NUMBER THEORY REVISION

Euclid's algorithm

Let a, b be integers but assume that $b > 0$. Then there exist integers q and $0 \leq r < b$ such that

$$a = bq + r.$$

This is fundamental.

If $r = 0$, we write $b|a$ ('b divides a').

Definition. Let a, b be integers. The Greatest Common Divisor $d = \gcd(a, b)$ is *defined to be* a <u>non-negative</u> integer satisfying the properties

- $d|a$ and $d|b$,
- if $e|a$ and $e|b$, then $e|d$.

Example. For any integer $n \geq 0$, then $\gcd(n, 0) = n$.This follows by unravelling the definition of gcd above.

Remark. The fact that gcd is defined to be non-negative integer helps us prove

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

One can show

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

This is the idea underlying Euclid's algorithm.

Proposition 1 asserts that if a, b be integers and d be a positive integer, then TFAE:

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

The proof is constructive and uses Bezout's identity/Euclid's algorithm: given a pair of integers a, b, there exist r, s in $\mathbb{Z}$ such that

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

Congreunces

Definition. If a and b are integers, write $a \equiv b \bmod N$ if N divides $a - b$.

Date: December 11, 2023.

Remark. This/the 'mod N congreunce' defines an equivalence relation on $\mathbb{Z}$.

Definition. By $\mathbb{Z}/N\mathbb{Z}$, we mean the set of equivalence ('mod N congruence') classes

$$[a]_N = \{r \in \mathbb{Z} \,|\, r \equiv a \bmod N\}$$

mod N, with addition $[a]_N + [b]_N = [a+b]_N$ and multiplication $[a]_N[b]_n = [ab]_N$, with respect to which $\mathbb{Z}/N\mathbb{Z}$ defines a ring.

Remark. When N is a prime number p, we write $\mathbb{F}_p$ for $\mathbb{Z}/p\mathbb{Z}$.

Be comfortable with the equivalent formulation

- $a \equiv b \bmod N$,
- $[a]_N = [b]_N$ in $\mathbb{Z}/N\mathbb{Z}$.

Proposition 8 is a 'mod N' analogue of Proposition 1 and asserts the following: given $a, N \in \mathbb{N}$ and $b \in \mathbb{Z}$, the following are equivalent

- the congruence equation $ax \equiv b \bmod N$ is soluble (i.e. can 'get rid of a' and find $x \equiv c \bmod N$ for some c)
- $\gcd(a, N)$ divides b.

In particular, if $\gcd(a, N) = 1$, we can always solve the congruence equation $ax \equiv b \bmod N$. This can be proved as follows (it contains the essence of what this proposition is about): since $\gcd(a, N) = 1$, it follows from Bezout/Euclid that there exist $r, s \in \mathbb{Z}$ such that $ar + sN = \gcd(a, N) = 1$. Hence $ar \equiv 1 \bmod N$. Multiplying r on the both side of the equation, $x \equiv 1x \equiv arx \equiv rb \bmod N$.

Q. If, on the other hand, $\gcd(a, N) > 1$, what do we do?

The Chinese Remainder Theorem ('how to solve a system of linear congruence equations'). Be comfortable with solving more than two equations.

Q. If one of them says $ax \equiv b \bmod N$, what do we do?

Euler's totient function and primitive roots

Euler's totient function $\phi : \mathbb{N} \to \mathbb{N}$ which sends $N \in \mathbb{N}$ to the number of integers $1 \leq z \leq N$ such that $\gcd(z, N) = 1$.

Proposition 14 shows that

$$|(\mathbb{Z}/N\mathbb{Z})^\times| = \phi(N),$$

where for a ring R, we mean by $R^\times$ the (multiplicative) subgroup of units, i.e. the set of r in R such that there exists s in R satisfying $rs = 1$.

