac{s_{n+1}}{t_{n+1}}$). By definition,

$$r_n = [a; a_1, \ldots, a_n] = \frac{a_n s_{n-1} + s_{n-2}}{a_n t_{n-1} + t_{n-2}}.$$

It then follows that

$$r_{n+1} = [a; a_1, \ldots, a_n + \frac{1}{a_{n+1}}] = \frac{\left(a_n + \dfrac{1}{a_{n+1}}\right) s_{n-1} + s_{n-2}}{\left(a_n + \dfrac{1}{a_{n+1}}\right) t_{n-1} + t_{n-2}} = \frac{a_{n+1}(a_n s_{n-1} + s_{n-2}) + s_{n-1}}{a_{n+1}(a_n t_{n-1} + t_{n-2}) + t_{n-1}} = \frac{a_{n+1} s_n + s_{n-1}}{a_{n+1} t_n + t_{n-1}} = \frac{s}{t}$$

□

We may now compute the convergents r_n in terms of the s_n's and t_n's.

Example. Compute the convergents of $[3; 7, 15, 1] = 3.1415929203\ldots$ (this is actually a 'truncated' infinite continued fraction $[3; 7, 15, 1, 292, 1, \ldots]$ of $\pi = 3.141592653589793\ldots$).

On one hand,

$$\begin{aligned} s_{-1} &= 1 \\ s_0 &= 3 \\ s_1 &= a_1 s_0 + s_{-1} = 7 \cdot 3 + 1 = 22 \\ s_2 &= a_2 s_1 + s_0 = 15 \cdot 22 + 3 = 333 \\ s_3 &= a_3 s_2 + s_1 = 1 \cdot 333 + 22 = 355. \end{aligned}$$

On the other hand,

$$\begin{aligned} t_{-1} &= 0 \\ t_0 &= 1 \\ t_1 &= a_1 t_0 + t_{-1} = 7 \cdot 1 + 0 = 7 \\ t_2 &= a_2 t_1 + t_0 = 15 \cdot 7 + 1 = 106 \\ t_3 &= a_3 t_2 + t_1 = 1 \cdot 106 + 7 = 113. \end{aligned}$$

Hence $r_1 = \dfrac{22}{7} = 3.14285714\ldots, r_2 = \dfrac{333}{106} = 3.1415094\ldots, r_3 = \dfrac{355}{113} = 3.141592...$

Theorem 34. Following the notation above,

- $s_n t_{n-1} - t_n s_{n-1} = (-1)^{n-1}$ for $n \geq 1$,
- $r_n - r_{n-1} = \dfrac{(-1)^{n-1}}{t_{n-1} t_n}$ for $n \geq 1$,
- $\gcd(s_n, t_n) = 1$.

Proof. To prove the first assertion, we use induction on n. To recall,

$$\begin{aligned} s_0 &= a, \\ s_1 &= aa_1 + 1, \\ t_0 &= 1, \\ t_1 &= a_1. \end{aligned}$$

Hence $s_1t_0 - t_1s_0 = (aa_1 + 1) - a_1a = 1 = (-1)^0$.

Suppose that $s_{n-1}t_{n-2} - t_{n-1}s_{n-2} = (-1)^{n-2}$ holds (we would like to prove $s_nt_{n-1} - t_ns_{n-1} = (-1)^{n-1}$). Since $s_n = a_ns_{n-1} + s_{n-2}$ and $t_n = a_nt_{n-1} + t_{n-2}$, we have

$$\begin{aligned} s_nt_{n-1} - t_ns_{n-1} &= (a_ns_{n-1} + s_{n-2})t_{n-1} - (a_nt_{n-1} + t_{n-2})s_{n-1} \\ &= s_{n-2}t_{n-1} - t_{n-2}s_{n-1} \\ &= -(-1)^{n-2} \\ &= (-1)^{n-1}. \end{aligned}$$

The second assertion follows immediately from $r_n = \dfrac{s_n}{t_n}$, $r_{n-1} = \dfrac{s_{n-1}}{t_{n-1}}$ and the first assertion.

The third assertion follows immediately from the first. □

Corollary 35. The convergents r_n's satisfy

$$r_0 < r_2 < r_4 < \cdots < r_5 < r_3 < r_1.$$

Proof. Applying Theorem 34 twice, we get

$$r_{n+2} - r_n = (r_{n+2} - r_{n+1}) + (r_{n+1} - r_n) = \frac{(-1)^{n+1}}{t_{n+2}t_{n+1}} + \frac{(-1)^n}{t_{n+1}t_n} = \frac{(-1)^n}{t_{n+2}t_{n+1}t_n}(t_{n+2} - t_n) = \frac{(-1)^n a_{n+2}}{t_{n+2}t_n}$$

since, by definition, $a_{n+2} = \dfrac{t_{n+2} - t_n}{t_{n+1}}$ at the last equality.

If n is even (resp. odd), the RHS is positive (resp. negative), so $r_{n+2} > r_n$ (resp. $r_n > r_{n+2}$).

It remains to show that

$$r_{2i} < r_{2j-1}$$

for every $i \geq 0$ and $j \geq 1$. Since the 'even' convergents increase, while 'odd' ones decrease, it follows from Theorem 34 to get $r_N - r_{N-1} = (-1)^{N-1}/t_{N-1}t_N < 0$ when $N = 2i + 2j$ is evidently even, which yields

$$r_{2i} < r_{2i+2j} < r_{2i+2j-1} < r_{2j-1},$$

as desired. □

6.2 Infinite continued fractions

As promised:

Theorem 36. Let $a, a_1, \ldots$ be a sequence of integers such that $a_n > 0$ if $n \geq 1$. Define, for every $n \geq 0$,

$$r_n = [a; a_1, \ldots, a_n].$$

Then the sequence $r_0, r_1, r_2, \ldots$, of rational numbers converges to a limit (not necessarily a rational number).

Proof. Since the $r_0, r_1, \ldots, r_n$ are the convergents to the finite continued fraction $[a; a_1, \ldots, a_n]$ for any fixed $n \geq 0$, all the results in the preceding section apply results of the preceding section:

- $r_n = \dfrac{s_n}{t_n}$,
- (Corollary 35) $r_0 < r_2 < r_4 < \cdots < r_5 < r_3 < r_1$,
- (Theorem 36) $r_n - r_{n-1} = \dfrac{(-1)^{n-1}}{t_{n-1}t_n}$.

Since the even terms

$$r_0, r_2, r_4, \ldots,$$

form an increasing sequence bounded above by r_1, it tends to a limit ρ. On the other hand, the odd terms

$$r_1, r_3, r_5, \ldots,$$

form a decreasing sequence bounded below by r_0 and it tends to a limit ρ'. Since $r_{2j} < r_{2j+1}$ (Theorem 34),

$$\rho = \lim_{j\to\infty} r_{2j} \leq \lim_{j\to\infty} r_{2j+1} = \rho'.$$

[Note that $\leq$ is not a typo; even if $r_{2j} < r_{2j+1}$ is maintained for every j, there is no way of knowing a priori this continues to hold in the limit] On the other hand,

$$|r_N - r_{N-1}| = \left|\frac{(-1)^{N-1}}{t_{N-1}t_N}\right| \to 0$$

as $N = 2j$ tends to ∞. It therefore follows

$$|\rho - \rho'| = |(\rho - r_N) + (r_N - r_{N-1}) + (r_{N-1} - \rho')| \leq |\rho - r_N| + |r_N - r_{N-1}| + |\rho' - \rho_{N-1}| \to 0$$

and one can deduce $\rho = \rho'$. □

Definition. We define the limit of the sequence of convergents to be the *value* of the infinite continued fraction $[a; a_1, \ldots]$.

We show that every real number (not just a rational number) has a continued fraction expansion:

Theorem 37. For every irrational number r, there exists a sequence of integers $a_0, a_1, \ldots$ with $a_n > 0$ if $n \geq 1$ such that the value of $[a; a_1, \ldots]$ is r.

Proof. Let $\rho_0 = r$ and $a = \lfloor \rho_0 \rfloor \in \mathbb{Z}$ so that $0 < \rho_0 - a < 1$. The number $\rho_1 = \dfrac{1}{\rho_0 - a} > 1$ is irrational (if not, r would be rational). We may continue this process: starting with an irrational number $\rho_n > 1$, we let $a_n = \lfloor \rho_n \rfloor \in \mathbb{N}$ and let ρ_{n+1} denote the irrational number $\dfrac{1}{\rho_n - a_n} > 1$. Then

$$a_1, a_2, \ldots,$$

are positive integers and

$$\rho_1, \rho_2, \ldots,$$

are irrational numbers > 1. The process continues without an end and $[a; a_1, \ldots,]$ define an infinite continued fraction (This 'algorithm' is called the *continued fraction algorithm*). It remains to check that the value of this continued fraction is indeed r (that we started with).

We prove by induction on $n \geq 1$ that

$$[a; a_1, \ldots, a_{n-1}, \rho_n] = r.$$

When $n = 1$, $r = a + \rho_0 - a = a + \dfrac{1}{\rho_1} = [a; \rho_1]$.

Suppose that $[a; a_1, \ldots, a_{n-1}, \rho_n] = r$ holds for $n \geq 1$ (we would like to show $[a; a_1, \ldots, a_n, \rho_{n+1}] = r$ holds). Since $\dfrac{1}{\rho_{n+1}} = \rho_n - a_n$,

$$\begin{aligned}
& [a; a_1, \ldots, a_n, \rho_{n+1}] \\
= \; & a + \cfrac{1}{a_1 + \cfrac{1}{\ddots\ a_{n-1} + \cfrac{1}{a_n + \cfrac{1}{\rho_{n+1}}}}} \\
= \; & a + \cfrac{1}{a_1 + \cfrac{1}{\ddots\ a_{n-1} + \cfrac{1}{a_n + \rho_n - a_n}}} \\
= \; & a + \cfrac{1}{a_1 + \cfrac{1}{\ddots\ a_{n-1} + \cfrac{1}{\rho_n}}} \\
= \; & [a; a_1, \ldots, a_{n-1}, \rho_n] \\
= \; & r,
\end{aligned}$$

the claim thus follows.

Let

$$\{r_n = [a; a_1, \ldots, a_n]\}$$

be the convergents of $[a; a_1, \ldots,]$. We know that the convergents tend to a limit ρ (by Theorem 36). We need to show that $\rho = r$, i.e., the continued fraction algorithm correctly defines the continued fraction of r. We know from the proof of Theorem 36 that

$$\rho \geq r_{2j}$$

and

$$r_{2j+1} \geq \rho,$$

it suffices to establish the same set of inequalities with ρ replaced by r holds. Indeed, if this is the case, letting $N = 2j$ for example,

$$|r - \rho| \leq |r_N - r_{N-1}| \to 0$$

as $j \to \infty$.

To prove that $r \geq r_{2j}$ and $r_{2j+1} \geq r$, we argue as follows.

Suppose n is even. From the algorithm, we see that $\lfloor \rho_n \rfloor = a_n$ so $a_n < \rho_n$. It therefore follows from the lemma below that

$$r_n = [a; a_1, \ldots, a_{n-1}, a_n] < [a; a_1, \ldots, a_{n-1}, \rho_n] = r.$$

Then case when n is odd follows similarly. □

Lemma 38. Suppose that $\gamma < \gamma'$ for positive real numbers γ, γ'. Then, for $n \geq 1$,

- $[a; a_1, \ldots, a_{n-1}, \gamma] < [a; a_1, \ldots, a_{n-1}, \gamma']$ when n is even,
- $[a; a_1, \ldots, a_{n-1}, \gamma] > [a; a_1, \ldots, a_{n-1}, \gamma']$ when n is odd.

Proof. Prove by induction on n. When $n = 1$,

$$a + \frac{1}{\gamma} = [a; \gamma] > [a; \gamma'] = a + \frac{1}{\gamma'}$$

holds because of the assumption $\gamma < \gamma'$. Suppose that the assertion holds with $n - 1$ in place of n.

Suppose firstly that $\boxed{n \text{ is even}}$ (i.e. $n - 1$ is odd). It then follows that

$$[a; a_1, \ldots, a_{n-1}, \gamma] = a + \frac{1}{[a_1; a_2, \ldots, a_{n-1}, \gamma]} < a + \frac{1}{[a_1; a_2, \ldots, a_{n-1}, \gamma']} = [a; a_1, \ldots, a_{n-1}, \gamma']$$

because by the inductive hypothesis we know

$$[a_1; a_2, \ldots, a_{n-1}, \gamma] = [\beta; \beta_1, \ldots, \beta_{n-2}, \gamma] > [\beta; \beta_1, \ldots, \beta_{n-2}, \gamma'] = [a_1; a_2, \ldots, a_{n-1}, \gamma'].$$

The case $\boxed{n \text{ is odd}}$ is similar. □

Remark. The proof is constructive. The algorithm is exactly the same as the one for finite continued fractions (for positive/negative rational numbers).

Example. What is the continued fraction of $\pi = 3.141592653589793$? Applying the continued fraction algorithm to get

$$
\begin{array}{rcl}
a = \lfloor 3.141592653589793 \rfloor = 3 & \Rightarrow & \rho_1 = \dfrac{1}{0.141592653589793} = 7.062513305931046 \\
& \swarrow & \\
a_1 = \lfloor 7.062513305931046 \rfloor = 7 & \Rightarrow & \rho_2 = \dfrac{1}{0.06251330593104577} = 15.99659440668572 \\
& \swarrow & \\
a_2 = \lfloor 15.99659440668572 \rfloor = 15 & \Rightarrow & \rho_3 = \dfrac{1}{0.99659440668572} = 1.003417231013372 \\
& \swarrow & \\
a_3 = \lfloor 1.003417231013372 \rfloor = 1 & \Rightarrow & \rho_4 = \dfrac{1}{0.003417231013372} = 292.6345910144503 \\
& \swarrow & \\
a_4 = \lfloor 292.6345910144503 \rfloor = 292 & \Rightarrow & \rho_5 = \dfrac{1}{0.6345910144503} = 1.575818089492172 \\
& \swarrow & \\
& \vdots &
\end{array}
$$

so the continued fraction looks like $[3; 7, 15, 1, 292, 1, \dots]$. The convergents are

$$3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103993}{33102}, \dots$$

The numbers are $\dfrac{22}{7}$ and $\dfrac{355}{113}$ are well-known approximations to π!

Example $r = 1 + \dfrac{1}{2}\sqrt{2}$.

$$
\begin{array}{rcl}
a = \lfloor 1 + \frac{1}{2}\sqrt{2} \rfloor = 1 & \Rightarrow & \rho_1 = \dfrac{1}{(1 + \frac{1}{2}\sqrt{2}) - 1} = \sqrt{2} \\
& \swarrow & \\
a_1 = \lfloor \sqrt{2} \rfloor = 1 & \Rightarrow & \rho_2 = \dfrac{1}{\sqrt{2} - 1} = 1 + \sqrt{2} \\
& \swarrow & \\
a_2 = \lfloor 1 + \sqrt{2} \rfloor = 2 & \Rightarrow & \rho_3 = \dfrac{1}{(1 + \sqrt{2}) - 2} = 1 + \sqrt{2} = \rho_2 \\
& \swarrow & \\
a_3 = a_2 & \Rightarrow & \rho_4 = \rho_3 = \rho_2 \\
& \swarrow & \\
& \vdots &
\end{array}
$$

Hence $1 + 1 + \dfrac{1}{2}\sqrt{2}$ is the value of $[1; 1, 2, 2, \dots]$.

Example. $r = \sqrt{15} - 3$.

$$
\begin{array}{rcl}
a = \lfloor \sqrt{15} - 3 \rfloor = 0 & \Rightarrow & \rho_1 = \dfrac{1}{\sqrt{15}-3} = \dfrac{\sqrt{15}+3}{6} \\
\swarrow & & \\
a_1 = \lfloor \dfrac{\sqrt{15}+3}{6} \rfloor = 1 & \Rightarrow & \rho_2 = \dfrac{1}{(\frac{\sqrt{15}+3}{6}) - 1} = \sqrt{15}+3 \\
\swarrow & & \\
a_2 = \lfloor \sqrt{15}+3 \rfloor = 6 & \Rightarrow & \rho_3 = \dfrac{1}{(\sqrt{15}+3)-6} = \dfrac{1}{\sqrt{15}-3} = \rho_1 \\
\swarrow & & \\
a_3 = a_1 & \Rightarrow & \rho_4 = \rho_2 \\
\swarrow & & \\
a_4 = a_2 & \Rightarrow & \rho_5 = \rho_3 = \rho_1 \\
\swarrow & & \\
\vdots & &
\end{array}
$$

Hence $\sqrt{15} - 3$ is the value of $[0; 1, 6, 1, 6, \dots]$.

