Number theory, geometry and algebra

J. B. Cooper
Johannes Kepler Universität Linz

Contents

1 Number theory

1.1 Prime numbers and divisibilty:

We shall work with the following number systems:

$$\begin{aligned} \mathbf{N} &= \{1, 2, 3, \dots\} \qquad (1)\\ \mathbf{N}_0 &= \{0, 1, 2, 3, \dots\} \qquad (2)\\ \mathbf{Z} &= \mathbf{N}_0 \cup \{-1, -2, -3, \dots\}. \qquad (3) \end{aligned}$$

We regard them as algebraic systems with the operations + und . of addition and multiplication.

$(\mathbf{N}, +)$ is a commutative semigroup (without unit);
$(\mathbf{N}, .)$ is a commutative semigroup with unit 1;
$(\mathbf{N}_0, +)$ is a commutative semigroup with unit 0;
$(\mathbf{Z}, +, .)$ is a ring.
(See the last chapter for these notions).

Divisibility: If we have two numbers $a, b \in \mathbf{Z}$, then we say that a is a **divisor** of b (written: $a|b$) if there exists $r \in \mathbf{Z}$ with $a = rb$. If further $a \neq b$, then a is a **proper divisor of** b. The following simple properties are self-evident:

a) If $a|b_i$ $(i = 1, \dots, n)$, $c_1, \dots, c_n \in \mathbf{Z}$, then $a|\sum c_i b_i$;
b) $b|a$ and $c \in \mathbf{Z} \Rightarrow bc|ac$;
c) $ac|bc$ $(a, b, c \in \mathbf{Z},\ c \neq 0) \Rightarrow a|b$;
d) $b|a \Rightarrow |b| \leq |a|$ $(a, b \in \mathbf{Z},\ a \neq 0)$;
e) $a|b$ and $b|a \Rightarrow |a| = |b|$ $(a, b \in \mathbf{Z} \setminus \{0\})$;
f) $a|b,\ b|c \Rightarrow a|c$ $(a, b, c \in \mathbf{Z})$.

The division algorithm: Suppose $a \in \mathbf{Z}$, $b \in \mathbf{N}$. Then there are unique numbers $q \in \mathbf{Z}$ and $r \in \mathbf{N}_0$ with $r < b$, so that $a = qb + r$

PROOF. Existence: q is the largest number in the set $\{t \in \mathbf{Z} : t \leq \frac{a}{b}\}$. For $q \leq \frac{a}{b} < q+1$ and so $qb \leq a < b(q+1)$. Then let $r = a - bq$ so that $0 \leq r < b$ and $a = qb + r$.

Uniqueness: Let $a = qb+r = q_1b+r_1$ with $r_1 \geq r$. Then $(q-q_1)b = r_1-r$. But $0 \leq r_1 - r < b$. Thus $q = q_1$ etc.

■

Applications: 1. b-adic development: We use a fixed $b \in \mathbf{N}$. Then we have: every $c \in \mathbf{N}$ has a unique representation of the form

$$c = a_0 + a_1 b + \cdots + a_n b^n$$

($n \in \mathbf{N}$, $a_0, \ldots, a_n \in \mathbf{N}_0$ with $a_i < b$ for each i). (This is the so-called b-adic representation of a— the most important special cases are $b = 2$ resp. $b = 10$).

PROOF. a_0 is the remainder from the division $a = q_0 b + a_0$. This means that $q_0 = \frac{a-a_0}{b}$. a_1 is determined as the remainder $q_0 = q_1 b + a_1$. Hence $q_1 = \frac{q_0 - a_1}{b}$. One continues in the obvious manner.

■

2) The greatest common divisor If $a, b \in \mathbf{Z}$, then: $d \in \mathbf{N}$ is the **greatest common divisor** of a, b (written: $\gcd(a, b)$), when the following hold:

a) $d|a$ and $d|b$

b) $\bigwedge_{d_1 \in \mathbf{N}} d_1|a$ and $d_1|b \Rightarrow d_1|d$

d can be calculated as follows (whereby we assume without loss of generality that a and b are positive):

Put

$$a = qb + a_1.$$

It is clear that $\gcd(b, a_1) = \gcd(a, b)$. Now let $a_1 < b_2$ and

$$b = q_1 a_1 + a_2 \tag{4}$$

$$\vdots \tag{5}$$

The procedure is continued until we arrive at two numbers a_i, a_{i+1}, for which

$$a_{i+1} = q_i a_i + 0,$$

i.e. a_{i+1} is a multiple of a_i. Then $\gcd(a, b) = \gcd(a_i, a_{i+1})$ and the latter is clearly a_i. This construction provides the additional fact that the greatest common divisor can be written in the form: $ra + sb$ for $r, s \in \mathbf{Z}$. d is in fact the smallest positive number which can be written in this form.