Example. $\phi(p) = p - 1$, hence $|\mathbb{F}_p^\times| = p - 1$, In fact,

$$\mathbb{F}_p^\times = \{[1], [2], \ldots, [p-1]\}.$$

Theorem 17 proves formulas for ϕ:

- If p is a prime and $r > 0$, then $\phi(p^r) = p^{r-1}(p-1)$.
- If a and b are coprime, i.e. $\gcd(a, b) = 1$, then $\phi(ab) = \phi(a)\phi(b)$.
- $\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$

Theorem 15 proves for $N \in \mathbb{N}$ and $z \in \mathbb{Z}$ such that $\gcd(z, N) = 1$ that

$$z^{\phi(N)} \equiv 1$$

mod n. This generalises (the proof of) Fermat's Little Theorem (Theorem 7): for any integer z co-prime to a prime number p, then $z^{p-1} \equiv 1 \bmod p$. Combined with the case $z \equiv 0 \bmod p$, the FLT asserts that

$$z^p \equiv z$$

mod p.

In theory, we can use the FLT for a primality test: Suppose that you are given an integer n (so large that it is difficult to check all possible divisors) and that you are interested in finding out whether it is a prime number or not. We can say that it is NOT a prime number if we can spot an integer z coprime to n such that z^{n-1} is NOT congruent to 1 mod n.

Example. 12 is NOT a prime number because $2^{11} = 2048 \equiv 8 \bmod 12$.

Definition. The order of an integer z is the smallest positive integer d such that $z^d \equiv 1 \bmod N$. If $[z] \in \mathbb{Z}/n\mathbb{Z}$ denotes the congruence class mod N represented by z, i.e., the set of integers congruent to z mod N, then the definition can be paraphrased as the smallest power such that $[z]^d = [z^d] = [1]$.

Theorem 15, Lemma 19 and Proposition 20 in combination prove for $z \in \mathbb{Z}$ such that $\gcd(z, N) = 1$ that if d is the order of z mod N, then d divides $\phi(N)$. If you are interested in computing the order of z mod N, we just need to search through the divisors of $\phi(N)$.

We specialise to 'congruence mod p'.

Definition. An integer z is a primitive root mod p if z has order $p - 1$. From the statement above, it means that z has max possible order.

We saw lots of examples of primitive roots mod p.

Is there, really, a primitive root mod p for any p? If so, how many? What about those integers of order $d < p - 1$ mod p? How many? Theorem 22 answers these questions. The number of integers $1 \leq z \leq p - 1$ of order $1 \leq d \leq p - 1$ is exactly $\phi(d)$.

A key observation: if z is a primitive root mod p (it exists and there are indeed $\phi(p) = p - 1$ of them according to Theorem 22), then

$$\mathbb{F}_p^\times = \{[1], \ldots, [p-1]\} = \{[z], [z]^2, \ldots, [z]^{p-2}\},$$

holds, i.e., a primitive root mod p defines a cyclic generator of the (multiplicative) group $\mathbb{F}_p^\times$.

Quadratic residues and non-residues

Suppose $\boxed{p > 2}$.

Definition. For an integer a not divisible by p, a is a quadratic residue mod p if there exists an integer z such that $z^2 \equiv a$ mod p, or equivalently the congruence equation $x^2 \equiv a$ mod p is soluble.

If $a \equiv b$ mod p, a is a quadratic residue mod p if and only if b is.

Definition.

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

Theorems 25 asserts if p is an odd prime, then:

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

(Rule 1) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$

(Rule 2) $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$

(Rule 3) $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$

(Rule 4) $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$ for any pair of distinct odd primes p and q.

Proposition 27 (Euler's criterion) shows that if p does not divide a, then $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \bmod p$. The key observation is used crucially in proving this.

We use Euler's criterion to solve congruence equations of the form $x^2 \equiv a \bmod p$ (the Legendre symbol only tells you the solubility).

Proposition 28 asserts if $p \equiv 3 \bmod 4$ and $\left(\frac{a}{p}\right) = 1$, then $a^{(p+1)/4}$ is a solution to the equation $x^2 \equiv a \bmod p$.