Example. The golden ratio $r = \dfrac{1+\sqrt{5}}{2}$. This is the value of $[1; 1, 1, \dots]$.

Finally, we show that the continued fraction expression is unique:

Theorem 39. Every irrational number is the value of a unique infinite continued fraction.

Proof. Let r be an irrational number. By Theorem 37, we may write r as $[a; a_1, \dots,]$. The goal is to show that $a, a_1, \dots$ are uniquely determined by r.

Then $r = a + \dfrac{1}{\rho_1}$ with $\rho_1 > 1$ is irrational; so $a = \lfloor r \rfloor$ and $\rho_1 = \dfrac{1}{r-a}$ are uniquely determined by r. Similarly, $a_1 = \lfloor \rho_1 \rfloor$ and $\rho_2 = \dfrac{1}{\rho_1 - a_1}$ are uniquely determined, and so on. □

Remark. Given Theorem 37, the proof of Theorem 39 seems redundant but note that Theorem 37 only proves that there *is* a way (an algorithm) to find a continued fraction expression for r (but it fails to prove that it is *the* way). Even though a is obtained as $\lfloor r \rfloor$ in Theorem 37, it does not automatically mean that, whatever method we come up with or decide to take, a– the first integer its expression– should *always* be $\lfloor r \rfloor$. Theorem 39 proves that it always should be.

To sum up, there is a bijection between

- the set of irrational numbers
- the set of *infinite* continue fractions $[a; a_1, \dots]$, where $a \in \mathbb{Z}$ and $a_1, \dots \in \mathbb{N}$.

Given an irrational number, the continued fraction algorithm (See the proof of Theorem **??**) defines an infinite continued fraction. Conversely, given a continued fraction $[a; a_1, \dots,]$, we may calculate the convergents $r_n = [a; a_1, \dots]$ and their limit gives the corresponding irrational number.

6.3 Periodic continued fraction

Definition. An infinite continued fraction $[a; a_1, \dots]$ is *periodic* if there exists l, N with $l > 0$ such that the sequence stabilizes at the N-th repeats itself with cycle of length l:

$$\begin{array}{ccccccc} a_N & = & a_{N+l} & = & a_{N+2l} & = & \cdots, \\ a_{N+1} & = & a_{N+l+1} & = & a_{N+2l+1} & = & \cdots, \\ \vdots & & \vdots & & \vdots & & \cdots \\ a_{N+l-1} & = & a_{N+2l-1} & = & a_{N+3l-1} & = & \cdots, \end{array}$$

or more succinctly

$$a_{n+l} = a_n$$

for all $n \geq N$; and we write the continued fraction as

$$[a; a_1, \dots, a_{N-1}, \overline{a_N, a_{N+1}, \dots, a_{N+l-1}}]$$

It is said to be *purely periodic* if $N = 0$, i.e.,

$$a_{n+l} = a_n$$

for all $n \geq 0$.

Example. $[2; 1, 2, 1, 2, 1, 2, \dots] = [\overline{2; 1}]$. What (irrational) real number r does this continued fraction represent? By definition, it suffices to solve the equation

$$x = [2; 1, x] = 2 + \frac{1}{1 + \frac{1}{x}} = 2 + \frac{x}{x+2} = \frac{3x+2}{x+1},$$

which is $x^2 - 2x - 2 = 0$. The solution (in $\mathbb{R}$) of the equation is $1 \pm \sqrt{3}$ but the expression suggests $r > 2$. So $r = 1 + \sqrt{3}$.

Example. What about $s = [3; 5, 2, 1, 2, 1, 2, 1, 2, \dots] = [3; 5, \overline{2, 1}]$? Let r denote the $\overline{2, 1}$-bit which we know, from the first example, to be $1 + \sqrt{3}$, so that

$$\begin{array}{rcl} s & = & [3; 5, r] \\ & = & 3 + \dfrac{1}{5 + \dfrac{1}{r}} \\ & = & 3 + \dfrac{r}{5r+1} \\ & = & \dfrac{16r+1}{5r+1} \\ & = & \dfrac{126 - \sqrt{3}}{39} \end{array}$$

These two real numbers are irrational while satisfying quadratic equations (s satisfies $(39s - 126)^2 = 3$).

It does not have be concrete numbers:

Example. Let n be a positive integer. Then

$$\sqrt{n^2+1} = [n; \overline{2n}].$$

To see this, we simply run the algorithm:

$$\begin{array}{rcl}
a = \lfloor \sqrt{n^2+1} \rfloor = n & \Rightarrow & \rho_1 = \dfrac{1}{\sqrt{n^2+1}-n} = \sqrt{n^2+1}+n \\
& \swarrow & \\
a_1 = \lfloor \sqrt{n^2+1}+n \rfloor = 2n & \Rightarrow & \rho_2 = \dfrac{1}{(\sqrt{n^2+1}+n)-2n} = \dfrac{1}{\sqrt{n^2+1}-n} = \sqrt{n^2+1}+n = \rho_1 \\
& \swarrow & \\
a_2 = a_1 & \Rightarrow & \rho_3 = \rho_2 = \rho_1 \\
& \swarrow & \\
& \vdots &
\end{array}$$

hence $\sqrt{n^2+1} = [n, 2n, 2n, \dots] = [n, \overline{2n}]$.

Conversely, if we are given $[n, \overline{2n}]$, can we work out the value r of the continued fraction? Let $s = [\overline{2n}] = [2n, 2n, \dots]$. By definition,

$$s = 2n + \frac{1}{s},$$

i.e., $s^2 - 2ns - 1 = 0$. Solving this equation in r, we have $s = \dfrac{2n \pm \sqrt{4n^2+4}}{2} = n \pm \sqrt{n^2+1}$. Since $s > 0$, it follows that $s = n + \sqrt{n^2+1}$. To compute r, we compute

$$r = n + \frac{1}{s} = n + \frac{1}{n+\sqrt{n^2+1}} = \frac{n^2 + n\sqrt{n^2+1}+1}{n+\sqrt{n^2+1}} = \frac{\sqrt{n^2+1}\left(n+\sqrt{n^2+1}\right)}{n+\sqrt{n^2+1}} = \sqrt{n^2+1}.$$

6.4 Diophantine approximation

We have seen that if r is the value of $[a; a_1, \dots]$ and $r_n = [a; a_1, \dots, a_n]$ is the n-th convergent to r, then the numbers r_n are rational numbers that tend to the limit r. In this section, we establish that they give best possible approximations to r.

What should be a *good* rational approximation $\dfrac{s}{t}$ to r?

- it should be close to r (of course),
- the denominator t is relatively small; there should be no rational number, with smaller denominator, that is close to r.

The goal of this section is to establish that the convergents to the continued fraction for r indeed satisfy these properties.

To this end, recall:

- $$r_n = \frac{s_n}{t_n},$$

  where

  $$s_n = a_n s_{n-1} + s_{n-2}$$

  for $n \geq 1, s_0 = a,\ s_{-1} = 1,\ s_{-2} = 0,$

  $$t_n = a_n t_{n-1} + t_{n-2},$$

  for $n \geq 1, t_0 = 1,\ t_{-1} = 0,\ t_{-2} = 1;$

- (Corollary 35) $r_0 < r_2 < r_4 < \cdots < r < \cdots < r_5 < r_3 < r_1$;

- if ρ (resp. ρ') is the limit $\lim_{j\to\infty} r_{2j}$ (resp. $\lim_{j'\to\infty} r_{2j'+1}$) of the strictly increasing (resp. decreasing) sequence $\{r_{2j}\}$ (resp. $\{r_{2j'+1}\}$), then

  $$\rho = \rho'.$$

- $|r - r_n| < |r_{n+1} - r_n|$ for every n. To see this, note that if n is even, then $r_n < r < r_{n+1}$, hence $|r - r_n| < |r_n - r_{n+1}|$; similarly if n is even, then $r_{n+1} < r < r_n$ and $|r - r_n| < |r_n - r_{n+1}|$.

Since Theorem 34 asserts $|r_{n+1} - r_n| = \dfrac{1}{t_n t_{n+1}}$, it follows that

$$|r - r_n| < |r_{n+1} - r_n| = \frac{1}{t_n t_{n+1}}.$$

Example. We saw that $r = \sqrt{15} - 3$ is the value of $[0; 1, 6, 1, 6, \dots]$. The convergents for $\sqrt{15} - 3$ are

$$r_0 = \frac{0}{1}, r_1 = \frac{s_1}{t_1} = \frac{1}{1}, r_2 = \frac{s_2}{t_2} = \frac{6}{7}, r_3 = \frac{7}{8}, r_4 = \frac{s_4}{t_4} = \frac{48}{55}, r_5 = \frac{s_5}{t_5} = \frac{55}{63}, r_6 = \frac{s_6}{t_6} = \frac{378}{433}, \dots.$$

How good is r_6, as a (rational) approximation to r?

$$|r - r_6| \leq \frac{1}{t_6 t_7} = \frac{1}{433 \cdot 496} = \frac{1}{214768} < \frac{1}{10^4},$$

hence accurate to the four decimal places.

We know that r_n is always a better approximation to r than r_{n-2} is (recall $r_{n-2} < r_n < r$ if n is even and $r < r_n < r_{n-2}$ if n is odd). What about r_n vs r_{n-1}? We will answer the question by showing that, after the first step, the approximation always gets better as n increases.

When r is the value of the continued fraction, $[a; a_1, \dots]$, we let r_n denote the n-th convergent $[a; a_1, \dots, a_n]$; let $\rho_n \in \mathbb{R}$ be the output of the continued fraction algorithm after n-steps:

$$a_{n-1} = \lfloor \rho_{n-1} \rfloor \Rightarrow \rho_n = \frac{1}{\rho_{n-1} - a_{n-1}}$$

and

$$r = [a; a_1, \dots, a_{n-1}, \rho_n]$$

(see the proof of Theorem 33) and we may see ρ_n as

$$\rho_n = [a_n; a_{n+1}, \dots].$$

Proposition 33 shows that for the finite continued fraction $r_{n+1} \stackrel{\text{def}}{=} [a; a_1, \dots, a_{n+1}]$ equals $\frac{s_{n+1}}{t_{n+1}}$ which, by definition, is $\frac{a_{n+1}s_n + s_{n-1}}{a_{n+1}t_n + t_{n-1}}$. Something similar holds for an infinite continued fraction:

Lemma 40. Let r be an irrational number.

$$r = [a; a_1, \dots, a_n, \rho_{n+1}] = \frac{\rho_{n+1}s_n + s_{n-1}}{\rho_{n+1}t_n + t_{n-1}}.$$

Proof. The first equality is established in the proof of Theorem 37. We prove the second equality by induction on n.

By definition, $r = a + r - a = a + \frac{1}{\rho_1} = \frac{a\rho_1 + 1}{\rho_1}$.

Suppose that $[a; a_1, \dots, a_{n-1}, \rho_n] = \frac{\rho_n s_{n-1} + s_{n-2}}{\rho_n t_{n-1} + t_{n-2}}$ holds. Since $[a; a_1, \dots, a_n, \rho_{n+1}] = [a; a_1, \dots, a_{n-1}, a_n + \frac{1}{\rho_{n+1}}]$ and

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

□

It follows

$$r - \frac{s_n}{t_n} = [a; a_1, \dots, a_n, \rho_{n+1}] - \frac{s_n}{t_n} = \frac{\rho_{n+1}s_n + s_{n-1}}{\rho_{n+1}t_n + t_{n-1}} - \frac{s_n}{t_n} = \frac{t_n s_{n-1} - s_n t_{n-1}}{t_n(\rho_{n+1}t_n + t_{n-1})} = \frac{(-1)(-1)^{n-1}}{t_n(\rho_{n+1}t_n + t_{n-1})},$$

by Theorem 34 and therefore

$$|r - \frac{s_n}{t_n}| = \frac{|(-1)^n|}{|t_n(\rho_{n+1}t_n + t_{n-1})|} = \frac{1}{t_n(\rho_{n+1}t_n + t_{n-1})} < \frac{1}{t_n t_{n+1}}$$

as

$$\rho_{n+1}t_n + t_{n-1} > a_{n+1}t_n + t_{n-1} = t_{n+1}$$

(recall that $a_{n+1} = \lfloor \rho_{n+1} \rfloor$, hence $\rho_{n+1} > a_{n+1}$).

Proposition 41. For every $n \geq 2$, we have

- $|t_n r - s_n| < |t_{n-1} r - s_{n-1}|$,
- $|r - r_n| < |r - r_{n-1}|$.

Proof. Using the argument above, we have

$$|t_n r - s_n| = t_n |r - \frac{s_n}{t_n}| = \frac{1}{\rho_{n+1}t_n + t_{n-1}}$$

and

$$|t_{n-1}r - s_{n-1}| = t_{n-1}|r - \frac{s_{n-1}}{t_{n-1}}| = \frac{1}{\rho_n t_{n-1} + t_{n-2}}.$$

To prove the first assertion, it therefore suffices to establish

$$\rho_{n+1}t_n + t_{n-1} > \rho_n t_{n-1} + t_{n-2}.$$

By definition, $a_n = \lfloor \rho_n \rfloor$, thus $\rho_n < a_n + 1$. It follows that

$$\rho_n t_{n-1} + t_{n-2} < (a_n + 1)t_{n-1} + t_{n-2} = a_n t_{n-1} + t_{n-2} + t_{n-1} = t_n + t_{n-1} < \rho_{n+1}t_n + t_{n-1},$$

since $\rho_{n+1} > 1$ by definition– recall that $\rho_{n+1} = \dfrac{1}{\rho_n - a_n} = \dfrac{1}{\rho_n - \lfloor \rho_n \rfloor}$ where $0 \leq \rho_n - \lfloor \rho_n \rfloor < 1$.

To prove the second assertion, we firstly observe from the first assertion, combined with $t_n > t_{n-1} \geq 1$, that

$$\frac{|t_n r - s_n|}{t_n} < \frac{|t_{n-1}r - s_{n-1}|}{t_n} < \frac{|t_{n-1}r - s_{n-1}|}{t_{n-1}}.$$

The LHS is $|r - \dfrac{s_n}{t_n}| = |r - r_n|$, while the RHS is $|r - \dfrac{s_{n-1}}{t_{n-1}}| = |r - r_{n-1}|$. □

Definition. We say that a rational number $\dfrac{s}{t}$ is a good approximation to r if

$$|r - \frac{s}{t}| < |r - \frac{s'}{t'}|$$

for any rational number $\dfrac{s'}{t'}$ with $t' < t$.

In other words, there is no rational number closer to r than $\dfrac{s}{t}$ with smaller denominator; if $\dfrac{s'}{t'}$ has smaller denominator than $\dfrac{s}{t}$, then it has to be further from r than $\dfrac{s}{t}$ is to r.

Theorem 42. Let r be an irrational number, $[a; a_1, \dots]$ be the continued fraction for r, and $r_n = [a; a_1, \dots, a_n] = \dfrac{s_n}{t_n}$ be the n-th convergent. Let $\rho = \dfrac{s}{t}$ be a rational number in its lowest terms. If $t < t_n$ whenever $n > 1$, then

$$|r - \frac{s_n}{t_n}| < |r - \frac{s}{t}| = |r - \rho|$$

holds.

Remark. The theorem asserts that the convergents r_n, for $n \geq 2$, are good approximations to r.

We need a lemma.

Lemma 43. Let r be an irrational number, $[a; a_1, \dots]$ be its continued fraction, $r_n = [a; a_1, \dots, a_n] = \dfrac{s_n}{t_n}$ be the n-th convergent. If s and t are integers satisfying $\gcd(s, t) = 1$ and $t < t_n$, then

$$|tr - s| \geq |t_{n-1}r - s_{n-1}|$$

holds, with equality if and only if $\frac{s}{t} = \frac{s_{n-1}}{t_{n-1}}$.