Example We calculate $\gcd(191, 35)$ as follows:

$$191 = 5.35 + 16 \quad (\to (35, 16)) \tag{6}$$

$$35 = 2.16 + 3 \quad (\to (16, 3)) \tag{7}$$

$$16 = 3.5 + 1 \quad (\to (3, 1)) \tag{8}$$

Hence $\gcd(191, 35) = 1$. Further

$$
\begin{aligned}
1 &= 16 - 5.3 && (9)\\
&= 16 - 5.(35 - 2.16) && (10)\\
&= 11.16 - 5.35 && (11)\\
&= 11(191 - 5.35) - 5.35 && (12)\\
&= 11.191 - 60.35 && (13)
\end{aligned}
$$

The following properties of the greatest common divisor are self-evident or follow easily from the above considerations:

1) $a, b, c \in \mathbf{Z}$, $m \in \mathbf{Z} \setminus \{0\} \Rightarrow \gcd(am, bm) = |m| \gcd(a, b)$

2) If $a, b, c \in \mathbf{Z}$ with $a \neq 0$ and $\gcd(a, b) = 1$, then $a|bc \Rightarrow a|c$.

(For there exist r, s with $ra + sb = 1$. Hence $rac + sbc = c$. Since a divides the left hand side, we have $a|c$.)

3) The equation $ax + by = c$ is solvable (i.e. given a, b, we can find x, y with this property) $\Leftrightarrow \gcd(a, b)|c$.

Definition: $a, b \in \mathbf{Z}$ are **relatively prime**, if $\gcd(a, b) = 1$.

In this case we can find $r, s \in \mathbf{Z}$ with $ra + sb = 1$. This implies that every $n \in \mathbf{Z}$ has a representation $r'a + s'b$.

We can carry over these considerations for pairs of numbers to general finite sequences thereof. Thus we define

$$\gcd(a_1, \dots, a_n)$$

to be the smallest $d \in \mathbf{N}$ which has a representation $\sum a_i r_i$

We can then calculate $\gcd(a_1, \dots, a_n)$ recursively in the obvious way since

$$\gcd(a_1, \dots, a_n) = \gcd(a_1,\ \gcd(a_2, \dots, a_n)).$$

$\{a_1, \dots, a_n\}$ are **relatively prime**, if $\gcd(a_1, \dots, a_n) = 1$. $\{a_1, \dots, a_n\}$ are **pairwise relatively prime**, if $\bigwedge_{i \neq j} \gcd(a_i, a_j) = 1$. (The second condition is stronger).

Definition: A whole number $p > 1$ is **prime**, if its only divisors are ± 1 and $\pm p$. Otherwise p is **composite**. The first few prime numbers are $2, 3, 5, 7, 11, 13 \dots$ If n is composite, then it clearly has a prime number as factor (this can be proved formally by induction).

Proof. 4 is non-prime and has the prime factor 2 Suppose that $n > 4$ and make the induction hypothesis that every composite number $n_1 < n$ has a prime factor. If n itself is not prime, then it has a positive factor n_1 with $1 < n_1 < n$. If n_1 is prime, then we are finished. If not, then it has a prime factor by the induction hypothesis.

■

The sieve of Eratosthenes: This is a method of calculating algorithmically all prime numbers up to a given natural number N. We write out the numbers $1, \dots, N$ and erase successively the multiples of 2 (except, of course, for 2 itself), then all remaining multiples of 3 and so on for each prime number $\leq \sqrt{N}$. The remaining numbers are the required primes.