Proposition 29 asserts if $p \equiv 1 \bmod 4$ and $\left(\frac{a}{p}\right) = -1$, then $a^{(p-1)/4}$ is a solution to the equation $x^2 \equiv -1 \bmod p$.

Both $a^{(p+1)/4}$ and $a^{(p-1)/4}$ are typically large! We discussed (repeatedly) how to simply these values $\boxed{\text{mod } p}$.

Finite continued fractions

Definition. Given $\alpha, \alpha_1, \ldots, \alpha_{N-1} \in \mathbb{Z}$ and $\alpha_N \in \mathbb{R}$ (where $N \geq 1$), we write

$$[\alpha; \alpha_1, \ldots, \alpha_N]$$

to mean

$$\alpha + \cfrac{1}{\alpha_1 + \cfrac{1}{\ddots + \cfrac{1}{\alpha_{N-1} + \cfrac{1}{\alpha_N}}}}$$

By definition, this is a rational number.

Proposition 31 shows conversely that, given a rational number in its lowest terms, one can write it as $[\alpha; \alpha_1, \ldots, \alpha_N]$ for some N, and Theorem 32 in fact shows that the continued fraction expression is indeed unique.

Proposition 31 is proved in two different ways, namely by induction and by an argument that involves Euclid's algorithm. We turn the latter into an algorithm:

$$\begin{array}{ccc}
\alpha = \lfloor r \rfloor & \Rightarrow & \rho_1 = \dfrac{1}{r-\alpha} \\
 & \swarrow & \\
\alpha_1 = \lfloor \rho_1 \rfloor & \Rightarrow & \rho_2 = \dfrac{1}{r_1-\alpha_1} \\
 & \swarrow & \\
 & \vdots & \\
 & \swarrow & \\
\alpha_{N-1} = \lfloor \rho_{N-1} \rfloor & \Rightarrow & \rho_N = \dfrac{1}{\rho_{N-1}-\alpha_{N-1}} \in \mathbb{N} \\
 & \swarrow & \\
\alpha_N = \lfloor \rho_N \rfloor = \rho_N & &
\end{array}$$

When the input r is a rational number, the algorithm stops as soon as we see a positive integer ρ_N.

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

$$\begin{array}{ccc}
\alpha = \lfloor \dfrac{87}{38} \rfloor = 2 & \Rightarrow & r_1 = \dfrac{1}{\frac{87}{38}-2} = \dfrac{38}{11} \\
 & \swarrow & \\
\alpha_1 = \lfloor \dfrac{38}{11} \rfloor = 3 & \Rightarrow & r_2 = \dfrac{1}{\frac{38}{11}-2} = \dfrac{11}{5} \\
 & \swarrow & \\
\alpha_2 = \lfloor \dfrac{11}{5} \rfloor = 2 & \Rightarrow & r_3 = \dfrac{1}{\frac{11}{5}-2} = \dfrac{5}{1} \in \mathbb{N} \\
 & \swarrow & \\
\alpha_3 = \lfloor \dfrac{5}{1} \rfloor = 5 = r_3 & &
\end{array}$$

The algorithm works for negative rationals! The only difference is that the first term 'α' will inevitably a negative integer.

Let $r = [\alpha, \alpha_1, \ldots, \alpha_N]$. For $0 \leq n \leq N$, define

$$\begin{array}{rcl} r_0 & = & [\alpha;] \\ r_1 & = & [\alpha; \alpha_1] \\ & \vdots & \\ r_n & = & [\alpha; \alpha_1, \ldots, \alpha_n] \\ & \vdots & \\ r_N & = & [\alpha; \alpha_1, \ldots, \alpha_N] \end{array}$$

and call them the convergents to r.