Proof of the lemma. This is due to Lagrange. Consider the following system of equations

$$\begin{aligned} s_{n-1}x + s_n y &= s \\ t_{n-1}x + t_n y &= t. \end{aligned}$$

Since $s_{n-1}t_n - s_n t_{n-1} = (-1)^n$ (Theorem 34), we see that it has a unique integer solution

$$(x, y) = ((-1)^n(st_n - ts_n), (-1)^n(ts_{n-1} - st_{n-1})).$$

$\boxed{x \text{ is non-zero}}$ Suppose $x = 0$. This is equivalent to $st_n - ts_n = 0$, i.e. $\frac{s}{t} = \frac{s_n}{t_n}$. While $\frac{s_n}{t_n}$ is in its lowest terms (Theorem 34), the equality express the same rational number with two distinct denominators (recall $t < t_n$ by assumption). This is a contradiction.

$\boxed{y \text{ is non-zero or } (s,t) = (s_{n-1}, t_{n-1})}$ Similar to the argument above. If $y = 0$, then $ts_{n-1} - st_{n-1} = 0$, hence $\frac{s}{t} = \frac{s_{n-1}}{t_{n-1}}$; note that as '$t < t_{n-1}$' is NOT assumed, it is not possible to eliminate this case. To sum up, y is either non-zero, or zero in which case $\frac{s}{t} = \frac{s_{n-1}}{t_{n-1}}$ and therefore $s = s_{n-1}$ and $t = t_{n-1}$ (again, because '$t < t_{n-1}$' is NOT assumed, this is allowed!) and consequently

$$|tr - s| = |t_{n-1}r - s_{n-1}|.$$

Suppose that y is non-zero. Our goal then is to establish that

$$|tr - s| > |t_{n-1}r - s_{n-1}|.$$

$\boxed{x \text{ and } y \text{ have opposite signs}}$ Suppose that $y < 0$. If $\underline{n \text{ is odd}}$, then $y = -(ts_{n-1} - st_{n-1}) < 0$, i.e., $ts_{n-1} - st_{n-1} > 0$, i.e., $\frac{s}{t} < \frac{s_{n-1}}{t_{n-1}} = r_{n-1}$. On the other hand, Theorem 34 proves that $r_n - r_{n-1} > 0$ (since $n-1$ is even), hence $\frac{s}{t} < \frac{s_{n-1}}{t_{n-1}} < \frac{s_n}{t_n}$, i.e., $st_n - ts_n < 0$. As a result, $x = (-1)^n(st_n - ts_n)$ (since n is odd). Similarly, if $\underline{n \text{ is even}}$, $ts_{n-1} - st_{n-1} < 0$, i.e., $\frac{s}{t} > \frac{s_{n-1}}{t_{n-1}} = r_{n-1}$. As Theorem 34 proves that $r_n - r_{n-1} < 0$ (since $n-1$ is odd), it follows that $\frac{s}{t} > \frac{s_{n-1}}{t_{n-1}} > \frac{s_n}{t_n}$, i.e., $st_n - ts_n > 0$. As a result $x = (-1)^n(st_n - ts_n) > 0$ (since n is even).

One can similarly show that if $y > 0$, then $x < 0$ (whether n is odd or even).

$\boxed{t_{n-1}r - s_{n-1} \text{ and } t_n r - s_n \text{ have opposite signs}}$ Observe that r lies between $r_{n-1} = \frac{s_{n-1}}{t_{n-1}}$ and $r_n = \frac{s_n}{t_n}$, hence $t_{n-1}r - s_{n-1}$ and $t_n r - s_n$ have opposite signs. For example, if n is even, then we have

$$\frac{s_n}{t_n} = r_n < r < r_{n-1} = \frac{s_{n-1}}{t_{n-1}}.$$

The first inequality yields $t_n r - s_n > 0$, while the second yields $t_{n-1}r - s_{n-1} < 0$. If n is odd, then we have $r_{n-1} < r < r_n$ and this yields $t_n r - s_n < 0$ and $t_{n-1}r - s_{n-1} > 0$.

Combining the last two items, we may then deduce that $x(t_{n-1}r - s_{n-1})$ and $y(t_n r - s_n)$ have the same signs. It hence follows that

$$\begin{aligned} |tr - s| &= |(t_{n-1}x + t_n y)r - (s_{n-1}x + s_n y)| \\ &= |x(t_{n-1}r - s_{n-1}) + y(t_n r - s_n)| \\ &= |x(t_{n-1}r - s_{n-1})| + |y(t_n r - s_n)| \\ &= |x||t_{n-1}r - s_{n-1}| + |y||t_n r - s_n| \\ &> |t_{n-1}r - s_{n-1}|. \end{aligned}$$

The last (strict) inequality follows since $|x| \geq 1$ (this follows since x is a non-zero integer), $|t_{n-1}r - s_{n-1}| > 0$ and $|t_n r - s_n| > 0$. □

Proof of Theorem 42. Suppose $t < t_n$. Then it follows from Lemma 43, Proposition 41

$$|r - \frac{s}{t}| = \frac{1}{t}|rt - s| \geq \frac{1}{t}|rt_{n-1} - s_{n-1}| > \frac{1}{t}|rt_n - s_n| > \frac{1}{t_n}|rt_n - s_n| = |r - \frac{s_n}{t_n}|.$$

□

Theorem 44. Let r be an irrational number, and let $s, t \in \mathbb{Z}$ with $t > 0$ and $\gcd(s, t) = 1$. Suppose that $|r - \frac{s}{t}| < \frac{1}{2t^2}$. Then $\frac{s}{t}$ is a convergent to r.

Proof. Let $[a; a_1, \dots]$ be the continued fraction for r (Theorem 37), and define $r_n = \frac{s_n}{t_n}$ as before. Choose n such that

$$t_{n-1} \leq t < t_n.$$

Lemma 43 states $|tr - s| \geq |t_{n-1}r - s_{n-1}|$, and it therefore follows that

$$|t_{n-1}r - s_{n-1}| \leq |tr - s| = t|r - \frac{s}{t}| < \frac{t}{2t^2} = \frac{1}{2t}$$

implies

$$|r - \frac{s_{n-1}}{t_{n-1}}| \leq \frac{1}{2t_{n-1}t}.$$

On one hand,

$$|\frac{s}{t} - \frac{s_{n-1}}{t_{n-1}}| = |\frac{s}{t} - r + r - \frac{s_{n-1}}{t_{n-1}}| \leq |r - \frac{s}{t}| + |r - \frac{s_{n-1}}{t_{n-1}}| < \frac{1}{2t^2} + \frac{1}{2t_{n-1}t} \leq \frac{2}{2t_{n-1}t} = \frac{1}{t_{n-1}t}.$$

On the other hand,

$$|\frac{s}{t} - \frac{s_{n-1}}{t_{n-1}}| = \frac{st_{n-1} - ts_{n-1}}{tt_{n-1}}.$$

It therefore follows that the numerator $st_{n-1} - ts_{n-1}$ has to be zero, i.e. $\frac{s}{t} = \frac{s_{n-1}}{t_{n-1}}$. □

6.5 Periodic continue fraction II

A *quadratic irrational* r is a real number of the form $s + t\sqrt{d}$ where $s \in \mathbb{Q}, t \in \mathbb{Q} - \{0\}$ and $d > 1$ is a square-free integer.

Quadratic irrationals are precisely the roots of irreducible degree 2 polynomials with $\mathbb{Q}$-coefficients.

Theorem 45. If a real number has a periodic continued fraction, then it is a quadratic irrational.

Proof. We firstly show that a (real) number with a purely periodic continued fraction is a quadratic number. To this end, let

$$r = [\overline{a; a_1, \ldots, a_{l-1}}].$$

The cases when $l = 1$ and $l = 2$ are left as exercises. We henceforth assume $l \geq 3$. If r was rational, the continued fraction would have finite length; hence r is irrational. By assumption,

$$r = [a; a_1, \ldots, a_{l-1}, r]$$

and it follows from Lemma 40 that the RHS equals

$$\frac{rs_{l-1} + s_{l-2}}{rt_{l-1} + t_{l-2}},$$

where $\frac{s_n}{t_n}$ is the n-th convergent of $[\overline{a; a_1, \ldots, a_{l-1}}]$. We then see that

$$t_{l-1}r^2 + (t_{l-2} - s_{l-1})r - s_{l-2} = 0$$

and therefore that r is a quadratic irrational.

We now show that any number with periodic continued fraction is a quadratic irrational.

Let ρ be the value of $[a; a_1, \ldots, a_N, \overline{a_{N+1}, \ldots, a_{N+l}}]$ for fixed N and l, and let r be the value of $[\overline{a_{N+1}; a_{N+2}, \ldots, a_{N+l}}]$. By the first part, r is a quadratic irrational of the form, say, $s + t\sqrt{d}$ with $s \in \mathbb{Q}, t \in \mathbb{Q} - \{0\}$. We have

$$\rho = [a; a_1, \ldots, a_N, r] = \frac{rs_N + s_{N-1}}{rt_N + t_{N-1}}$$

by Lemma 40, where $\frac{s_n}{t_n}$ is the n-th convergent of ρ. Substituting $r = s + t\sqrt{d}$, we have

$$\rho = \frac{s_N s + s_{N-1} + s_N t\sqrt{d}}{t_N s + t_{N-1} + t_N t\sqrt{d}} = \cdots + \frac{s_{N-1}t(s_N - t_N)}{(t_N s + t_{N-1})^2 - (t_N t)^2 d}\sqrt{d} \in \mathbb{Q} + \mathbb{Q}\sqrt{d}$$

and ρ is quadratic irrational. □

The converse also holds, but it will not be discussed any further:

Theorem 46. A real number has periodic continued fraction if and only if it is quadratic irrational.

In fact, for a square-free positive integer d, $\sqrt{d}$ always has the continued fraction expression of the form

$$\sqrt{d} = [a; \overline{a_1, a_2, \ldots, a_2, a_1, 2a}].$$

7 Pell's equation

In this section, we will study Pell's equation:

$$x^2 - dy^2 = \pm 1.$$

We will see that continued fractions give us constructive methods of finding (integer) solutions to these equations. In doing so, we will describe the solutions in two different ways (in Part I and Part II).

7.1 Part I

Theorem 47 Suppose that $d \in \mathbb{N}$ is not a square and suppose that there is a pair of positive integers s and t satisfying $s^2 - dt^2 = \pm 1$. Then $\frac{s}{t}$ is a convergent to $\sqrt{d}$ (in the sense that it is of the form $r_n = \frac{s_n}{t_n}$ for some n).

Proof. If $s^2 - dt^2 = \pm 1$, then $(s + \sqrt{d}t)(s - \sqrt{d}t) = \pm 1$, hence

$$|\sqrt{d} - \frac{s}{t}| = \frac{1}{t(s + t\sqrt{d})}$$

holds. The denominator

$$t(s + t\sqrt{d}) = t^2(\frac{s}{t} + \sqrt{d}) = t^2(\sqrt{d \pm \frac{1}{t^2}} + \sqrt{d}) \geq t^2(\sqrt{d-1} + \sqrt{d}) > 2t^2$$

since $d \geq 2$ (as d is positive and not a square). It then follows from Theorem 44 that $\frac{s}{t}$ is a convergent to $\sqrt{d}$. □

Remark. The theorem asserts that positive integer solutions to Pell's equation $x^2 - dy^2 = \pm 1$ are necessarily convergents to the continued fraction of $\sqrt{d}$:

$$\left\{\text{The } \underline{\text{positive}} \text{ integer solutions to } x^2 - dy^2 = \pm 1\right\} \subset \left\{\text{the convergents } r_n = \frac{s_n}{t_n} \text{ to } \sqrt{d}\right\}$$

(recall that s_n and t_n are both positive if $n \geq 1$) But not all convergents are solutions to the Pell equation. Do we know which convergents?

Example. $x^2 - 2y^2 = \pm 1$.

The continued fraction of $\sqrt{2}$ is $[1; \overline{2}]$. The convergents are:

$$r_1 = \frac{3}{2}, r_2 = \frac{7}{5}, r_3 = \frac{17}{12}, r_4 = \frac{41}{29}, \ldots$$

These, as it turns out, define solutions to $x^2 - 2y^2 = \pm 1$:

$$3^2 - 2 \cdot 2^2 = 1, 7^2 - 2 \cdot 5^2 = -1, 17^2 - 2 \cdot 12^2 = 1, 41^2 - 2 \cdot 29^2 = -1, \ldots$$

It might be reasonable to expect that the convergent r_n, when n is even (resp. odd), define solutions to $x^2 - 2y^2 = -1$ (resp. $x^2 - 2y^2 = 1$).

Example. $x^2 - 3y^2 = \pm 1$.

The continued fraction of $\sqrt{3}$ is $[1; \overline{1, 2}]$. The convergents are:

$$r_1 = \frac{2}{1}, r_2 = \frac{5}{3}, r_3 = \frac{7}{4}, r_4 = \frac{19}{11}, \ldots$$

This time, not all of them are solutions to the Pell equation:

$$2^2 - 31^2 = 1, 5^2 - 3 \cdot 3^2 = -2, 7^2 - 3 \cdot 4^2 = 1, 19^2 - 3 \cdot 11^2 = -2, \ldots$$

Again, it might not be so far-fetched to conjecture that the r_n, where n is odd, define solutions to $x^2 - 3y^2 = 1$, while it is likely that $x^2 - 3y^2 = -1$ does not have any solutions. This can be checked by passing to $\mathbb{F}_3$. If there was a solution, say (s, t), then

$$s^2 \equiv -1 \equiv 2$$

mod 3, however the Legendre symbol $\left(\frac{2}{3}\right) = -1$, a contradiction!

The following theorem singles out exactly which convergents to $\sqrt{d}$ indeed define positive integer solutions to $x^2 - dy^2 = \pm 1$:

Theorem 48 Suppose that $d \in \mathbb{N}$ is not a square. Suppose that $\sqrt{d} = [a; \overline{a_1, \ldots, a_l}]$. Let $\frac{s_n}{t_n}$ be the n-th convergent of the continued fraction of $\sqrt{d}$. Then

$$s_n^2 - dt_n^2 = \pm 1$$

if and only if $n = Nl - 1$ for some $N = 1, 2, 3, \ldots$.

Moreover,

$$s_{Nl-1}^2 - dt_{Nl-1}^2 = (-1)^{Nl}.$$

Remark. As advertised, Theorem 48 proves that, if $\sqrt{d} = [a; \overline{a_1, \ldots, a_l}]$,

$$\left\{\text{The } \underline{\text{positive}} \text{ integer solutions to } x^2 - dy^2 = \pm 1\right\} = \left\{\text{the 'convergents' } (s_{Nl-1}, t_{Nl-1}),\ N = 1, 2, \ldots, \text{ to } \sqrt{d}\right\}$$

Proof. NON-EXAMINABLE. □

Example. $x^2 - 2y^2 = \pm 1$. In this case, $l = 1$ and every $r_n = \frac{s_n}{t_n}$ is a solution for $n = 0, 1, 2, \ldots$; and $s_n^2 - 2t_n^2 = (-1)^{n+1}$.

Example. $x^2 - 3y^2 = \pm 1$. In this case, $l = 2$ and every r_n, when $n = 2N - 1$, i.e. when n is odd, defines a solution to $x^2 - 3y^2 = \pm 1$. In fact $s_{2N-1}^2 - 3t_{2N-1}^2 = (-1)^{2N} = 1$ and $x^2 - 3y^2 = -1$ does not have any solutions (as expected).

As we have seen in the second example, when l is even, $(-1)^{Nl} = 1$, and the following follows immediately from the theorem:

Corollary 49 Suppose that $d \in \mathbb{N}$ is not a square. Suppose that $\sqrt{d} = [a; \overline{a_1, \ldots, a_l}]$. If l is even, the equation

$$x^2 - dy^2 = -1$$

has no solutions.

7.2 Part II

Definition. We define a partial order on the set of solutions to equation $x^2 - dy^2 = \pm 1$: if (s,t) and (s',t') are two distinct solutions, we then define

$$(s,t) < (s',t')$$

if $x + y\sqrt{d} < s' + t'\sqrt{d}$ in $\mathbb{R}$ (SL says this is equivalent to $s < s'$ and $t < t'$). The fundamental solution is the minimum positive solution in this sense.

By Theorem 48, we know that the fundamental solution to $x^2 - dy^2 = \pm 1$ is $(x,y) = (s_{l-1}, t_{l-1})$ where $\frac{s_{l-1}}{t_{l-1}}$ is the $(l-1)$-st convergent.

We will see that the fundamental solution generates all positive integer solutions (x,y).