Example: $N = 25$

$$2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\ 16\ 17\ 18\ 19\ 20\ 21\ 22\ 23\ 24\ 25$$

Step 1:

$$2\ 3\ \underline{4}5\ \underline{6}\ 7\ \underline{8}\ 9\ \underline{10}\ 11\ \underline{12}\ 13\ \underline{14}\ 15\ \underline{16}\ 17\ \underline{18}\ 19\ \underline{20}\ 21\ \underline{22}\ 23\ \underline{24}\ 25$$

Step 2:

$$2\ 3\ 5\ 7\ \underline{9}\ 11\ 13\ 17\ \underline{18}\ 19\ \underline{21}\ 23\ \underline{24}\ 25$$

Step 3:

$$2\ 3\ 5\ 7\ 11\ 13\ 17\ 19\ 23\ \underline{25}$$

The result: 2 3 5 7 11 13 17 19 23.

We now show that there are infinitely many prime numbers. Suppose that there are only finitely many, say $p_1, \dots, p_k$. Consider the number

$$N = p_1 \dots p_k + 1.$$

There are two possibilities—either N is prime and then we have found a prime number which is not on our list, or N has a prime factor p. Since p is a factor of N and members of our list cannot have this property, we have again found a prime which is not on the list. □

We can improve this result as follows:

Proposition 1 $\sum_{p \in P} \frac{1}{p} = \infty$ *where P denotes the family of all prime numbers).*

Proof. Suppose that the sum is finite. Then

$$\prod_{p \in P} (1 + \frac{1}{p}) \geq \sum_{n \in \mathbf{N}_{ns}} \frac{1}{n}$$

where $\mathbf{N}_{ns} = \{n \in \mathbf{N} : \bigwedge_{p \in P} p^2 \nmid n\}$. Hence it suffices to show that $\sum_{n \in \mathbf{N}_{ns}} \frac{1}{n} = \infty$. But

$$(\sum_{n \in \mathbf{N}_{ns}} \frac{1}{n})(\sum_{n \in \mathbf{N}} \frac{1}{n^2}) \geq \sum_{n \in \mathbf{N}} \frac{1}{n}$$

and

$$\sum \frac{1}{n^2} < \infty \text{ bzw. } \sum \frac{1}{n} = \infty.$$

■

Lemma 1 *If p is a prime and $p|ab$, where $a, b \in \mathbf{Z}$, then: $p|a$ or $p|b$.*

Proof. If $p \nmid a$, then $\gcd(p, a) = 1$, i.e. there are r, s with $pr + as = 1$ and so $pbr + abs = b$. We then have $p|b$ (since $p|pbr + abs$).

■

Corollar 1 *If p is prime and $a_1, \dots, a_n \in \mathbf{Z}$, then*

$$p|a_1 \dots a_n \Rightarrow \bigvee_i p|a_i.$$

With these preliminaries, we can now prove the so-called **fundamental theorem of number theory**:

Proposition 2 *Suppose that $n \in \mathbf{N}$ and write $p_1, p_2, \dots$ for the sequence of prime numbers. Then there exists for each i a uniquely determined $\alpha_i \in N_0$, so that $n = \prod p_i^{\alpha_i}$. (It is clear that $\alpha_i = 0$ for almost all values of i, i.e. the product finite).*

Proof. We prove firstly the existence (by induction). The case $n = 1$ is clear. Suppose that the proposition is valid for each $n_1 < n$ and consider the case n. If n is prime, then it is itself such a factorisation. If not, then it has a prime number p as factor. Put $n_1 = \frac{n}{p}$. n_1 has a suitable factorisation and hence so has $n = n_1 p$.

Uniqueness: Suppose that $n \in \mathbf{N}$ has two factorisations:

$$n = \prod_i p_i^{\alpha_i} = \prod_i p_i^{\beta_i}.$$

We show that $\alpha_i = \beta_i$ for each i. If this were not the case, then there would exist i with $\alpha_i \neq \beta_i$, say $\alpha_i < \beta_i$. We cancel the factor p^{α_i} from both factorisations and arrive at the equation

$$\prod_{j \neq i} p_j^{\alpha_j} = p_i^{\beta_i - \alpha_i} \prod_{j \neq i} p_i^{\beta_j}.$$

But p_i is a factor of the right-hand side and so:

$$p| \prod_{j \neq i} p_j^{\alpha_j} \text{ and hence } p_i | p_j^{\alpha_j}$$

for any j. This is a contradiction.