It turns out that it is possible to rewrite this in terms of pairs s_n, t_n for $0 \leq n \leq N$, which are defined as

- $s_{-1} = 0, s_0 = \alpha$,
$$s_n = \alpha_n s_{n-1} + s_{n-2},$$
- $t_{-1} = 0$, $t_0 = 1$,
$$t_n = \alpha_n t_n + t_{n-2}.$$

Proposition 33 asserts that, for $0 \leq n \leq N$, we may write $r_n = [\alpha; \alpha_1, \ldots, \alpha_n]$ as

$$r_n = \frac{s_n}{t_n}.$$

Definition. The (positive) rational number r_n is called the n-th convergent to r.

Theorem 34 proves how the s_n' and the t_n's are related.

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

As a corollary, we see that 'even (resp. odd) indexed' convergents define an increasing (resp. decreasing) sequence of positive integers:

$$r_0 < r_2 < r_4 < \cdots < r_{2i} < r_{2j-1} < \cdots < r_5 < r_3 < r_1.$$

What if we know more than finitely many

$$\alpha, \alpha_1, \dots, \alpha_N$$

and know that infinitely many

$$\alpha, \alpha_1, \dots$$

positive integers (α may be negative)? Does it make sense to think about $[\alpha; \alpha_1, \dots]$? Would it make sense to consider the limit of $[\alpha; \alpha_1, \dots, \alpha_N]$ as N tends to ∞?

Theorem 36 answers 'Yes'! More precisely, the convergents $r_n = [\alpha; \alpha_1, \dots, \alpha_n] = \dfrac{s_n}{t_n}$ do converge to a limit in $\mathbb{R}$ (no longer in $\mathbb{Q}$ as in the finite case) as $n \to \infty$.

Theorem 37 characterises what these infinite length continued fractions look like. It proves that every irrational number can be written as $[\alpha; \alpha_1, \dots]$, and Theorem 39 says that the expression is unique.

Amazingly, just as in the case of finite continued fractions, the algorithm we use work for irrationals!

Example. $r = \sqrt{2}$.

$$\begin{array}{ccc}
\alpha = \lfloor \sqrt{2} \rfloor = 1 & \Rightarrow & \rho_1 = \dfrac{1}{\sqrt{2}-1} = 1 + \sqrt{2} \\
 & \swarrow & \\
\alpha_1 = \lfloor 1 + \sqrt{2} \rfloor = 2 & \Rightarrow & \rho_2 = \dfrac{1}{(1+\sqrt{2}) - 2} = \dfrac{1}{\sqrt{2}-1} = \rho_1 \\
 & \swarrow & \\
\alpha_2 = \alpha_1 & \Rightarrow & \rho_3 = \rho_2 = \rho_1 \\
 & \swarrow & \\
 & \vdots &
\end{array}$$

Hence $\sqrt{2} = [\alpha; \alpha_1, \alpha_2, \dots] = [1; 2, 2, \dots]$.

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

$$\begin{array}{ccc}
\alpha = \lfloor\sqrt{3}\rfloor = 1 & \Rightarrow & \rho_1 = \dfrac{1}{\sqrt{3}-1} = \dfrac{1+\sqrt{3}}{2} \\
 & \swarrow & \\
\alpha_1 = \lfloor\dfrac{1+\sqrt{3}}{2}\rfloor = 1 & \Rightarrow & \rho_2 = \dfrac{1}{\frac{1+\sqrt{3}}{2}-1} = \dfrac{2}{\sqrt{3}-1} = 1+\sqrt{3} \\
 & \swarrow & \\
\alpha_2 = \lfloor 1+\sqrt{3}\rfloor = 2 & \Rightarrow & \rho_3 = \dfrac{1}{(\sqrt{3}+1)-2} = \dfrac{1}{\sqrt{3}-1} = \rho_1 \\
 & \swarrow & \\
\alpha_3 = \alpha_1 & \Rightarrow & \rho_4 = \rho_2 \\
 & \swarrow & \\
 & \vdots &
\end{array}$$

Hence $\sqrt{3} = [\alpha; \alpha_1, \alpha_2, \dots] = [1; 1, 2, 1, 2, \dots]$.