Example. $x^2 - 2y^2 = \pm 1$. As we saw already, $(s_0, t_0) = (1,1), (s_2, t_2) = (7,5), \ldots$ define solutions to

$$x^2 - 2y^2 = -1,$$

while $(s_1, t_1) = (3,2), (s_3, t_3) = (17,12), \ldots$ define solutions to

$$x^2 - 2y^2 = 1.$$

An eagle-eyed reader might notice a pattern– if (v_n, w_n) is the n-th (positive integer) solution, then $(v_{n+1}, w_{n+1}) = (v_n + 2w_n, v_n + w_n)$ is the $(n+1)$-st, and

$$v_{n+1} + w_{n+1}\sqrt{2} = (v_n + w_n\sqrt{2})(1 + \sqrt{2})$$

in other words,

$$(v_n + w_n\sqrt{2}) = (v_1 + w_1\sqrt{2})^n = (1 + \sqrt{2})^n.$$

This is not a coincidence, as we shall see shortly.

Example. $x^2 - 3y^2 = \pm 1$. The first few solutions to $x^2 - 3y^2 = 1$ are $(v_1, w_1) = (s_1, t_1) = (2,1), (v_2, w_2) = (s_3, t_3) = (7,4), (v_3, w_3) = (s_5, t_5) = (26,15), \ldots$ and solutions necessarily satisfy

$$v_n + w_n\sqrt{3} = (v_1 + w_1\sqrt{3})^n = (2 + \sqrt{3})^n.$$

Lemma 50 Let $(s,t) = (v_1, w_1)$ be the fundamental solution to the Pell equation

$$x^2 - dy^2 = \pm 1.$$

For $n = 1, 2, \ldots$, define $(v_n, w_n) \in \mathbb{N} \times \mathbb{N}$ by the equation

$$v_n + w_n\sqrt{d} = (s + t\sqrt{d})^n.$$

Then

$$v_n = \frac{1}{2}\left((s + t\sqrt{d})^n + (s - t\sqrt{d})^n\right)$$

and

$$w_n = \frac{1}{2\sqrt{d}}\left((s + t\sqrt{d})^n - (s - t\sqrt{d})^n\right).$$

Remark. Note that (v_n, w_n) is different from (s_n, t_n) that defines the n-th convergent r_n any longer.

Proof. Induction on n. □

Theorem 51 Let $(s, t) = (v_1, w_1)$ be the fundamental solution to the equation

$$x^2 - dy^2 = \pm 1$$

and let

$$\epsilon = s^2 - dt^2 \in \{\pm 1\}.$$

As before, for $n = 1, 2, \ldots$, define $(v_n, w_n) \in \mathbb{N} \times \mathbb{N}$ by the equation

$$v_n + w_n\sqrt{d} = (s + t\sqrt{d})^n.$$

Then

$$v_n^2 - dw_n^2 = \epsilon^n.$$

Remark. Theorem 51 proves that

$$\left\{\text{The } \underline{\text{positive}} \text{ integer solutions to } x^2 - dy^2 = \pm 1\right\} \supset \{(v_n, w_n)\}.$$

Proof of Theorem 51. Applying Lemma 50, we obtain

$$v_n - w_n\sqrt{d} = \frac{1}{2}\left((s + t\sqrt{d})^n + (s - t\sqrt{d})^n\right) - \frac{\sqrt{d}}{2\sqrt{d}}\left((s + t\sqrt{d})^n - (s - t\sqrt{d})^n\right) = (s - t\sqrt{d})^n.$$

Hence

$$v_n^2 - dw_n^2 = (v_n - w_n\sqrt{d})(v_n + w_n\sqrt{d}) = (s - t\sqrt{d})^n(s + t\sqrt{d})^n = (s^2 - dt^2)^n.$$

□

The following theorem establishes the 'converse of Theorem 51', i.e. $\{(v_n, w_n)\}$ sees all the positive integer solutions to $x^2 - dy^2 = \pm 1$, i.e.

$$\left\{\text{The } \underline{\text{positive}} \text{ integer solutions to } x^2 - dy^2 = \pm 1\right\} = \{(v_n, w_n)\}.$$

Theorem 52 Suppose that $d \in \mathbb{N}$ is not a square. Suppose that (v, w) is a solution to the Pell equation

$$x^2 - dy^2 = \pm 1.$$

Then there exists $n \geq 1$ such that $(v, w) = (v_n, w_n)$ as above.

Proof. Let (s, t) be the fundamental solution to $x^2 - dy^2 = \pm 1$. Suppose there exists a pair (v, w) of integers such that

- $v \geq 1$ and $w \geq 1$,
- $v^2 - dw^2 = \pm 1$,
- (v, w) is not (v_n, w_n) for any $n \geq 1$, where (v_n, w_n) is a pair of integers satisfying

$$v_n + w_n\sqrt{d} = \left(s + t\sqrt{d}\right)^n$$

and

$$v_n^2 - dw_n^2 = \pm 1$$

The assertion follows if this set of assumption leads to a contradiction.

There exists a unique N such that

$$(s + t\sqrt{d})^N < v + w\sqrt{d} < (s + t\sqrt{d})^{N+1},$$

because the interval $(s + t\sqrt{d})^{N+1} - (s + t\sqrt{d})^N = (s + t\sqrt{d})^N(s + t\sqrt{d} - 1)$ gets bigger as N increases [the point is that $v + w\sqrt{d}$ is bounded strictly by powers of $s + t\sqrt{d}$; this only occurs if (v, w) is not (v_n, w_n)!].

Multiplying all by $\dfrac{1}{(s + t\sqrt{d})^N}$, we have

$$1 < V + W\sqrt{d} < s + t\sqrt{d}$$

where

$$V + W\sqrt{d} = \frac{1}{(s + t\sqrt{d})^N}(v + w\sqrt{d}) = \frac{1}{v_N + w_N\sqrt{d}}(v + w\sqrt{d}) = \frac{v_N - w_N\sqrt{d}}{v_N^2 - dw_N^2}(v + w\sqrt{d}).$$

It is straightforward to check that

$$V - W\sqrt{d} = \frac{v_N + w_N\sqrt{d}}{v_N^2 - dw_N^2}(v - w\sqrt{d}).$$

We check
$\boxed{V^2 - dW^2 = \pm 1, \text{ i.e. } (V, W) \text{ is a solution for } x^2 - dy^2 = \pm 1}$. Simply substituting above,

$$V^2 - dW^2 = (V + W\sqrt{d})(V - W\sqrt{d}) = \frac{1}{(v_N^2 - dw_N^2)^2}(v^2 - dw^2)(v_N^2 - dw_N^2) = \frac{v^2 - dw^2}{v_N^2 - dw_N^2} = \pm 1$$

by Theorem 51.

$\boxed{-1 < V - W\sqrt{d} < 1}$ Since $|V + W\sqrt{d}||V - W\sqrt{d}| = |\pm 1| = 1$ and $V + W\sqrt{d} > 1$, it follows that $|V - W\sqrt{d}| < 1$.

Consequently,

$$2V = (V + W\sqrt{d}) + (V - W\sqrt{d}) > 1 - 1 = 0,$$

hence $V > 0$ and

$$2W\sqrt{d} = (V + W\sqrt{d}) - (V - W\sqrt{d}) > 1 - 1 = 0,$$

hence $W > 0$. This contradicts the minimality of the fundamental solution $s + t\sqrt{d}$. □

To sum up, we prove that, when$\sqrt{d} = [a, \overline{a_1, \ldots, a_l}]$,

$$\left\{\text{The } \underline{\text{positive}} \text{ integer solutions to } x^2 - dy^2 = \pm 1\right\} \quad = \quad \begin{cases} \{(s_{nl-1}, w_{nl-1})\} \text{ (Part I)} \\ \{(v_n, w_n)\} \text{ (Part II)} \end{cases}$$

In particular, we see

- The fundamental solution (s, t) is (v_1, w_1);
- $(v_n, w_n) = (s_{nl-1}, t_{nl-1})$ [note that we see this equality rather indirectly, without comparing (v_n, w_n) and (s_{nl-1}, t_{nl-1}) directly] and $v_n^2 - dw_n^2 = (-1)^{nl}$;

and as a result

$$\epsilon = s^2 - dt^2 = v_1^2 - dw_1^2 = (-1)^l.$$

If l is *odd*, then

- (v_n, w_n), for n even, are solutions to the Pell equation

$$x^2 - dy^2 = +1$$

- (v_n, w_n), for n odd, are solutions to the Pell equation

$$x^2 - dy^2 = -1.$$

Example. $x^2 - 3y^2 = 1$. The fundamental solution is $(s, t) = (2, 1)$. Hence

$$\begin{array}{ccccc} v_2 + w_2\sqrt{3} & = & (2 + \sqrt{3})^2 & = & 7 + 4\sqrt{3}, \\ v_3 + w_3\sqrt{3} & = & (2 + \sqrt{3})^3 & = & 26 + 15\sqrt{3}, \\ v_4 + w_4\sqrt{3} & = & (2 + \sqrt{3})^2 & = & 97 + 56\sqrt{3}, \\ \vdots & & \vdots & & \vdots \end{array}$$

Example. The continued fraction of $\sqrt{13}$ is $[3; \overline{1, 1, 1, 1, 6}]$ with $l = 5$. Hence the positive integer solutions to $x^2 - 13y^2 = \pm 1$ are the convergent (s_{5N-1}, t_{5N-1}) where $r_{5N-1} = \dfrac{s_{5N-1}}{t_{5N-1}}$ is

the $5N-1$-st convergent. The 'smallest' solution is $(s_4,t_4)=(18,5)$ and $18^2-13\cdot 5^2=-1$.

Example. $x^2-41y^2=-1$. The fundamental solution to $x^2-41y^2=-1$ is $(s_1,t_1)=(32,5)$ and the fundamental solution to $x^2-41y^2=1$ is (s_2,t_2) is computed by

$$v_2+w_2\sqrt{41}=(32+5\sqrt{41})^2=2049+320\sqrt{41}.$$

Example. The continued fraction of $\sqrt{61}$ is $[7;\overline{1,4,3,1,2,2,1,3,4,1,14}]$ with period $l=11$. It follows from Theorem 48 that the solutions to $x^2-\sqrt{61}y^2=\pm 1$ are

$$(s_{10},t_{10})=(29718,3805),(s_{21},t_{21}),(s_{32},t_{32}),\dots$$

(satisfying $s_{11N-1}^2-dt_{11N-1}^2=(-1)^{11N}$), and the fundamental solution to $x^2-dy^2=-1$ is

$$(s_{10},t_{10})=(29718,3805)$$

while the fundamental solution to $x^2-dy^2=1$ is

$$(s_{21},t_{21})=(1766319049,226153980).$$

Theorem 51 ascertains that $(29718+3805\sqrt{61})^2=1766319049+226153980\sqrt{61}$.

7.3 Appendix: A proof of Theorem 48 (NON-EXAMINABLE)

Let d be a square-free integer. Let $\{a_n\}$ (resp. $\{\rho_n\}$) be positive integers (resp. real numbers) appearing in the continued fraction algorithm for $\sqrt{d}$, i.e.,

$$\sqrt{d}=[a;a_1,\dots,a_{n-1},\rho_n]$$

(by definition, $a_{n-1}=\lfloor\rho_{n-1}\rfloor$ and $\rho_n=\dfrac{1}{\rho_{n-1}-a_{n-1}}$, hence $\rho_{n-1}=a_{n-1}+\dfrac{1}{\rho_n}$; furthermore, since $\sqrt{d}$ is irrational, ρ_n is non-zero for any n).

Definition. Define integers R_n and S_n inductively as follows:

- $R_0=1$ and $S_0=0$,
- $S_{n+1}=a_{n-1}R_n-S_n$,
- $R_{n+1}=\dfrac{d-S_{n+1}^2}{R_n}$.

The following is the key lemma:

Lemma We have

- R_n and S_n are both integers,
- R_n divides $d-S_n^2$,
- $\rho_{n-1}=\dfrac{S_n+\sqrt{d}}{R_n}$

Proof of the lemma. We prove this by induction. The case when $n = 0$ holds trivially. Suppose that the assertion holds for n.

$\boxed{S_{n+1} \in \mathbb{Z}}$ By definition, $S_{n+1} = a_{n-1}R_n - S_n$ where a_{n-1} is an integer by definition and R_n and S_n are integers by the inductive hypothesis.

$\boxed{R_{n+1} \in \mathbb{Z}}$ By definition,

$$R_{n+1} = \frac{d - S_{n+1}^2}{R_n} = \frac{d - (a_{n-1}R_n - S_n)^2}{R_n} = \frac{d - S_n^2}{R_n} + 2a_{n-1}S_n - a_{n-1}R_n.$$

The assertion follows since $\dfrac{d - S_n^2}{R_n}$ is an integer by the inductive hypothesis.

$\boxed{R_{n+1} \text{ divides } d - S_{n+1}^2}$ This follows immediately from $R_{n+1}R_n = d - S_{n+1}^2$.

$\boxed{\rho_n = \dfrac{S_{n+1} + \sqrt{d}}{R_{n+1}}}$ Since $\rho_{n-1} = a_{n-1} + \dfrac{1}{\rho_n}$, it follows from the inductive hypothesis that

$$a_{n-1} + \frac{1}{\rho_n} = \frac{S_n + \sqrt{d}}{R_n}.$$

It follows that

$$\rho_n = \frac{R_n}{S_n + \sqrt{d} - a_{n-1}R_n} = \frac{R_n}{\sqrt{d} - S_{n+1}} = \frac{R_n(\sqrt{d} + S_{n+1})}{d - S_{n+1}^2} = \frac{R_n(\sqrt{d} + S_{n+1})}{R_nR_{n+1}} = \frac{\sqrt{d} + S_{n+1}}{R_{n+1}},$$

as desired. □

Proposition A-1 Let d be a square-free integer. Let $\dfrac{s_n}{t_n}$ be the n-th convergent to $\sqrt{d}$. We then have

$$s_n^2 - dt_n^2 = (-1)^{n+1}R_{n+1}$$

and $R_{n+1} > 0$.

Proof. Since $\sqrt{d} = [a; a_1, \ldots, a_{n-1}, \rho_n] = \dfrac{\rho_n s_n + s_{n-1}}{\rho_n t_n + t_{n-1}}$,

$$\sqrt{d}\,(\rho_n t_n + t_{n-1}) = \rho_n s_n + s_{n-1}.$$

Substituting $\rho_n = \dfrac{S_{n+1} + \sqrt{d}}{R_{n+1}}$ from the lemma, we have

$$\sqrt{d}\left(\frac{S_{n+1} + \sqrt{d}}{R_{n+1}}t_n + t_{n-1}\right) = \frac{S_{n+1} + \sqrt{d}}{R_{n+1}}s_n + s_{n-1}.$$

Multiplying R_{n+1} and rearranging, we have

$$\sqrt{d}\,(S_{n+1}t_n + R_{n+1}t_{n-1} - s_n) = S_{n+1}s_n + R_{n+1}s_{n-1} - dt_n.$$

Since $\sqrt{d}$ is irrational, it follows that

$$S_{n+1}t_n + R_{n+1}t_{n-1} - s_n = 0 \Leftrightarrow s_n = S_{n+1}t_n + R_{n+1}t_{n-1},$$

and

$$S_{n+1}s_n + R_{n+1}s_{n-1} - dt_n = 0 \Leftrightarrow dt_n = S_{n+1}s_n + R_{n+1}s_{n-1}.$$

It therefore follows that

$$s_n^2 - dt_n^2 = s_n(S_{n+1}t_n + R_{n+1}t_{n-1}) - t_n(S_{n+1}s_n + R_{n+1}s_{n-1}) = R_{n+1}(s_nt_{n-1} - t_ns_{n-1}) = R_{n+1}(-1)^{n-1}$$

by Theorem 34. □

Proposition A-2 Suppose that $\sqrt{d} = [a; \overline{a_1, \ldots, a_l}]$ for some $l \geq 1$. Let $\{R_n\}$ be as above. Then $R_n = 1$ if and only if l divides n.

Corollary. Theorem 48 follows.