■

$\prod p_i^{\alpha_i}$ is called **prime factorisation** of n.

If

$$a = \prod_i p_i^{\alpha_i} \text{ and } b = \prod_i p_i^{\beta_i}$$

then

$$\gcd(a, b) = \prod_i p_i^{\gamma_i}$$

whereby $\gamma_i = min(\alpha_i, \beta_i)$. Analogously, the number

$$\prod_i p_i^{\delta_i},$$

where $\delta_i = max(\alpha_i, \beta_i)$ is the smallest positive whole number which has both a and b as factor. It is thus called the least common multiple of a and b, written: $\operatorname{lcm}(a, b)$.

We use the above factorisation to prove the following result:

Suppose that $a, b, c \in \mathbf{N}$ are such that $c = (ab)^n (n \in \mathbf{N})$, whereby $\gcd(a, b) = 1$. Then a and b are also n-th powers i.e. $\bigvee_{c_1, c_2 \in \mathbf{N}} a = c_1^n \wedge b = c_2^n$.

For let $a = \prod_{p \in P_1} p^{\alpha_p}$ and $b = \prod_{p \in P_2} p^{\alpha_p}$, where P_1 (resp. P_2) are sets of primes. Then $P_1 \cap P_2 = \emptyset$ (since $\gcd(a, b) = 1$). Tthe condition $ab = c^n$ easily implies that $p \in P_1 \cup P_2 \Rightarrow n | \alpha_p$. It is then clear that a and b are n-th powers.

Further applications of the division algorithm: 1) Let $g_0, g_1, \dots$ be a sequence of positive whole numbers with $g_0 = 1$, $g_n \geq 2 (n \geq 1)$. Then every $x \in \mathbf{N}$ has a unique representation:

$$x = a_n (g_n \cdot \dots \cdot g_1) + a_{n-1} (g_{n-1} \cdot \dots \cdot g_n) + \dots + a_1 g_1 + a_0,$$

with $a_n \neq 0$, $0 \leq a_i \leq g_i - 1$.
2) An **egyptian representation** of a rational number $r \in]0,1[$ is one of the form

$$\frac{1}{n_1} + \cdots + \frac{1}{n_k},$$

with $n_i \in \mathbf{N}$, for instance

$$\frac{2}{3} = \frac{1}{2} + \frac{1}{6} \tag{14}$$

$$\frac{3}{10} = \frac{1}{5} + \frac{1}{10} \tag{15}$$

$$\frac{3}{7} = \frac{1}{3} + \frac{1}{11} + \frac{1}{231} \tag{16}$$

$$= \frac{1}{4} + \frac{1}{7} + \frac{1}{28} \tag{17}$$

The last example shows that such a representation need not be unique. However, we have

Proposition 3 *Every rational number $r \in]0,1[$ has a representation*

$$r = \frac{1}{d_0} + \frac{1}{d_0 d_1} + \cdots + \frac{1}{d_0 \ldots d_n},$$

$(d_0, d_1, \ldots, d_n \in \mathbf{N})$.

PROOF. Let $r = \frac{p}{q}$. The d's are determined recursively by means of the following scheme: Put

$$q = d_0 p - p_1 \text{ with } 0 \leq p_1 < p \quad (\text{ so that } \frac{p}{q} = \frac{1}{d_0} + \frac{1}{d_0}(\frac{p_1}{q})) \tag{18}$$

$$q = d_1 p_1 - p_2 \text{ with } 0 \leq p_2 < p_1 \tag{19}$$

$$\vdots \tag{20}$$

$$q = d_r p_r \tag{21}$$

(the algorithm breaks off, when ? divides ?). Then

$$\frac{p}{q} = \frac{1}{d_0} + \frac{1}{d_0 d_1} + \cdots + \frac{1}{d_0 \ldots d_i} + \frac{p_{i+1}}{q d_0 \ldots d_i}$$

as can easily be shown by induction.