We often see ‘periodic’ continued fractions (e.g. $\sqrt{2}, \sqrt{3}$). Write

$$[\alpha; \alpha_1, \dots, \alpha_{N-1}, \overline{\alpha_N, \dots, \alpha_{N+l-1}}]$$

if

$$\begin{array}{lcccc}
[\alpha; \alpha_1, \dots, \alpha_{N-1}, & \alpha_N, & \alpha_{N+1}, & \dots, & \alpha_{N+l-1}, \\
 & \| & \| & & \| \\
 & \alpha_{N+l}, & \alpha_{N+l+1}, & \dots & \alpha_{N+2l-1}, \\
 & \| & \| & & \| \\
 & \alpha_{N+2l} & \dots & & \dots \quad]
\end{array}$$

Examples. $\sqrt{2} = [1; \overline{2}], \sqrt{3} = [1; \overline{1, 2}]$.

On the other hand, if we are given a periodic continued fraction, we can work out what it looks like as a real number.

Example. $[\overline{1; 2}]$.

Let $r = [\overline{1;2}]$. Then

$$r = 1 + \cfrac{1}{2 + \cfrac{1}{[\overline{1;2}]}} = \cfrac{1}{2 + \cfrac{1}{r}} = \frac{3r+1}{2r+1}.$$

It therefore follows that $r(2r+1) = 3r+1$, i.e., $2r^2 - 2r - 1 = 0$. By the quadratic formula, the solutions for the quadratic equation is

$$\frac{1 \pm \sqrt{3}}{2}.$$

Since $r > 0$, $r = \dfrac{1 \pm \sqrt{3}}{2}$.

This simple example contains the essence of Theorem 46 which asserts that a real number has a periodic continued fraction if and only if it is a quadratic irrational, i.e., it is of the form $s + t\sqrt{d}$ where $s, t \in \mathbb{Q}$, t is non-zero and $d > 1$ is square-free.

Given an irrational number r, we are interested in finding a 'good' approximation in $\mathbb{Q}$ to r. What should we mean by 'good'?

Definition. Let r be an irrational number. A rational number $\dfrac{s}{t}$ is a good approximation to r if

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

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

Remark. It means that 'there is no rational number closer to r than $\dfrac{s}{t}$ is to r with smaller denominator. If $\dfrac{s'}{t'}$ has a smaller denominator than that of $\dfrac{s}{t}$, it has to be further from r than $\dfrac{s}{t}$ is to r. This is exactly what it says in the the definition.

In fact, Theorem 43 proves that if r is irrational, $[\alpha;\alpha_1,\dots]$ is its continued fraction and r_n is the n-th convergent $[\alpha;\alpha_1,\dots,\alpha_n]$, then r_n is a good approximation to r for $n \geq 2$, i.e.,

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

for any $\frac{s'}{t'}$ with $t' < t_n$.

Theorem 44 asserts that a rational number sufficiently close to an irrational number r is inevitably a convergent to r: let r be an irrational number and let $s, t \in \mathbb{Z}$, $t > 0$ and $\gcd(s,t) = 1$. If

$$|r - \frac{s}{t}| < \frac{1}{2t^2},$$

then $\frac{s}{t}$ is a convergent to r.

Pell equations

Given an integer d that is not a square, we are interested in solving the equation

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

more precisely, finding a pair (s,t) of positive integers satisfying $s^2 - dt^2 = 1$ or -1.

Remark Why d should not be a square? If $d = 0$, the equation is $x^2 = \pm 1$ and that is boring. If d is a square $d = b^2 > 1$ say, then the equation is $x^2 - dy^2 = x^2 - b^2y^2 = (x+by)(x-by) = \pm 1$. Therefore $(x+by, x-by)$ is either $(1,1)$, $(-1,1)$, $(1,-1)$ or $(-1,-1)$. Each one of the possibilities would then allows us to determine what x, y and d should be.