Proof of Corollary (Theorem 48).

$$s_n^2 - dt_n^2 = (-1)^n \Leftrightarrow R_{n+1} = 1 \Leftrightarrow l|(n+1) \Leftrightarrow n = l-1, 2l-1, \ldots.$$

□

Proof of Proposition A-2. Suppose that l divides n. We show that for any multiple kl of l, $R_{kl} = 1$. Firstly, since $\sqrt{d} = [a; \rho_1]$ by definition, we have $\rho_1 = [\overline{a_1; a_2, \ldots, a_l}]$. Similarly, since $\sqrt{d} = [a; a_1, \ldots, a_l, \rho_{l+1}]$, we also have $\rho_{l+1} = [\overline{a_1; a_2, \ldots, a_l}]$; indeed, for any integer $k \geq 1$, we have

$$\rho_{kl+1} = [\overline{a_1; a_2, \ldots, a_l}] = \rho_1$$

(easy to check by induction). By the lemma above, it therefore follows that

$$\frac{S_{kl+1} + \sqrt{d}}{R_{kl+1}} = \frac{S_1 + \sqrt{d}}{R_1},$$

and, as a result, that $R_{kl+1} = R_1$ and $S_{kl+1} = S_1$. By definition,

$$R_1 = d - S_1^2 = d - S_{kl+1}^2 = R_{kl}R_{kl+1} = R_{kl}R_1.$$

Since $R_1 > 0$, we have $R_{kl} = 1$ as desired.

Conversely, suppose that $R_n = 1$. It follows from the lemma that

$$a_n + \frac{1}{\rho_n} = \rho_{n-1} = \frac{S_n + \sqrt{d}}{R_n} = S_n + \sqrt{d} = S_n + a + \frac{1}{\rho_1}.$$

While a_n (on the leftmost), S_n and a (on the rightmost) are all integers, both $\frac{1}{\rho_n}$ and $\frac{1}{\rho_1}$ are fractions < 1 and therefore the equality $\rho_n = \rho_1$ needs to hold. This implies that l divides n (as l is the period length). □

8 Sums of squares

8.1 $x^2 + y^2 = p$

Let p be a prime. The basic question we want to understand in this section is whether

$$x^2 + y^2 = p$$

has a solution in $(x, y) \in \mathbb{N} \times \mathbb{N}$. For example,

$$\begin{aligned} 2^2 + 1^2 &= 5 \\ 3^2 + 2^2 &= 13 \\ 4^2 + 1^2 &= 17 \\ 5^2 + 2^2 &= 29. \end{aligned}$$

On the other hand,

$$x^2 + y^2 = 7$$

is not soluble in $\mathbb{Z} \times \mathbb{Z}$; for if it were, there would be $(m, n) \in \mathbb{Z} \times \mathbb{Z}$ such that $m^2 + n^2 = 7$, but the table

$z \pmod 4$	$z^2 \pmod 4$
0	0
1	1
2	0
3	1

shows $m^2 + n^2$ would never be $7 \equiv 3 \bmod 4$.

In fact,

Proposition 53. Let p be a prime congruent to 3 mod 4. Then

$$x^2 + y^2 = p$$

has no solutions in $(x, y) \in \mathbb{Z} \times \mathbb{Z}$.

Proof. The table above shows that, for any pair of integers r, s, the sum $r^2 + s^2$ is congruent to $0, 1$, or $2 \bmod 4$. If (r, s) are a solution for $x^2 + y^2 = p$, then p would be congruent to $0, 1$ or 2, but this contradicts the assumption that p is congruent to 3. □

On the other hand, the following theorem gives us a good handle on primes representable as sums of squares:

Theorem 54. If $\left(\dfrac{-1}{p}\right) = 1$, then the equation

$$x^2 + y^2 = p$$

is soluble.

To prove the theorem, we firstly prove

Lemma 55. Let $r \in \mathbb{R}$ and $N \in \mathbb{N}$. Then there exists $s/t \in \mathbb{Q}$ with $\gcd(s,t) = 1$ and $1 \leq t \leq N$ such that

$$\left| r - \frac{s}{t} \right| \leq \frac{1}{t(N+1)}.$$

Remark. Indeed, we show that we may take $\frac{s}{t}$ to be a convergent $\frac{s_n}{t_n}$ to r for some n.

Proof. Let $r = [a; a_1, \dots]$ be the continued fraction of r and $r_n = \frac{s_n}{t_n}$ be the n-th convergent (recall that t_n is an increasing sequence). It follows from Theorem 34 that

$$|r - r_n| = |r - \frac{s_n}{t_n}| \leq |r_{n+1} - r_n| \leq \frac{1}{t_n t_{n+1}}.$$

There are two cases to proceed:

Suppose that $t_n \leq N$ for every n. In this case, the increasing sequence t_n is bounded from above, i.e., t_n stabilises for sufficiently large n, i.e., the continued sequence is indeed finite and r is a rational number $r = [a; a_1, \dots, a_n]$ for some n. Letting $\frac{s}{t} = \frac{s_n}{t_n}$, we have

$$|r - \frac{s}{t}| = 0 \leq \frac{1}{t(N+1)}.$$

Suppose that there exists n such that

$$t_n \leq N < t_{n+1}$$

(whether r is a rational or not) Since N and t_{n+1} are both integers, it follows that $N + 1 \leq t_{n+1}$. Letting $\frac{s}{t} = \frac{s_n}{t_n}$,

$$|r - \frac{s}{t}| = |r - \frac{s_n}{t_n}| \leq \frac{1}{t_n t_{n+1}} \leq \frac{1}{t(N+1)}.$$

□

Proof of Theorem 54. Since -1 is a quadratic residue mod p, there exists an integer z such that $z^2 \equiv -1 \bmod p$. Applying the lemma with $r = \frac{z}{p}$ and $N = \lfloor \sqrt{p} \rfloor$, we find $\frac{s}{t} \in \mathbb{Q}$ such that $1 \leq t \leq \lfloor \sqrt{p} \rfloor < \sqrt{p}$ and

$$|r - \frac{s}{t}| = |\frac{z}{p} - \frac{s}{t}| \leq \frac{1}{t(N+1)} < \frac{1}{t\sqrt{p}}$$

since $N = \lfloor \sqrt{p} \rfloor < \sqrt{p} < N + 1$.

Let $u = ps - zt$. Then, since $t > 0$ and $p > 0$,

$$|u| = tp \left| \frac{z}{p} - \frac{s}{t} \right| < \frac{tp}{t\sqrt{p}} = \sqrt{p}.$$

$\boxed{(u,t) \text{ is a solution we are looking for}}$ It follows from the last inequality that

$$0 < u^2 + t^2 < p + p = 2p$$

as we know $t \leq N < \sqrt{p}$. On the other hand,

$$u^2 + t^2 = (ps - zt)^2 + t^2 \equiv z^2t^2 + t^2 \equiv (z^2 + 1)t^2$$

mod p. However, by assumption, $z^2 \equiv -1 \bmod p$, hence $u^2 + t^2 \equiv 0 \bmod p$. The only possibility for the 'real' value of $u^2 + t^2$ therefore is p, i.e., $u^2 + t^2 = p$. □

Corollary 56. Let p be a prime congruent to 1 mod 4. Then

$$x^2 + y^2 = p$$

has an integer solution in x and y.

Proof. It follows from the theorem that if $\left(\frac{-1}{p}\right) = 1$, then $x^2 + y^2 = p$ is soluble. By assumption, p is odd and Rule 2 in Theorem 25 asserts that $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \bmod 4$. □

8.2 Hermite's algorithm

The proof of Theorem 54 can be made into an algorithm for finding x, y such that $x^2 + y^2 = p$.

Step 1: find z such that $z^2 \equiv -1 \bmod p$.

Step 2 (inductive): compute the n-th $r_n = \frac{s_n}{t_n}$ and the $(n+1)$-th convergents $r_{n+1} = \frac{s_{n+1}}{t_{n+1}}$ of the continued fraction $\frac{z}{p}$. If

$$t_n < \sqrt{p} < t_{n+1},$$

then the algorithm stops and $(x, y) = (t_n, ps_n - zt_n)$ defines a solution for $x^2 + y^2 = p$. If this does not hold,

Step 3: see if $t_{n+1} < \sqrt{p} < t_{n+2}$ works. If it does, the algorithm stops. If not...

Example. Let $p = 13$.

Step 1: we simply spot that $5^2 \equiv -1 \bmod 13$.

Step 2: To find the convergents of $\frac{5}{13}$, we see

$$
\begin{array}{rcl}
a = \lfloor \frac{5}{13} \rfloor = 0 & \Rightarrow & r_1 = \frac{1}{\frac{5}{13} - 0} = \frac{13}{5} \\
& \swarrow & \\
a_1 = \lfloor \frac{13}{5} \rfloor = 2 & \Rightarrow & r_2 = \frac{1}{\frac{13}{5} - 2} = \frac{5}{3} \\
& \swarrow & \\
a_2 = \lfloor \frac{5}{3} \rfloor = 1 & \Rightarrow & r_3 = \frac{1}{\frac{5}{3} - 1} = \frac{3}{2} \\
& \swarrow & \\
a_3 = \lfloor \frac{3}{2} \rfloor = 1 & \Rightarrow & r_4 = \frac{1}{\frac{3}{2} - 1} = 2 \in \mathbb{N} \\
& \swarrow & \\
a_4 = \lfloor r_4 \rfloor = r_4 & &
\end{array}
$$

Hence the convergents are

$$\frac{s_0}{t_0} = 0, \frac{s_1}{t_1} = [0;2] = \frac{1}{2}, \frac{s_2}{t_2} = [0;2,1] = \frac{1}{3}, \frac{s_3}{t_3} = [0;2,1,1] = \frac{2}{5}.$$

The algorithm stops at $n = 2$ as

$$t_2 = 3 < \sqrt{13} < t_3 = 5$$

and $(x,y) = (3, 13 \cdot 1 - 5 \cdot 3) = (3,-2)$, or $(3,2)$ defines a solution. Indeed, $3^2 + 2^2 = 13$.

Example. $p = 2017$.

Step 1: $229^2 \equiv -1 \bmod 2017$.

To do this, we make appeal to Proposition 29. Since $\left(\frac{5}{2017}\right) = -1$ (trial and error), it follows from Proposition 29 that $5^{\frac{2017-1}{4}} = 5^{504}$ defines a solution to $x^2 \equiv -1 \bmod 2017$. Simplify $5^{504} \bmod 2017$, we get $229 \bmod 2017$.

Step 2: to find the convergents of $\frac{229}{2017}$, we see that

$$
\begin{array}{rcl}
a = \lfloor \frac{229}{2017} \rfloor = 0 & \Rightarrow & r_1 = \frac{1}{\frac{229}{2017} - 0} = \frac{2017}{229} \\
& \swarrow & \\
a_1 = \lfloor \frac{2017}{229} \rfloor = 8 & \Rightarrow & r_2 = \frac{1}{\frac{2017}{229} - 8} = \frac{229}{185} \\
& \swarrow & \\
a_2 = \lfloor \frac{229}{185} \rfloor = 1 & \Rightarrow & r_3 = \frac{1}{\frac{229}{185} - 1} = \frac{185}{44} \\
& \swarrow & \\
a_3 = \lfloor \frac{185}{44} \rfloor = 4 & \Rightarrow & r_4 = \frac{1}{\frac{185}{44} - 4} = \frac{44}{9} \\
& \swarrow & \\
a_4 = \lfloor \frac{44}{9} \rfloor = 4 & \Rightarrow & r_5 = \frac{1}{\frac{44}{9} - 4} = \frac{9}{8} \\
& \swarrow & \\
& \vdots &
\end{array}
$$

Hence the convergents are

$$\frac{s_0}{t_0} = 0, \frac{s_1}{t_1} = \frac{1}{8}, \frac{s_2}{t_2} = \frac{1}{9}, \frac{s_3}{t_3} = \frac{5}{44}, \frac{s_4}{t_4} = \frac{21}{185}, \dots$$

The algorithm stops at $n = 3$ as

$$t_3 = 44 < \sqrt{2017} < t_4 = 185.$$

and $(x, y) = (44, 2017 \cdot 5 - 229 \cdot 44) = (44, 9)$ is a solution. Indeed, $9^2 + 44^2 = 2017$.

8.3 More sums of squares

Let $n \in \mathbb{N}$. We can write it as $n = a^2 b$ where $a, b \in \mathbb{N}$ and b is *square free*, in the sense that if p is a prime that divides b, then p^2 does not divide b). More precisely, it follows from the Fundamental Theorem of Algebra that we may write n as

$$n = \prod_p p^{r_p} = \prod_{r_p = 2s_p} p^{2s_p} \prod_{r_p = 2s_p+1} p^{2s_p+1} = \left(\prod_p p^{s_p} \right)^2 \prod_{r_p = 2s_p+1} p$$

and we may take $a = \prod_p p^{s_p}$ (where p ranges over all prime factors of n) and $b = \prod_{r_p = 2s_p+1} p$ (where p ranges over the prime factors of n for which r_p is odd).

For example,

$$1440 = 2^5 3^2 5 = 12^2 \cdot 10.$$

We shall refer to b as the square-free part of n.

Theorem 57 (Fermat & Euler) A positive integer n is the sum of <u>two</u> squares (of integers) if and only if the square-free part ρ of n has no prime factors congruent to 3 mod 4.

Proof. Suppose firstly that ρ has no prime factors congruent to 3 mod 4.

If $\rho = \pm 1$, then n is a square and it is evidently a sum of squares (since 0^2 is a sqaure). We may then suppose that $\rho > 1$.Since the product of sums of squares is, again, a sum of squares [if $\alpha = r^2 + s^2$ and $\beta = t^2 + u^2$, then $\alpha\beta = (r^2 + s^2)(t^2 + u^2) = (rt - su)^2 + (ru + st)^2$], it suffices to establish that any prime factor p of ρ is a sum of squares. In theory, there are 4 cases mod 4 to deal with:

$\boxed{p \equiv 0}$ This can occur since p is a prime.

$\boxed{p \equiv 1}$ This follows from Corollary 56.

$\boxed{p \equiv 2}$ The only possibility for p (prime *and* congruent to 2 mod 4) is 2. Clearly 2 is a sum of squares $2 = 1^2 + 1^2$!

$\boxed{p \equiv 3}$ This is excluded by assumption.

Conversely, suppose that $n = r^2 + s^2$ for some integers r, s. It suffices to prove that if p divides the square-free part of n, then p is not congruent to 3 mod 4. It is equivalent to establishing that if p is a prime factor of n and is congruent to 3 mod 4, then p is a factor of the 'square-part' of n, i.e. , if p is a prime factor of n, is congruent to 3 mod 4 and p^N is the maximal p-power divisor of n, then N is even. We shall prove the latter by induction on 1.

If $n = 1$, then the assertion holds (as $n = 1$ has no non-trivial divisors).

Suppose that the assertion holds for a positive integer $< n$. Suppose that $p \equiv 3$ mod 4 and $p|n$. The goal is to show that p divides n an even number of times.

$\boxed{p|r \text{ and } p|s}$ Suppose WLOG that p does not divide r. Therefore there exists t such that $rt \equiv 1$ mod p. On the other hand, since $p|n$, it follows from $n = r^2 + s^2$ that

$$r^2 + s^2 \equiv 0$$

mod p. Multiplying the congruence by t^2, we obtain

$$0 \equiv t^2(r^2 + s^2) = (rt)^2 + (st)^2 = 1 + (st)^2,$$

i.e. $(st)^2 \equiv -1$ mod p. In other words, $\left(\frac{-1}{p}\right) = 1$ but this contradicts Rule 2 in Theorem 25 (this is where we use $p \equiv 3$ mod 4).

Let $r = pu$ and $s = pv$. Substitute those back into the equation, we have

$$n = r^2 + s^2 = p^2(u^2 + v^2).$$

Since $(u^2 + v^2) < n$, it follows from the inductive hypothesis that p divides $(u^2 + v^2)$ an even number of times. The same remains true for $p^2(u^2 + v^2)$. □

Example. $585 = 3^2 \cdot 5 \cdot 13$. The square-free part is $5 \cdot 13$ and both prime factors 5 and 13 are congruent to 1 mod 4, in particular NOT congruent to 3 mod 4. According to the theorem, we should be able to express 585 as a sum of four integer squares. In fact,

$$5 = 2^2 + 1^2 = r^2 + s^2$$

and

$$13 = 2^2 + 3^2 = t^2 + u^2$$

And it follows from the formula in the proof of the theorem that

$$5 \cdot 13 = (r^2+s^2)(t^2+u^2) = (rt-su)^2 + (ru+st)^2 = (2\cdot 3 - 1\cdot 2)^2 + (2\cdot 2 + 3\cdot 1)^2 = 4^2 + 7^2.$$

It therefore follows that

$$585 = 3^2(4^2+7^2) = (3\cdot 4)^2 + (3\cdot 7)^2 = 12^2 + 21^2.$$

Remark. We were 'lucky' that we could easily spot that squares for 5 and 13 respectively in the example. What should we do if numbers are much bigger? Note that a positive integer n will NOT be a sum of squares if there is a prime congruent to 3 mod 4 that divides the square-free part. So if we know that no prime factor of the square-free part of n is congruent to 3 mod 4 (the theorem ascertains that n *is* a sum of squares), what we need to do is to write all prime factors congruent to 1 mod 4 (the only prime congruent to 2 mod 4 is 2 and it is $1^2 + 1^2$, while there is no prime congruent to 0 mod 4) as sums of two squares, for which we may make appeal to Hermite's algorithm, and use the product formula in the proof of the theorem.