■

We round up this chapter with some remarks on the set Q of rational numbers which we define as the quotient space $\mathbf{Z} \times \mathbf{Z}^*$ ($\mathbf{Z}^* = \mathbf{Z} \setminus \{0\}$) under the equivalence relation

$$(m,n) \sim (m_1,n_1) \Leftrightarrow mn_1 = nm_1.$$

Q is a field (see the last chapter) with the operations

$$[(m,n)] + [(m_1,n_1)] = [(mn_1 + nm_1, nn_1)]; \tag{22}$$
$$[(m,n)].[(m_1,n_1)] = [(mm_1, nn_1)]. \tag{23}$$

We provide Q with an ordering as follows:

$$[(m,n)] \leq [(m_1,n_1)] \Leftrightarrow mn_1 \leq nm_1$$

where we chose representations with $n > 0,\ n_1 > 0$. Each element $q \in Q$ has a unique representation $\pm \frac{m}{n}$, with $n \in \mathbf{N}_0,\ n \in \mathbf{N}$ where $(m,n) = 1$ (and so has a uniue representation of the form

$$\pm \prod_i p_1^{\alpha_1} \dots\ p_r^{\alpha_r} \dots \quad (\alpha_i \in \mathbf{Z})).$$

Then $q \in \mathbf{Z} \Leftrightarrow \alpha_i \geq 0$ for each i. This representation can be written as follows:

$$q = \prod_{p \in P} p^{w(p)},$$

where P is the set of prime numbers and $w(p)$ is the index of p in the representation of q. We have:
1) $(\bigwedge_{p\in P} w_p(q) = w_p(q_1)) \Rightarrow q = \pm q_1$
2) $w_p(qq_1) = w_p(q) + w_p(q_1)$
3) $w_p(q + q_1) = min(w_p(q), w_p(q_1))$ falls $q + q_1 \neq 0)$.

b-adic representation: Let $g \in \mathbf{N}$, $g > 1$. Each rational number $q \in]0,1[$ has a representation

$$q = \frac{c_1}{g} + \frac{c_2}{g^2} + \dots,$$

where $0 \leq c_i < g$. This can be demonstrated as follows: Let $q = \frac{a}{b}$. We determine the coefficients according to the following scheme which uses the division algorithm. $a = r_0$. c_1, r_1 are determined by the equation $gr_0 = c_1 b + r_1$, c_2, r_2 by $gr_1 = c_2 b + r_2$ and so on (whereby $0 \leq r_i < b$ and $0 \leq c_i < g$).

Such a representation is **finite** if there is an N so that $c_n = 0$ for $n > N$. In the case of an infinite representation i.e. of the form

$$0, c_1 c_2 \dots\ c_n (g-1)(g-1)(g-1)\dots$$

then we can rewrite it in the finite form.

$$(= 0, c_1. \dots .(c_n + 1)00\dots).$$

1.2 Integritätsbereiche

Wir bringen eine abstrakte Version des Fundamentalsatzes über Primzahlen. Der obige Beweis läßt sich fast wörtlich übertragen.

Im folgenden bezeichnet R einen **kommutativen** Ring mit Einheit. $x \in R$ ist ein **Nullteiler**, falls $y \in \mathbf{R}$ mit $x \cdot y = 0$ existiert. x ist eine **Einheit**, falls $y \in R$ mit $x \cdot y = 1$ existiert. R ist ein **Integritätsbereich**, falls R keine Nullteiler besitzt.

Falls $a, b \in R$ (mit $a \neq 0$), dann ist a **ein Teiler** von b, falls c existiert, mit $a = bc$ (gesch. $a|b$). Falls $a|b$ und $b|a$ dann sind a und b **assoziert**. Mit (a) bezeichnen wir das von a erzeugte Hauptideal d.h. $\{ar\colon r \in R\}$. Es gilt damit: $a|b \Leftrightarrow (a) \supset (b)$, bzw. a und b sind assoziert $\Leftrightarrow (a) = (b)$.

Das Element u ist genau dann eine Einheit, wenn $(u) = R$ gilt. Falls $a = bu$ wobei u eine Einheit ist, dann sind a und b assoziert. Das umgekehrt gilt, falls R ein Integritätsbereich ist.

Definition: Ein Element $c (\neq 0)$ aus R ist **irreduzierbar**, falls gilt: Wenn c eine Darstellung $a \cdot b$ hat, dann ist entweder a oder b eine Einheit. $p \in R$ ist **prim** falls gilt: $p|ab \Rightarrow p|a$ oder $p|b$.