Remark. Why $s > 0$ and $t > 0$? If (s,t) is a solution, so is any of $(-s,t), (s,-t), (-s,-t)$.

Remark. Why $d > 0$? If $d < 0$, then $-d > 0$. In this case, $x^2 + (-d)y^2 = -1$ has no solutions. On the other hand, if $d < -1$, then $-d > 1$ and the solutions to $x^2 + (-d)y^2 = 1$ are $(x, y) = (1, 0)$ and $(-1, 0)$; if $d = -1$, then the solutions to $x^2 + (-d)y^2 = x^2 + y^2 = 1$ are $(1, 0), (-1, 0), (0, 1), (0, -1)$.

It makes sense to consider only a non-square positive integer d and look for positive integer solutions to $x^2 - dy^2 = \pm 1$.

Theorem 47 asserts that

$$\left\{\text{The (positive) integer solutions to } x^2 - dy^2 = \pm 1\right\} \subset \{(s_n, t_n)\},$$

where s_n, t_n is defined by the n-th convergent $r_n = \dfrac{s_n}{t_n}$ to $\sqrt{d}$.

We did two examples: $x^2 - dy^2 = \pm 1$ where $d = 2$ or 3– we computed the convergents

$$r_0 = \frac{s_0}{t_0},\ r_1 = \frac{s_1}{t_1},\ r_2 = \frac{s_2}{t_2}, \dots$$

to $\sqrt{d}$ and checked which ones indeed defines solutions to $x^2 - dy^2 = \pm 1$. The point was that NOT all of them did! It is therefore natural for us to ask if we can single out exactly which convergents $r_n = \dfrac{s_n}{t_n}$ satisfy $s_n^2 - dt_n^2 = \pm 1$.

For a square-free positive integer d, it is known that

$$\sqrt{d} = [\alpha; \overline{\alpha_1, \alpha_2, \dots, \alpha_2, \alpha_1, 2\alpha}].$$

Theorem 48 (I've written the proof in the notes; the proof is NON-EXAMINABLE) asserts that if $\sqrt{d} = [\alpha; \overline{\alpha_1, \dots, \alpha_l}]$ (a form a lot less precise than above, but this still does the job), then

$$\begin{aligned}
&\left\{\text{The (positive) integer solutions to } x^2 - dy^2 = \pm 1\right\} \\
&= \{(s_{Nl-1}, t_{Nl-1}) \mid N = 1, 2, \dots\} \\
&= \{(s_{l-1}, t_{l-1}), (s_{2l-1}, t_{2l-1}), (s_{3l-1}, t_{3l-1}), \dots\}.
\end{aligned}$$

Moreover,

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

In other words, not only we can single out solutions to $x^2 - dy^2 = \pm 1$, we can also single out exactly which one are solutions to $x^2 - dy^2 = +1$ and which ones are to $x^2 - dy^2 = -1$!

We pushed forward and introduced the concept of the fundamental solution to $x^2 - dy^2 = \pm 1$.

Definition. If (s, t) and (s', t') are solutions to $x^2 - dy^2 = \pm 1$, then define

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

if $s + t\sqrt{d} < s' + t'\sqrt{d}$ in $\mathbb{R}$. It then, and only then, makes sense to define the smallest solution (with respect to $<$ defined above) to $x^2 - dy^2 = \pm 1$ to be the fundamental solution. By definition, the fundamental solution does exist.

Remark. Theorem 48 proves that, if $\sqrt{d} = [\alpha; \overline{\alpha_1, \ldots, \alpha_l}]$, the solutions $\underline{\text{are}}$ $(s_{l-1}, t_{l-1}), (s_{2l-1}, t_{2l-1}), \ldots$. Since we know by definition,

$$0 < s_{l-1} < s_{2l-1} < \ldots$$

and

$$0 < t_{l-1} < t_{2l-1} < \ldots,$$

the fundamental solution $\underline{\text{is}}$ (s_{l-1}, t_{l-1})!