Theorem 58 (Legendre & Gauss) Every positive integer can be written as a sum of <u>three</u> squares (of integers) except for those of the form $4^r(8z+1)$ for $r, z \geq 0$.

Theorem 59 (Lagrange) Every positive integer can be written as a sum of <u>four</u> squares (of integers).

Proof (NON-EXAMINABLE). I learned the proof from A. Baker; it illustrates Fermat's 'infinite descent argument'.

Firstly, by the formal identity

$$\begin{aligned}(x^2+y^2+z^2+w^2)(s^2+t^2+u^2+v^2) &= (xs+yt+zu+wv)^2 + (xt-ys+wu-zv)^2 \\ &+ (xu-zs+yv-wt)^2 + (xv-ws+zt-yu)^2\end{aligned}$$

that the product of two sums of four squares is again a sum of four squares. Therefore the theorem follows if we can prove that every prime number is a sum of four squares. In fact, since $2 = 1^2 + 1^2 + 0^2 + 0^2$, it suffice to prove it for an <u>odd</u> prime number.

Consider the set

$$X = \{x^2 \,|\, 0 \leq x \leq \frac{p-1}{2}\}.$$

The elements in the set are NOT congruent to each other mod p. Indeed, if $s^2 \equiv t^2 \bmod p$ where $0 \leq s, t \leq \frac{p-1}{2}$, then p would divide $s^2 - t^2 = (s+t)(s-t)$; however, since $0 \leq s+t \leq p-1$

and $-\frac{p-1}{2} \leq s-t \leq \frac{p-1}{2}$, it is evident that p does not divide $s+t$ nor $s-t$.

Similarly, the elements of the set

$$Y = \{-1-y^2 \,|\, 0 \leq y \leq \frac{p-1}{2}\}$$

are NOT congruent to each other neither (this can be proved similarly). Each of these two sets contain $1+\frac{p-1}{2}$ elements and therefore $|X|+|Y| = p+1$ elements in total. Therefore if we consider their residues mod p, there exists at least one pair of elements $(x,y) \in (X,Y)$ whose resides mod p coincide (the pigenhole principle), i.e.,

$$x^2 \equiv -1-y^2$$

mod p. Furthermore, since $x < \frac{p}{2}$ and $y < \frac{p}{2}$,

$$0 < x^2+y^2+1 < 2\left(\frac{p}{2}\right)^2 + 1 < p^2.$$

We may therefore let $x^2+y^2+1 = kp$ for some $0 < k < p$.

We now define ℓ to be the least positive integer such that $p\ell = s^2+t^2+u^2+v^2$ for some $s,t,u,v \in \mathbb{Z}$– we may and will find the smallest positive integer of the form $s^2+t^2+u^2+v^2$ that is divisible by p, and ℓ is simply its quotient by p.

Let s,t,u,v be a set of integers satisfying $p\ell = s^2+t^2+u^2+v^2$.

$\boxed{\ell \leq k < p}$ This follows by definition.

$\boxed{\ell \text{ is odd}}$ Suppose that ℓ is even. Then $s^2+t^2+u^2+v^2$ is even, hence either 0, 2, or 4 of $\{s,t,u,v\}$ are even. If at least two of them are even, we may relabel them if necessary to assume that s and t are even. In this case, $s+t, s-t, u+v, u-v$ are indeed all even! In fact, even if s,t,u,v are all odd, $s+t, s-t, u+v$ and $u-v$ are all even. Granted,

$$\left(\frac{s+t}{2}\right)^2 + \left(\frac{s-t}{2}\right)^2 + \left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2 = \frac{s^2+t^2+u^2+v^2}{2} = p\frac{\ell}{2} \in \mathbb{N}$$

contradicting the minimality of ℓ. Therefore ℓ is odd.

It remains to establish that $\boxed{\ell = 1}$. To this end, suppose that $\ell > 1$. Let $\bar{s}$ denote the residue of s by ℓ, i.e., the unique integer $0 \leq \bar{s} \leq \ell-1$ congruent to s; in fact, it is possible to choose $\bar{s}$ such that $0 \leq \bar{s} < \frac{\ell}{2}$ (since ℓ is odd, $\bar{s} = \frac{\ell}{2}$ cannot hold). Similarly define $\bar{t}, \bar{u}, \bar{v}$, and let

$$N = \bar{s}^2+\bar{t}^2+\bar{u}^2+\bar{v}^2.$$

$\boxed{\ell \text{ divides } N}$ Since ℓ divides $s^2+t^2+u^2+v^2$, it follows that $\bar{s}^2+\bar{t}^2+\bar{u}^2+\bar{v}^2$ is congruent to 0 mod ℓ.

$\boxed{N > 0}$ If $N = 0$, then $\bar{s} = \bar{t} = \bar{u} = \bar{v} = 0$, i.e., ℓ divides s, t, u and v. It would then follow that ℓ^2 divides s^2, t^2, u^2 and v^2 and consequently it divides $s^2 + t^2 + u^2 + v^2$. As the latter is ℓp, this would mean that ℓ divides p but by definition $\ell < p$ and this cannot possibly happen.

It follows that

$$N < 4\left(\frac{\ell}{2}\right)^2 = \ell^2$$

and therefore $N = r\ell$ for some integer $0 < r < \ell$. By the formal identity, the product $(r\ell)(p\ell)$ of $r\ell = \bar{s}^2 + \bar{t}^2 + \bar{u}^2 + \bar{v}^2$ and $p\ell = s^2 + t^2 + u^2 + v^2$ is again a sum of four squares. As the product is divisible by ℓ^2, it is easy to see that each of the four squares is in fact divisible by ℓ^2. Dividing through by ℓ^2, we then see that rp is a sum of four squares, but this contradicts the minimality of ℓ. □

9 Algebraic number theory

Definition. Let α be a complex number.

- α is an *algebraic number* if there is a non-zero polynomial $f(x) \in \mathbb{Q}[x]$ such that $f(\alpha) = 0$;
- α is an *algebraic integer* if there is a non-zero monic polynomial $f(x) \in \mathbb{Z}[x]$ such that $f(\alpha) = 0$.
- α is a *transcendental number* if it is NOT an algebraic number.

By a monic polynomial $f(x)$, it means that the coefficient of the highest power (=degree of f) of x is exactly 1.

Remark. By definition, an algebraic integer is an algebraic number.

Example. A rational number is an algebraic number. A rational number $r \in \mathbb{Q}$ is a root of the monic polynomial $x - r \in \mathbb{Q}[x]$. Similarly, an integer is an algebraic integer.

Example. $\alpha = \sqrt{2}$ is an algebraic integer. It is a root of the polynomial $x^2 - 2$ which is monic and has coefficients in $\mathbb{Z}$.

Example. $\alpha = \dfrac{1}{\sqrt{2}}$ is an algebraic number. Is this an algebraic integer? If $\alpha = \dfrac{1}{\sqrt{2}}$, then $\alpha^2 = \dfrac{1}{2}$, hence α is a root of the polynomial $x^2 - \dfrac{1}{2} \in \mathbb{Q}[x]$ monic but not not all coefficients are in $\mathbb{Z}$; alternatively, we may think of α as a root of the polynomial $2x^2 - 1$ with integer coefficients but it is not monic. It seems likely α is not an algebraic *integer* (the argument above is not good enough to conclude α is not an algebraic integer– we have not eliminated the possibility that there *might* be a strange monic polynomial with integer coefficients with α its root.

Example. π is a transcendental number, i.e., not an algebraic number. This is a theorem of Lindermann about 150 years ago. 'Transcendental number theory' is what A. Baker got a Fields medal (1970) for.

Proposition 60. A rational number is an algebraic integer if and only if it is an integer.

Proof. It suffices to prove that if a rational number $r = \frac{s}{t}$, with $\gcd(s,t) = 1$, is an algebraic number, then $r \in \mathbb{Z}$, i.e., $t = 1$.

By definition, r satisfies

$$r^n + c_{n-1}r^{n-1} + \cdots + c_1 r + c = 0.$$

Substituting $r = \frac{s}{t}$ and subsequently multiplying by t^n, we obtain

$$s^n + c_{n-1}s^{n-1}t + \cdots + c_1 s t^{n-1} + c t^n = 0.$$

If we write $s^n = -(c_{n-1}s^{n-1}t + \cdots + c_1 s t^{n-1} + c t^n)$, one sees that t divides the RHS and therefore also divides the LHS, s^n.

Suppose that $t > 1$ (the goal is to deduce a contradiction). Let p be a prime factor of t (which exists because $t > 1$). Since t divides s^n, the prime factor p divides s^n and it follows from Lemma 4 that p divides s. However, since p divides t, it follows that $p|\gcd(s,t)$. But $\gcd(s,t) = 1$ and this is a contradiction. □

Definition. Let α be an algebraic number. The minimal polynomial of α is the non-zero, monic polynomial $f(x) \in \mathbb{Q}[x]$ of smallest possible degree, such that $f(\alpha) = 0$.

Remark. What do we know about the minimal polynomial f of an algebraic number α?

Existence The minimal polynomial exists.

If $g(x)$ is a polynomial in $\mathbb{Q}[x]$ such that $g(\alpha) = 0$, then f necessarily divides g; indeed by 'division algorithm', there exist q and r in $\mathbb{Q}[x]$ such that $g = qf + r$ with $\deg r < \deg f$, and it follows from $g(\alpha) = 0 = f(\alpha)$ that $r(\alpha) = 0$, contradicting the minimality of degree of f!

On the other hand, f should be irreducible–it can not be factorised as a product of polynomials in $\mathbb{Q}[x]$ of smaller degrees. Indeed, if it was not irreducible (i.e. reducible) in $\mathbb{Q}[x]$, then it would contradict the minimality of the degree of f.

How do these all add up to explain the existence of f?

Since α is algebraic, there is a polynomial $f(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 \in \mathbb{Q}[x]$ such that c_n is non-zero and $f(\alpha) = 0$. Then $\frac{1}{c_n} f \in \mathbb{Q}[x]$ is a such monic polynomial. So the minimal polynomial is an irreducible 'factor/divisor' of the 'defining' polynomial of which α is a root, but it is not always easy to spot one!

Example. The minimal polynomial of $\alpha = \frac{1}{\sqrt{2}}$ is $x^2 - \frac{1}{2}$.

Example. The minimal polynomial of $\alpha \in \mathbb{Q}$ is $x - \alpha$.

Example. What about the minimal polynomial of $\alpha = \sqrt[3]{2}$? It is easy to check that α satisfies $\alpha^3 - 2$. In order for us to claim that it is indeed minimal for α, we need to know that $x^3 - 2$ irreducible in $\mathbb{Q}[x]$.

(NON-EXAMINABLE) A slick way of saying that, there is a minimal polynomial for an algebraic number, is that the ring $\mathbb{Q}[x]$ of polynomials with rational coefficients is a UFD (Unique Factorisation Domain), hence a PID (Principal Integral Domain))– we can run Euclid's algorithm with polynomials with rational coefficients. For example, a 11-th root of unity $\cos\frac{2\pi}{11}+i\sin\frac{2\pi}{11}$ is, by definition, a root of the polynomial $x^{11}-1$ but

$$x^{11}-1=(x-1)(x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x^1+1)$$

suggests that $x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x^1+1$ is a good candidate for a minimal polynomial. How do we know that it is irreducible?

Uniqueness the minimal polynomial is unique. Indeed, if f and g were two distinct monic polynomials of α, then α would be a root of $h=f-g$ with $\deg h<\deg f=\deg g$. This contradicts the minimality of the degree of f (and g)– the key point is that the minimal polynomial is monic!

Theorem 61 (Gauss's lemma) The algebraic number α is an algebraic integer if and only if its minimal polynomial has integer coefficients.

Example. Indeed, we can make appeal to Gauss's lemma to establish that $\alpha=\frac{1}{\sqrt{2}}$ is NOT an algebraic integer. As we saw earlier, α is a root of the polynomial $x^2-\frac{1}{2}$. By Gauss' lemma, we are home if we show that this is the minimal polynomial of α. In fact, since $x^2-\frac{1}{2}$ is monic, it suffices to show that there is no monic polynomial of degree $<\deg(x^2-\frac{1}{2})=2$ of which α is a root. But if α is a root of a degree 1 polynomial with rational coefficient, then α should be a rational number.

9.1 Irreducible polynomials over the rationals

I would call the following Gauss' lemma.

Theorem. Let f be a polynomial in $\mathbb{Z}[x]$ and suppose that it is monic. Suppose furthermore that there exist g,h in $\mathbb{Q}[x]$ such that $\deg g<\deg f$, $\deg h<\deg f$, and

$$f=gh.$$

Then there exist $g',h'\in\mathbb{Z}[x]$ such that g' (resp. h') is a $\mathbb{Q}$-multiple of g (resp. h) and

$$f=g'h'.$$

If f is a polynomial in $\mathbb{Z}[x]$, then

$$\boxed{f \text{ is reducible in } \mathbb{Z}[x] \Rightarrow f \text{ is reducible in } \mathbb{Q}[x]}$$

This is equivalent to the statement that

$$\boxed{f \text{ is irreducible in } \mathbb{Q}[x] \Rightarrow f \text{ is irreducible in } \mathbb{Z}[x]}$$

Gauss's lemma proves the (non-trivial) converse, assuming that f is monic. In other words, if f is a monic polynomial in $\mathbb{Z}[x]$,

$$\boxed{f \text{ is irreducible in } \mathbb{Z}[x] \Rightarrow f \text{ is irreducible in } \mathbb{Q}[x]}$$

Just because it is not possible to factorise f in $\mathbb{Z}[x]$ does not necessarily mean that it is not possible in $\mathbb{Q}[x]$, but Gauss' lemma asserts this is indeed the case. It asserts equivalently (assuming f is a polynomial in $\mathbb{Z}[x]$ and, in particular, monic) if f is reducible in $\mathbb{Q}[x]$, then it is reducible in $\mathbb{Z}[x]$.

A non-monic polynomial which is irreducible in $\mathbb{Z}[x]$ but not irreducible in $\mathbb{Q}[x]$ (a complement to Gauss' lemma) is, for example, $6x^2 - 5x + 1$. This is evidently irreducible in $\mathbb{Z}[x]$ but it factors as $6(x - \frac{1}{2})(x - \frac{1}{3})$ in $\mathbb{Q}[x]$.

(NON-EXAMINABLE) We will not prove this lemma but we use it to prove Theorem.

Firstly, we show that if the minimal polynomial (in $\mathbb{Q}[x]$) of α is an element of $\mathbb{Z}[x]$, then α is an algebraic number. This follows by definition, as if α is an algebraic number and its minimal polynomial has integer coefficients, then α is an algebraic integer.

Conversely, we show if α is an algebraic integer, then the minimal polynomial of α is an element of $\mathbb{Z}[x]$. Suppose that α is an algebraic integer. Let g be its minimal polynomial– we know that it is an element of $\mathbb{Q}[x]$ but the goal is to show that it is indeed an element of $\mathbb{Z}[x]$. By assumption, there exists a monic polynomial f in $\mathbb{Z}[x]$ such that $f(\alpha) = 0$. Seeing it as an element of $\mathbb{Q}[x]$, it follows that g divides f. To establish the divisibility, suppose that $f = gq + r$ with $\deg r < \deg g$. If r is non-zero, then it follows from $f(\alpha) = 0$ that $r(\alpha) = 0$, contradicting the minimality of g (you have seen this argument before!).

If $\deg f = \deg g$, then, while it is in theory possible that f differs from g by a non-zero scalar in $\mathbb{Q}$, both f and g are monic and therefore $f = g$. In particular f is an element of $\mathbb{Z}[x]$.