Beispiel: In $\mathbf{Z}_6$ ist $2 = 2 \cdot 4$. Daher ist 2 zwar prim, aber nicht irreduzibel.

Bemerkung: p ist genau dan prim, wenn $I = (p)$ ein Primideal ist (ein Ideal $I \subset \mathbf{R}$ ist prim, wenn gilt: $ab \in I \Rightarrow a \in I$ oder $b \in I$). c ist irreduzibel $\Leftrightarrow I = (c)$ maximal in der Familie aller echten Hauptidealen.

Man sieht leicht, daß jedes Primelement irreduzierbar ist. Falls R ein Hauptidealbereich ist, (d.h. ein Bereich, in dem jedes Ideal ein Hauptideal ist), dann gilt:

$$p \quad \text{prim} \quad \Leftrightarrow p \quad \text{irreduzierbar}$$

Definition: Ein Integritätsbereich R ist ein **Bereich mit eindeutiger Faktorisierung**, falls gilt:

- 1) Jedes $a \in \mathbf{R}$ hat eine Darstellung $a = c_1, \ldots, c_n$, wobei die c_i irreduzierbar sind.

- 2) Falls a zwei solche Darstellungen

  $$a = c_1 \ldots c_n = d_1 \ldots d_m$$

  hat, dann gilt: $n = m$ und es gibt eine Umnummerierung so, daß c_i und d_i assoziert sind $(i = 1, \ldots, n)$.

Proposition 4 *Falls R ein Hauptidealbereich ist und*

$$I_1 \subset I_2 \subset I_3 \subset \ldots$$

eine steigende Kette von Idealen ist, dann ist diese Kette stationär, d.h. es existiert n so daß $I_m = I_n$ $(m \geq n)$.

Proposition 5 *Jedes Hauptideal ist ein Bereich mit eindeutiger Faktorisierung.*

Beispiel: Es gibt Bereiche mit eindeutiger Faktorisierung, die aber keine Hauptidealbereiche sind.

Definition: Ein Ring R ist **euklidisch**, falls eine Funktion $\phi\colon \mathbf{R}\backslash\{0\} \to \mathbf{N}$ existiert, sodaß

- a) $a, b \in R$, $ab \neq 0 \Rightarrow \phi(a)\phi(ab)$;

- b) $a, b \in R$, $b \neq 0 \Rightarrow$ es existieren $q, r \in \mathbf{R}$ mit $r = 0$ oder $\phi(r) < \phi(b)$, sodaß $a = qb + r$

Falls R ein Integritätsbereich ist, dann heißt R ein **euklidischer Bereich.**

Beispiel: $\mathbf{Z}$, mit $\phi(n) = |n|$
Ein Körper K mit $\phi(x) = 1$;
Der Ring $K(X)$ der Polynome über ein Körper K, mit $\phi(p) = \deg p$;
Der Ring der Gauß'schen Zahlen mit $\phi(m + in) = m^2 + n^2$

Proposition 6 *Jeder euklidischer Ring ist ein Hauptideal. Damit ist jeder euklidischer Bereich ein Bereich mit eindeutiger Faktorisierung.*

1.3 The ring $\mathbf{Z}_m$, residue classes

$\mathbf{Z}$ with the operation of addition ist an abelian group and with the two operations of addition and multiplication it is a commutative ring with unit.

Proposition 7 *A subset $A \subset R$ is a subgroup of $(\mathbf{Z}, +) \Leftrightarrow A$ is an ideal in the ring $(\mathbf{Z}, +, .)$. We then have*

$$\bigvee_{m \in \mathbf{N}_0} A = m\mathbf{Z},$$

(i.e. A is a principal ideal).

Proof. $\Leftarrow$ is clear.

$\Rightarrow$: If $x \in A$, $n \in \mathbf{N}_0$, then

$$nx = (x + \dots\ + x) \in A \quad ((x + \dots\ + x) \ \ n\text{-mal})$$

For $n \in \{-1, -2, \dots\}$ we have $nx = -(-nx) \in A$.