Furthermore, $\boxed{\text{the fundamental solution generates all the solutions to } x^2 - dy^2 = \pm 1}$. Theorem 51 and Theorem 52 made precise what this means.

Theorem 51 proves that if $(s, t) = (s_{l-1}, t_{l-1})$ is the fundamental solution to $x^2 - dy^2 = \pm 1$ and if we define (v_n, w_n) by

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

then $v_n^2 - dw_n^2 = \pm 1$, i.e., (v_n, w_n) defines a solution to $x^2 - dy^2 = \pm 1$. Furthermore, we know for which n defines $v_n^2 - dw_n^2 = +1$ and which does $v_n^2 - dw_n^2 = -1$: if $\epsilon = s^2 - dt^2 \in \{\pm 1\}$,then $v_n^2 - dw_n^2 = \epsilon^n$. To sum up,

$$\{(v_n, w_n)\} \subset \Big\{\text{The (positive) integer solutions to } x^2 - dy^2 = \pm 1\Big\}$$

This does not quite prove what $\boxed{\text{the fundamental..}}$ says. But Theorem 52 concludes it; it proves that if (v, w) is a solution to $x^2 - dy^2 = \pm 1$, then $(v, w) = (v_n, w_n)$ for some n. In other words,

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

Combining these two, we have another way of completely describing the solutions to the Pell equation:

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

Example. $x^2 - 61y^2 = \pm 1$. As

$$\sqrt{61} = [7; \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}],$$

the length $l = 11$ and therefore the fundamental solution is $(s, t) = (s_{11-1}, t_{11-1}) = (s_{10}, t_{10})$. Unfortunately, there is no easy way to compute the first/smallest solution. Starting with (s_0, t_0), we recursively compute convergents. It turns out that $(s_{10}, t_{10}) = (29718, 3805)$ and

$$s_{10}^2 - 61t_{10}^2 = 29718^2 - 61 \cdot 3805^2 = -1.$$

According to Theorem 51,

$$v_1 + w_1\sqrt{61} = (s + t\sqrt{61})^1$$

so $(v_1, w_1) = (s, t) = (s_{10}, t_{10})$, but

$$v_2 + w_2\sqrt{d} = (s + t\sqrt{61})^2 = (29718 + 3805\sqrt{61})^2 = ...$$

and

$$v_2 - 61w_2^2 = (-1)^2 = 1.$$

To sum up, (v_1, w_1) is the smallest solution (the fundamental solution) to $x^2 - 61y^2 = \pm 1$; (v_2, w_2) is the second smallest solution to $x^2 - 61y^2 = \pm 1$ while it is also the smallest solution to $x^2 - 61y^2 = 1$ (as $v_1^2 - 61w_1^2 = -1$ and this is the only smaller solution to $x^2 - dy^2 = \pm 1$).

Furthermore, we know from Theorem 48 that $(v_2, w_2) = (s_{2\cdot 11-1}, t_{2\cdot 11-1}) = (s_{21}, t_{21})$. It is easy to compute (v_2, w_2) using Theorem 51, but it is far more laborious to compute (s_{21}, t_{21}) by hand, simply following the definition (as you might have notice in Assessed coursework 4)! It is possible to extrapolate the trick and compute (s_n, t_n) when n is very big (computing such a thing is useful in approximating numbers), especially when l is very big.

Sums of squares

If p is a prime number, can we express p as a sum of two integer-squares?, i.e., can we solve $x^2 + y^2 = p$ in $(x, y) \in \mathbb{N} \times \mathbb{N}$?

Proposition 53 asserts that if $p \equiv 3 \bmod 4$, then $x^2+y^2 = p$ has no solutions in $(x, y) \in \mathbb{N} \times \mathbb{N}$.

Theorem 54/Corollary 56 asserts that if $p \equiv 1 \bmod 4$, then $x^2 + y^2 = p$ has a solution in $(x, y) \in \mathbb{N} \times \mathbb{N}$.