Suppose that $\deg f > \deg g$. In this case, there exists $h \in \mathbb{Q}[x]$ such that $\deg h < \deg f$ and $f = gh$. Since f and g is monic, so is h. On the other hand, it follows from the second Gauss' lemma that there exist g', h' in $\mathbb{Z}[x]$ which differ from g and h by scalars respectively such that $f = g'h'$. If we let $g' = cg$, then h' should be of the form $\frac{1}{c}h$ and this cannot possibly be an element of $\mathbb{Z}[x]$ (for example, the coefficient of the top degree term in h' is $1/c$) unless $c = \pm 1$. This means that $g \in \mathbb{Z}[x]$.

It is, hopefully, clear by now that it is important to know whether a polynomial in $\mathbb{Q}[x]$ is irreducible or not. Knowingly, there are two ways of deciding the irreducibility of a polynomial in $\mathbb{Z}[x]$ (if one is lucky).

Let

$$f = f(x) = x^n + c_{n-1}x^{n-1} \cdots + c_1 x + c_0 \in \mathbb{Z}[x]$$

be a monic polynomial with integer coefficients. We will know that f is irreducible in $\mathbb{Z}[x]$ (hence it is irreducible in $\mathbb{Q}[x]$ by Gauss) if we can

- find a prime number p such that if we let $\overline{f} = x^n + [c_{n-1}]_p x^{n-1} + \cdots + [c_1]_p x + [c_0]_p \in \mathbb{F}_p[x]$, then $\overline{f}$ is irreducible in $\mathbb{F}_p[x]$;
- find a prime number p such that p divides all c_j but p^2 does not divide c_0; if this holds, we say that f is *Eisenstein at* p.

The former is useful because it reduces our search for factors to finitely many computations. The latter is often called *Eisenstein criterion*. These are not sufficient conditions, i.e. failure to spot a such p does not mean that f is NOT irreducible in $\mathbb{Q}[x]$.

Example. We may reverse-engineer the first criterion and work out all irreducible polynomials in $\mathbb{Z}_2[x]$ first.

Degree 1: x and $x+1=x-1$.

Degree 2: x^2+x+1. To see this, we firstly observe that the monic polynomials in $\mathbb{Z}_2[x]$ of degree 2 are $x^2+x, x^2+x+1, x^2, x^2+1$; we may then eliminate the reducible ones.

Degree 3: x^3+x+1 and x^3+x^2+1. To see this, we remove from the list of 8 monic polynomials in $\mathbb{Z}_2[x]$ of degree 3 all reducible polynomials which are necessarily of the form

- either (irreducible of degree 1) $\times$ (irreducible of degree 2)
- or (irreducible of degree 1) $\times$ (irreducible of degree 1) $\times$ (irreducible of degree 1).

(NON-EXAMINABLE) We may 'inductively' complete a list of all monic irreducible polynomials in $\mathbb{F}_p[x]$ of degree n as follows. Firstly, we list all p^n monic polynomials in $\mathbb{F}_p[x]$ of degree n. For every partition $n_1+n_2+\cdots+n_k=n$ of n by positive integers, we consider all polynomials of the form (irreducible of degree n_1) $\times\cdots\times$ (irreducible of degree n_k) using the list of degree $<n$ and remove them from the list. We repeat the process for all possible partitions as above and what remains is the list of irreducible polynomials of degree n. Can you compute how many irreducibles in the list? Indeed, the number of monic irreducible polynomials of degree n in $\mathbb{F}_p[x]$ is computed by

$$\frac{1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)p^d$$

where d ranges over all integers in $[1,n]$ dividing n and μ is the Möbius function defined for a positive integer z as

$$\mu(z)=\begin{cases}1 & \text{if } z \text{ is square-free with even number of prime factors,}\\ -1 & \text{if } z \text{ is square-free with odd number of prime factors,}\\ 0 & \text{if } z \text{ is not square-free, i.e. has a squared prime factor.}\end{cases}$$

In fact, μ is related to primitive integers earlier!

Anyway, x^2+x+1 is irreducible in $\mathbb{F}_2[x]$. Any $f\in\mathbb{Z}[x]$ such that $\overline{f}=x^3+x+1$ in $\mathbb{F}_2[x]$ is irreducible. For example, $x^3+x+1, x^3+3x+1, x^3+3x+3, x^3+x+3, \ldots$ (there are of course infinitely many such polynomials in $\mathbb{Z}[x]$).

Example. Let $f(x)=x^3-2$. Then

- $\overline{f}=x^3$ is reducible in $\mathbb{F}_2[x]$.
- $\overline{f}=x^3+1=(x+1)(x^2-x+1)$ is reducible in $\mathbb{F}_3[x]$.
- $\overline{f}=(x-3)(x^2+3x+9)$ is reducible in $\mathbb{F}_5[x]$.
- $\overline{f}$ is irreducible in $\mathbb{F}_7[x]$. To see this, we argue as follows. If $\overline{f}$ was reducible in $\mathbb{F}_7[x]$, then it would factor as (irreducible of degree 1)$\times$(irreducible of degree 2) or (irreducible of degree 1)$\times$

(irreducible of degree 1) $\times$ (irreducible of degree 1). In either case, $\overline{f}$ would have a linear factor i.e. $x - a$ for some a in $\mathbb{F}_7$. In other words, it would be the case that $f(a) = 0$ for some a in $\mathbb{F}_7$. However,

a	0	1	2	3	4	5	6
$f(a)$	5	6	6	4			

hence no a in $\mathbb{F}_7$ would be a root of $f(x)$! This is a contradiction.

Remark. Just because it is *not* possible to find p such that $\overline{f} \in \mathbb{F}_p[x]$ is irreducible does NOT mean that f is *not* irreducible! For example, let $f(x) = x^4 - 10x^2 + 1$. It turns out (check it if you are interested!) that $\overline{f} \in \mathbb{F}_p[x]$ is reducible for any prime p, but f itself is actually irreducible in $\mathbb{Z}[x]$!

Example. $X^3 - 2$ is Eisenstein at 2. In fact, $x^n = 2$, for any $n \geq 2$, is Eisenstein at 2. It therefore follows that $x^n - 2$ is irreducible in $\mathbb{Z}[x]$.

One can reverse-engineer and come up with Eisenstein polynomials at p (i.e. monic irreducible polynomials) in $\mathbb{Q}[x]$ rather easily. This was important in the development of algebraic number theory.

Example. $x^{19} + 6x^{10} - 9x^4 + 75$ is Eisenstein at 3.

(NON-EXAMINABLE until the end of the section) Let us prove the legitimacy of the two irreducible criteria.

Proposition. (Reduction-mod-p-criterion). Let f be a monic polynomial in $\mathbb{Z}[x]$. If $\overline{f} \in \mathbb{F}_p[x]$ is irreducible, then f is irreducible in $\mathbb{Z}[x]$.

Proof. Suppose that f is reducible in $\mathbb{Z}[x]$ and there exist g, h in $\mathbb{Z}[x]$ such that $\deg g < \deg f$, $\deg h < \deg f$ and $f = gh$. Since f is monic, the top degree terms in g and h have coefficients both 1 or both -1. We may therefore assume WLOG that g and h are monic. Since $f = gh$, we have $\overline{f} = \overline{g}\overline{h}$ in $\mathbb{F}_p[x]$. However, since $\overline{f}$ is assumed to be irreducible, either $(\deg \overline{g}, \deg \overline{h}) = (0, \deg \overline{f})$ or $(\deg \overline{g}, \deg \overline{h}) = (\deg \overline{f}, 0)$ holds. Since g and h are monic, this implies either $\deg h = \deg f$ or $\deg g = \deg h$ contradicting assumptions on g and h. $\square$

Proposition. (Eisenstein criterion). Let f be a monic polynomial in $\mathbb{Z}[x]$. If f is Eisenstein at p, then f is irreducible in $\mathbb{Z}[x]$.

Proof. Suppose that f is reducible in $\mathbb{Z}[x]$ and there exist g, h in $\mathbb{Z}[x]$ such that $\deg g < \deg f$, $\deg h < \deg f$ and $f = gh$. Since f is monic, we may assume that g and h are monic. If we let $f = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$, then $\overline{f} = x^n$ and therefore $x^n = \overline{g}\overline{h}$ in $\mathbb{F}_p[x]$. Since $\overline{g}$ and $\overline{h}$ are still monic in $\mathbb{F}_p[x]$, we see that $\overline{g} = x^r$ and $\overline{h} = x^s$ for some integers r and s satisfying $r + s = n$. From this, it follows that g should be of the form

$$g = x^r + c_{r-1}(g)x^{r-1} + \cdots + c_1(g)x + c_0(g)$$

where p divides all $c_j(g)$, while h is of the form

$$h = x^s + c_{s-1}(h)x^{s-1} + \cdots + c_1(h)x + c_0(h)$$

where p divides all $c_j(h)$. However, $c_0 = c_0(g)c_0(h)$ and the RHS is divisible by p^2. This contradicts f being Eisenstein at p. $\square$

Both irreducibility criteria can be generalised slightly where it is no longer necessary to assume f is monic from the outset. Let $f = c_n x^n + \cdots c_1 x + c_0$ be a polynomial in $\mathbb{Z}[x]$ and assume c_n is non-zero (i.e. f is of degree n).

- Suppose that c_n is not divisible by p (e.g. $c_n = 1$, i.e. f is monic). Then if $\overline{f}$ is irreducible in $\mathbb{F}_p[x]$, then f is irreducible in $\mathbb{Z}[x]$.
- Suppose that $p \nmid c_n$. If f is Eisenstein at p, i.e. $p|c_j$ for every j but $p^2 \nmid c_0$, then f is irreducible.

The proofs in the monic case hold almost verbatim– we just have to multiply $\overline{f}$ by the inverse of c_n that exists by assumption. In both cases, to deduce the irreducibility of f in $\mathbb{Q}[x]$ from that of $\mathbb{Z}[x]$, it is necessary to have Gauss' lemma that works for non-monic polynomial. We conclude this section by stating a generalised Gauss.

Definition. A polynomial $f = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 + c_0 \in \mathbb{Z}[x]$ is said to be primitive if $\gcd(c_0, c_1, \ldots, c_n) = 1$.

Example. A monic polynomial is primitive.

Theorem. Let $f \in \mathbb{Z}[x]$ be a primitive polynomial. Suppose that there exist g, h in $\mathbb{Q}[x]$ such that $\deg g < \deg f$, $\deg h < \deg f$, and

$$f = gh.$$

Then there exist $g', h' \in \mathbb{Z}[x]$ such that g' (resp. h') is a $\mathbb{Q}$-multiple of g (resp. h) and

$$f = g'h'.$$

9.2 Quadratic number fields

Let α be an algebraic number. By definition, there exists a non-trivial polynomial with coefficients in $\mathbb{Q}$ of which α is a root. As an irreducible factor of this polynomial, there exists a minimal polynomial f in $\mathbb{Q}[x]$ of degree, say n.

Definition. Let $\mathbb{Q}(\alpha) \subset \mathbb{C}$ denote the smallest field extension of $\mathbb{Q}$ containing α.

By definition, $\mathbb{Q}(\alpha)$ is a field and therefore closed under addition and multiplication. It contains, for example, elements such as $2\alpha, 3\alpha, \ldots$ and $\alpha^2, \alpha^3, \ldots$ (or any sum/multiple of these elements!). As a vector space over $\mathbb{Q}$, the field $\mathbb{Q}(\alpha)$ is generated by the linearly independent elements $1, \alpha, \ldots, \alpha^{n-1}$.

This is an example of a *number field*. Algebraic number theory was (initially) defined as the study of properties/structure of number fields.

In theory, one can keep adding algebraic numbers: let $\overline{\mathbb{Q}}$ denote the field extension of $\mathbb{Q}$ containing all algebraic numbers. Understanding $\overline{\mathbb{Q}}$ is one of the goals of modern number theory (e.g. the Langlands program).

In what follows, we consider the case when $n = 2$.

Definition. An integer d is a square-free if p is a prime that divides d, then p^2 does not divide d. Evidently, $\sqrt{d}$ is not a rational number.

Definition. Let $\mathbb{Q}(\sqrt{d})$ denote the smallest field extension of $\mathbb{Q}$ that contains $\sqrt{d} \notin \mathbb{Q}$. More concretely,

$$\mathbb{Q}(\sqrt{d}) = \{s + t\sqrt{d} \,|\, s, t \in \mathbb{Q}\}$$

and it is a field with respect to addition:

$$(s + t\sqrt{d}) + (s' + t'\sqrt{d}) = (s + s') + (t + t')\sqrt{d} \in \mathbb{Q}(\sqrt{d})$$

and multiplication

$$(s + t\sqrt{d})(s' + t'\sqrt{d}) = (ss' + dtt') + (st' + s't)\sqrt{d} \in \mathbb{Q}(\sqrt{d});$$

if $s + t\sqrt{d}$ is a non-zero element of $\mathbb{Q}(\sqrt{d})$ and, in particular if t is non-zero, then it has multiplicative inverse:

$$\frac{s}{s^2 - dt^2} - \frac{t}{s^2 - dt^2}\sqrt{d} \in \mathbb{Q}(\sqrt{d})$$

[note that the denominator $s^2 - dt^2$ is never zero because of the assumption that d is square-free!]

Definition. Let d be a square-free integer. The set of elements α in $\mathbb{Q}(\sqrt{d})$ which are algebraic integers defines a ring. The ring is called the ring of integers of $\mathbb{Q}(\sqrt{d})$.

Proposition 62 Let d be a square-free integer. Then the ring of integers of $\mathbb{Q}(\sqrt{d})$ is

- $\mathbb{Z}[\sqrt{d}] = \{s + t\sqrt{d} \,|\, s, t \in \mathbb{Z}\}$ if $d \equiv 2, 3 \bmod 4$,
- $\mathbb{Z}\left[\dfrac{1 + \sqrt{d}}{2}\right] = \{s' + t'\dfrac{1 + \sqrt{d}}{2} \,|\, s', t' \in \mathbb{Z}\}$ if $d \equiv 1 \bmod 4$.

Proof (NON-EXAMINABLE). Let α be an element $s + t\sqrt{d}$ of $\mathbb{Q}[\sqrt{d}]$. It is a root of the polynomial

$$x^2 - 2sx + (s^2 - dt^2).$$

For α to be an algebraic integer in $\mathbb{Q}[\sqrt{d}]$, the coefficients $2s$ and $s^2 - dt^2$ both need to be integers. These conditions boil down to both s and t being integers if $d \equiv 2$ or $3 \bmod 4$, or (s, t) being of the form $(s' + \dfrac{t'}{2}, \dfrac{t'}{2})$ for some integers s', t' if $d \equiv 1 \bmod 4$. In the latter case,

$$s + t\sqrt{d} = (s' + \frac{t'}{2}) + \frac{t'}{2}\sqrt{d} = s' + t'\frac{1 + \sqrt{d}}{2}.$$

To elaborate more on 'These conditions boil down to...', we ask the following question: amongst the elements in $\mathbb{Q}(\sqrt{d})$, which α's are roots of the monic polynomial in $\mathbb{Z}$-coefficients? The question boils down to the following question:

find the set Σ of all pairs $(s, t) \in \mathbb{Q} \times \mathbb{Q}$ satisfying $2s \in \mathbb{Z}$ and $s^2 - dt^2 \in \mathbb{Z}$ simultaneously.