We can axiomatise the proof of Theorem 54 to solve the equation when $p \equiv 1 \bmod 3$ (Hermite's algorithm).

Step 1: find $z \in \mathbb{Z}$ such that $z^2 \equiv -1 \bmod p$.

Step 2: compute convergents $\frac{s_n}{t_n}$ to $\frac{z}{p}$ and find n satisfying

$$t_n < \sqrt{p} < t_{n+1}.$$

Then $(x, y) = (t_n, ps_n - zt_n)$ or $(t_n, zt_n - ps_n)$ defines a solution.

More sums of squares.

Theorem 57 asserts that a positive integer n is the sum of $\underline{\text{two}}$ squares if the square free part of n has no prime factors congruent to 3 mod 4.

We can turn the proof into an algorithm that solves $x^2 + y^2 = n$ for $x, y \in \mathbb{N} \times \mathbb{N}$ *when the square-free part of n has no prime factors congruent to* 3 *mod* 4. Indeed, the identity

$$(r^2 + s^2)(t^2 + u^2) = (rt - su)^2 + (ru + st)^2$$

allows us to reduce the computation to solving $x^2 + y^2 = p$ for every prime factor p in the square-free part of n. By definition, p is either 2 (in which case $1^2 + 1^2 = 2$) or is congruent to 1 mod 4 and we may make appeal to Hermite's algorithm to solve the equation $x^2 + y^2 = p$.

Theorem 59 asserts that every positive integer is a sum of four squares. I wrote a proof but it is NON-EXAMINABLE.

Algebraic Number Theory

Definition. A complex number α is an algebraic number (resp. algebraic integer) if there exists $f(x) \in \mathbb{Q}[x]$ (resp. $f(x) \in \mathbb{Z}[x]$) such that $f(\alpha) = 0$ (resp. such that f is monic and $f(\alpha) = 0$).

Proposition 60 asserts that, in $\mathbb{Q}$, the algebraic numbers are the integers.

Definition. If α is an algebraic number, the minimal polynomial is a non-zero monic polynomial $f(x)$ in $\mathbb{Q}[x]$ of smallest possible degree such that $f(\alpha) = 0$.

The minimal polynomial exists and it does so uniquely: given an algebraic number α, there exists a polynomial with rational coefficients of which α is a root; the minimal polynomial is the irreducible polynomial (with rational coefficients) of smallest possible degree that divides of the polynomial. We need to check a candidate polynomial is really minimal/irreducible!

Gauss' lemma (Theorem 61) allows us to characterise algebraic integers in terms of minimal polynomials: an algebraic number is an algebraic integer if and only if its minimal polynomial has integer coefficients.

Gauss's lemma is often useful in proving a given algebraic number is NOT an algebraic integer (to prove a given algebraic number is an algebraic integer, it is only necessary to spot a monic polynomial with integer coefficients and a redundant to check if it is minimal).

Within the field $\mathbb{Q}(\sqrt{d}) = \{s + t\sqrt{d} \,|\, s, t \in \mathbb{Q}\}$, it is possible to describe exactly the ring of algebraic integers (Proposition 63)

$$\mathbb{Z}[\sqrt{d}]$$

if $d \equiv 2$ or 3 mod 4 and

$$\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$$

if $d \equiv 1$ mod 4.

Definition. Let R be a ring. An element r in R is a unit if there exists s in R such that $rs = 1$.

We are interested in understanding the units in the ring of integers of $\mathbb{Q}(\sqrt{d})$.

Proposition 66 asserts that if d is a square-free integer congruent to 2 or 3 mod 4, an algebraic integer $\alpha = s + t\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ is a unit if and only if $s^2 - dt^2 = \pm 1$, i.e., (s, t) is a solution to Pell's equation $x^2 - dy^2 = \pm 1$.

This proposition says that one can use what we learned in the section about Pell's equation to understand the units.