It turns out that

- if $d \equiv 2$ or $d \equiv 3 \bmod 4$, then

$$\Sigma = \{(s,t) \in \mathbb{Q} \times \mathbb{Q} \,|\, s \in \mathbb{Z}, t \in \mathbb{Z}\}$$

and therefore the ring of integers in $\mathbb{Q}(\sqrt{d})$ is

$$\{s + t\sqrt{d} \,|\, (s,t) \in \mathbb{Q}\} =: \mathbb{Z}[\sqrt{d}],$$

- if $d \equiv 1 \bmod 4$, then

$$\Sigma = \{(s,t) = (s' + \frac{t'}{2}, \frac{t'}{2}) \in \mathbb{Q} \times \mathbb{Q} \,|\, s' \in \mathbb{Z}, t' \in \mathbb{Z}\}.$$

and therefore the ring of integers in $\mathbb{Q}(\sqrt{d})$ is

$$\left\{\frac{s'+t'}{2} + \frac{t'}{2}\sqrt{d} \,|\, s', t' \in \mathbb{Z}\right\} = \left\{s' + t'\frac{1+\sqrt{d}}{2} \,|\, s', t' \in \mathbb{Z}\right\} =: \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$$

Suppose $\boxed{d \equiv 2 \bmod 4}$ or $\boxed{d \equiv 3 \bmod 4}$. The inclusion

$$\Sigma \supset \{(s,t) \in \mathbb{Q} \times \mathbb{Q} \,|\, s \in \mathbb{Z}, t \in \mathbb{Z}\}$$

is clear. To prove

$$\Sigma \subset \{(s,t) \in \mathbb{Q} \times \mathbb{Q} \,|\, s \in \mathbb{Z}, t \in \mathbb{Z}\},$$

we argue as follows. Let $2s = r \in \mathbb{Z}$– we only know that $s \in \mathbb{Q}$. One of the goals is to prove that $2|r$, in order for us to conclude $s \in \mathbb{Z}$. As

$$s^2 - dt^2 = \frac{r^2 - 4dt^2}{4} \in \mathbb{Z},$$

it follows that $r^2 - 4dt^2 \in 4\mathbb{Z}$. We deduce from this that, while we do not know if $t \in \mathbb{Z}$ yet, we do know that $t = \frac{u}{2}$ for some $u \in \mathbb{Z}$. To see this, we argue as follows. Since $r^2 - 4dt^2$ is, in particular, an integer and r^2 is an integer, $4dt^2 = d(2t)^2$ is an integer. It suffices to show that $2t$ is an integer. As $2t$ is a rational, we may write $2t$ as $\frac{a}{b}$ for a pair of integers a, b such that $\gcd(a,b) = 1$ and b is non-zero; the goal is to show that $b = 1$. Suppose that $b > 1$. In this case, there exists a prime number p that divides b. Since $\frac{da^2}{b^2}$ is an integer, it therefore follows that p^2 divides da^2. However, since $\gcd(a,b) = 1$, it follows that p^2 divides d. This contradicts the assumption that d is a square-free integer.

Substitute $t = \frac{u}{2}$ back into the equation above, we have

$$r^2 - du^2 \in 4\mathbb{Z}.$$

Recall that

$z \bmod 4$	$z^2 \bmod 4$
1	1
2	0
3	1
4	0

We prove $\boxed{u^2 \equiv 0 \bmod 4}$. Suppose $u^2 \equiv 1 \bmod 4$. Then $du^2 \equiv 2$ (resp. $du^2 \equiv 3$) mod 4 if $d \equiv 2$ (resp. $d \equiv 3$). It follows from $r^2 - du^2 \in 4\mathbb{Z}$ that $r^2 \equiv 2$ (resp. $r^2 \equiv 3$) mod 4. According to the table, this is not possible.

$\boxed{r^2 \equiv 0 \bmod 4}$ This follows immediately from $r^2 - du^2 \in 4\mathbb{Z}$ and $du^2 \equiv 0 \bmod 4$ from above.

According to the table, $r^2 \equiv 1 \bmod 4$ implies $r \equiv 0$ or $\equiv 2 \bmod 4$. In either case, $2|r$. Similarly for u.

The case for $d \equiv 1$ is similar but slightly harder. Suppose $\boxed{d \equiv 1 \bmod 4}$. We show the following two sets are equal:

$$\Sigma = \{(\alpha, \beta) \in \mathbb{Q} \times \mathbb{Q} \,|\, \alpha + \beta \in \mathbb{Z}, \alpha - \beta \in \mathbb{Z}\}.$$

On the other hand, the equality

$$\{(\alpha, \beta) \in \mathbb{Q} \times \mathbb{Q} \,|\, \alpha + \beta \in \mathbb{Z}, \alpha - \beta \in \mathbb{Z}\} = \{(s' + \frac{t'}{2}, \frac{t'}{2}) \in \mathbb{Q} \times \mathbb{Q} \,|\, s' \in \mathbb{Z}, t' \in \mathbb{Z}\}$$

holds by relating $(\alpha - \beta, \alpha + \beta)$ to $(s', s' + t')$, or equivalently relating (α, β) to $(s' + \frac{t'}{2}, \frac{t'}{2})$.

It therefore remains to check the equality $\Sigma = \{(\alpha, \beta) \in \mathbb{Q} \times \mathbb{Q} \,|\, \alpha + \beta \in \mathbb{Z}, \alpha - \beta \in \mathbb{Z}\}$.

The inclusion

$$\Sigma \supset \{(\alpha, \beta) \in \mathbb{Q} \times \mathbb{Q} \,|\, \alpha + \beta \in \mathbb{Z}, \alpha - \beta \in \mathbb{Z}\}$$

is easy. To prove the inclusion

$$\Sigma \subset \{(\alpha, \beta) \in \mathbb{Q} \times \mathbb{Q} \,|\, \alpha + \beta \in \mathbb{Z}, \alpha - \beta \in \mathbb{Z}\},$$

we argue as in the first case. We let $2s = r \in \mathbb{Z}$,

$$r^2 - dt^2 \in 4\mathbb{Z}$$

forces t to be of the form $t = \frac{u}{2}$ for some $u \in \mathbb{Z}$. We then have

$$r^2 - du^2 \in 4\mathbb{Z}.$$

As $d \equiv 1 \bmod 4$ this time, $r^2 \equiv 1 \bmod 4$ if and only if $u^2 \equiv 1 \bmod 4$. According to the table, this implies that $2|(r - u)$ and $2|(r + u)$. We then observe that $\alpha - \beta = \frac{r - u}{2} \in \mathbb{Z}$ and $\alpha + \beta = \frac{r + u}{2} \in \mathbb{Z}$ as desired. □

9.3 Units in the ring of integers in $\mathbb{Q}(\sqrt{d})$

Definition. Let R be a ring. An element r in R is said to be a unit if there exists s in R such that $rs = 1$.

Example. The units in $\mathbb{Z}$ are ± 1. If r and s are integers such that $rs = 1$, the only possibilities for (r, s) are $(1, 1)$ or $(-1, -1)$.

Example. $\pm 1, \pm\sqrt{-1}$ are units in $\mathbb{Z}[\sqrt{-1}]$. This is because $1 \cdot 1 = 1$, $(-1) \cdot (-1) = 1$, $\sqrt{-1} \cdot (-\sqrt{-1}) = 1$. Indeed, they are the units in $\mathbb{Z}[\sqrt{-1}]$ (to be explained shortly).

Remark. It might be useful for us to understand what units in $R = \mathbb{Z}[\sqrt{d}]$ look like. Let $r = s + t\sqrt{d}$ and $R = S + T\sqrt{d}$ where $s, t, S, T \in \mathbb{Z}$. The condition $rR = 1$ would then imply that (1) $sS + tTd = 1$ and (2) $tS + sT = 0$. It follows from (1) $\times\, s -$ (2) $\times\, td$ that $S(s^2 - dt^2) = s$ and from (1) $\times\, t -$ (2) $\times\, s$ that $T(s^2 - dt^2) = t$. To sum up,

$$S + T\sqrt{d} = \frac{s}{s^2 - dt^2} + \frac{t}{s^2 - dt^2}\sqrt{d}$$

and therefore both $\frac{s}{s^2 - dt^2}$ and $\frac{t}{s^2 - dt^2}$ should be integers. In fact, $s^2 - dt^2$ should be 1 or -1 because of this. See the forthcoming proposition [it is, of course, possible to prove this directly].

Definition. Given $\alpha = s + t\sqrt{d} \in \mathbb{Q}(\sqrt{d})$, let

$$\overline{\alpha} = s - t\sqrt{d} \in \mathbb{Q}(\sqrt{d})$$

and we call it the *conjugate* of α.

Lemma 65. Let d be a square-free integer and $\alpha, \beta \in \mathbb{Q}(\sqrt{d})$.

- $\alpha = \beta$ if and only if $\overline{\alpha} = \overline{\beta}$.
- $\alpha\overline{\alpha} \in \mathbb{Z}$ if $\alpha = s + t\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$.
- $\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}$.

Proof. This is straightforward. □

Proposition 66 Suppose that d is a square-free integer and $d \equiv 2, 3 \bmod 4$ (hence the ring of integers in $\mathbb{Q}(\sqrt{d})$ is $\mathbb{Z}[\sqrt{d}]$). An integer $\alpha = r + s\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ is a unit if and only if $|\alpha\overline{\alpha}| = 1$, or equivalently,

$$s^2 - dt^2 = \pm 1,$$

i.e., (s, t) is a solution of Pell's equation $x^2 - dy^2 = \pm 1$.

Remark(NON-EXAMINABLE). When $d \equiv 1 \bmod 4$, the ring of integers in $\mathbb{Q}(\sqrt{d})$ is $\mathbb{Z}[\frac{1 + \sqrt{d}}{2}]$. An element in $\mathbb{Z}[\frac{1 + \sqrt{d}}{2}]$ is of the of form $r = s + t\frac{1 + \sqrt{d}}{2} = s' + t'\sqrt{d}$ where $s' = s + \frac{t}{2} = \frac{2s + t}{2} \in \mathbb{Q}$ and $t' = \frac{t}{2} \in \mathbb{Q}$. Analogous to the argument in the previous remark, it follows from $rR = 1$ for $R = S + T\frac{1 + \sqrt{d}}{2} = S' + T'\sqrt{d} \in \mathbb{Z}[\frac{1 + \sqrt{d}}{2}]$ that

$$S' + T'\sqrt{d} = \frac{s'}{s'^2 - dt'^2} + \frac{t}{s'^2 - dt'^2}\sqrt{d} = \frac{s}{u} + \frac{t}{u}\frac{1 + \sqrt{d}}{2},$$

where $u = \frac{1}{4}((2s+t)^2 - dt^2) = s^2 + st + \frac{1-d}{4}t^2 \in \mathbb{Z}$ (since $d \equiv 1$ implies $\frac{1-d}{4} \in \mathbb{Z}$). Demanding $\frac{s}{u} \in \mathbb{Z}$ and $\frac{t}{u} \in \mathbb{Z}$ simultaneously is equivalent to $u = \pm 1$ [if not, there would be a prime $p > 1$ that divides u. And any power of p would then divide s and t, which is evidently a contradiction] which is markedly different from what we get when $d \equiv 2$ or $3 \bmod 4$. Note that $u = \pm 1$ is equivalent to $(2s+t)^2 - dt^2 = \pm 4$.

Proof. Suppose that $\alpha = s + t\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ is a unit. Then there exists $\beta \in \mathbb{Z}[\sqrt{d}]$ such that $\alpha\beta = 1$. By the first part of Lemma 65, $\overline{\alpha\beta} = \overline{1} = 1$. It follows from the third part of Lemma 65 then that

$$1 = (\alpha\beta)(\overline{\alpha\beta}) = \alpha\overline{\alpha}\beta\overline{\beta}.$$

From the second part of Lemma 65, $\beta\overline{\beta} \in \mathbb{Z}$ and $\alpha\overline{\alpha} = s^2 - dt^2 \in \mathbb{Z}$ and the equality in fact claims that $\alpha\overline{\alpha} = r^2 - ds^2$ is a unit in $\mathbb{Z}$. Since the units in $\mathbb{Z}$ are ± 1, we then conclude that $r^2 - ds^2 = \pm 1$.

Conversely, suppose that $r^2 - ds^2 = \pm 1$. Then $(\alpha\overline{\alpha})^2 = (r^2 - ds^2)^2 = 1$. In other words,

$$\alpha(\alpha\overline{\alpha}\,\overline{\alpha}) = 1,$$

which says that α is a unit in $\mathbb{Z}[\sqrt{d}]$. □

Example. The units in $\mathbb{Z}[\sqrt{-1}]$ are $\pm 1, \pm\sqrt{-1}$. By Proposition 66, the units in $\mathbb{Z}[\sqrt{-1}]$ are $s + t\sqrt{-1}$ such that $s^2 + t^2 = 1$ for integers s and t. The only possible pairs (s,t) are $(\pm 1, 0)$ and $(0, \pm 1)$.

Example. $\mathbb{Z}[\sqrt{3}]$ has infinitely many units. Indeed, the units in $\mathbb{Z}[\sqrt{3}]$ are of the form $(2+\sqrt{3})^n$ for n in $\mathbb{N}$. Since we know that the fundamental solution to the Pell's equation $x^2 - 3y^2 = \pm 1$ is $(s,t) = (2,1)$ and Theorem 51 asserts that every positive integer solution (v_n, w_n) is given by $v_n + w_n\sqrt{d} = (s + t\sqrt{3})^n$. Can you find more?

Example(NON-EXAMINABLE). The units in $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$ are $\{s + t\frac{1+\sqrt{-3}}{2} \mid s^2 + st + t^2 = 1\}$. To solve the equation $s^2 + st + t^2 = 1$ in $s, t \in \mathbb{Z}$, we firstly make appeal to the quadratic formula and see that $s = \frac{-t \pm \sqrt{t^2 - 4(t^2-1)}}{2} = \frac{-t \pm \sqrt{-3t^3+4}}{2}$. For s to be an integer, there are two cases to follow:

$\boxed{t \text{ is even}}$ In this case, $-3t^2 + 4$ should be of the form $(2a)^2$ for some $a \in \mathbb{Z}$. It then follows from the equation $(2a)^2 + 3t^2 = 4$ that $t^2 = 0$ (as t is meant to be an even integer) and therefore that $-3t^3 + 4 = 4$ and $s = \frac{\pm\sqrt{4}}{2} = \pm 1$ as a result.

$\boxed{t \text{ is odd}}$ In this case, $-3t^2 + 4$ is should be of the form $(2a+1)^2$ for some $a \in \mathbb{Z}$. It then follows form the equation $(2a+1)^2 + 3t^2 = 4$ that $t^2 = 1$ (as t is meant to be an odd integer), i.e. $t = \pm 1$. If $t = 1$, then $-3t^3 + 4 = 1$ and $s = \frac{-1 \pm \sqrt{1}}{2} = 0$ or 1; if $t = -1$, thenm $-3t^2 + 4 = 1$ and $s = \frac{1 \pm \sqrt{1}}{2} = 1$ or 0.

In conclusion, the units in $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$ are elements of the form $s + t\frac{1+\sqrt{-3}}{2}$ where (s,t) is $(1,0),(-1,0),(0,1),(-1,1),(1,-1)$ or $(0,-1)$.

The following theorem proves the structure of solutions of Pell's equation $x^2 - dy^2 = \pm 1$.

Theorem 67 (Dirichlet's unit theorem for a real quadratic field; NON-EXAMINABLE) Let d be a square-free positive integer congruent to 2 or 3 mod 4. The group of units in the ring $\mathbb{Z}[\sqrt{d}]$ of integers of $\mathbb{Z}[\sqrt{d}]$ is isomorphic to

$$\{\pm 1\} \times \mathbb{Z}.$$

This is a distilled form of what Dirichlet actually proved for F in 1846 (apparently, P. G. L. Dirichlet came up with a proof during a concert in the Sistine Chapel in Rome).

Theorem 68 (Dirichlet's unit theorem for a number field; NON-EXAMINABLE) Let F be a number field. Let $r_{\mathbb{R}}$ (resp. $2r_{\mathbb{C}}$) be the number $|\mathrm{Hom}_{\mathbb{Q}}(F,\mathbb{R})|$ (resp. $|\mathrm{Hom}_{\mathbb{Q}}(F,\mathbb{C})|$) of real embeddings (of pairs of complex conjugate embeddings). The group of units in F is finitely generated by $r = r_{\mathbb{R}} + r_{\mathbb{C}} - 1$ generators of infinite order, i.e., the group of units in F is isomorphic to

$$\mu \times \mathbb{Z}^r$$

where μ is the finite cyclic group of roots of unity.

Remark. If $r_{\mathbb{R}} > 0$, then $\mu = \{\pm 1\}$ as ± 1 are the only roots of unity in $\mathbb{R}$. Even if $r_{\mathbb{R}} = 0$, we still have $\mu = \{\pm 1\}$; for example $\mathbb{Z}[\sqrt{-1}]$ has units $\{\pm 1, \pm\sqrt{-1}\}$.

Remark. The unit group is finite if and only if $r = 0$, i.e., $(r_{\mathbb{R}}, r_{\mathbb{C}}) = (1,0)$ or $(0,1)$, i.e, $F = \mathbb{Q}$ or $\mathbb{Q}(\sqrt{d})$ with $d < 0$. When d is a negative square-free integer, the group of units in the ring of integers of $\mathbb{Q}(\sqrt{d})$ is $\{\pm 1\}$ except when $d = -1$ in which case it is $\{\pm 1, \pm\sqrt{-1}\}$ and when $d = -3$ and there are 6 units in $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$.