For the second part we can assume that $A \neq \{0\}$. Let m be the smallest positive element of A. We show that $A = m\mathbf{Z}$. First of all $m\mathbf{Z} \subset A$. If $x \in A$ and we write

$$x = qd + d_1,$$

with $0 \leq d_1 < d$ (the division algorithm). Then: $d_1 = x - qd \in A$, and so $d_1 = 0$.

■

This implies that the only quotient groups of $(\mathbf{Z}, +)$ resp. quotient rings of $(\mathbf{Z}, +, .)$ are $\mathbf{Z}_m$ $(m \in \mathbf{N}_0)$, where

$$\mathbf{Z}_m = \mathbf{Z}/m\mathbf{Z}$$

(i.e. $\mathbf{Z}|_\sim$, where $x \sim y \Leftrightarrow m|x - y$.) We write $x = y$ (mod$m$), if $m|x - y$.

The following properties are then evident:

$a = b$ (mod$m$) and $t|m \Rightarrow a = b$ (mod$|t|$),

$a = b$ (mod$m$), $r \in \mathbf{Z} \Rightarrow ra = rb$ (mod$m$),

$a = b$ (mod$m$), $c = d$ (mod$m$) $\Rightarrow a + c = b + d$ (mod$m$), $a - c = b - d$ (mod$m$), $ac = bd$ (mod$m$),

$ar = br$ (mod$m$) $\Rightarrow a = b$ (mod$(\frac{m}{d})$ (where $d =$ gcd(r, m)),

$ar = br$ (mod$m$) $\Rightarrow a = b$ (mod$m$), provided that gcd$(m, r) = 1$,

$a = b$ (mod$m$) $\Rightarrow ac = bc$ (modcm) $(c > 0)$,

$a = b$ (mod$m$) $\Rightarrow$ gcd$(a, m) =$ gcd(b, m),

$a = b$ (mod$m$), $0 \leq |b - a| < m \Rightarrow a = b$,

$a = b$ (mod$m$), $a = b$ (mod$n$) $\Rightarrow a = b$ (modmn), if m and n are relatively prime.

Applications: I. Rules for divisibility: Suppose that n is a natural number with decimal expansion $\sum_{i=0}^{p} a_i \, 10^i$. One verifies easily that $n = n_1$ (mod3) (even (mod9)), whereby $n_1 = \sum_{i=0}^{k} a_i$, i.e. n is divisible by 3 (resp. 9) if and only if the same holds for n_1. Similarly,

$$n = a_0 \pmod 2 \tag{24}$$

$$n = 10a_1 + a_3 \pmod 4 \tag{25}$$

$$n = a_0 \pmod 5 \tag{26}$$

$$n = \sum_{i=0}^{k} (-1)^{[\frac{i}{3}]} 10^{i-3[\frac{i}{3}]} \pmod 7 \tag{27}$$

$$n = 100a_2 + 10a_1 + a_0 \pmod 8 \tag{28}$$

$$n = \sum_{i=0}^{k} (-1)^i a_i \pmod{11} \tag{29}$$

$$n = \sum (-1)^{[\frac{i}{3}]} 10^{i-3[\frac{i}{3}]} \pmod{13} \tag{30}$$

(For

$$\begin{aligned} 10^3 &= -1 \pmod 7 && (31)\\ 10^3 &= -1 \pmod{13} && (32)\\ 10 &= -1 \pmod{11} && (33)\\ 10^2 &= 1 \pmod{11} \quad \text{usw.)} && (34) \end{aligned}$$

Fermat and Mersenne Primes: Fermat Primes: Fermat was of the opinion that all natural numbers of the form $2^{2^n}+1$ are prime. In fact, $641|2^{2^5}+1$ as we shall now show. $641 = 5.2^7+1$, i.e. $5.2^7 = -1 \pmod{641}$. Hence $5.2^{28} = 1 \pmod{641}$. But $641 = 625 + 16$ and so $5^4 = -2^4 \pmod{641}$. Hence: $-2^{32} = -2^4.2^{28} = 5.2^{28} = 1 \pmod{641}$, i.e. $641|2^{2^5}+1$.

Prime numbers of the form $2^{2^n}+1$ are called **Fermat Primes**. We remark that if a prime number p is of the form 2^r+1, then r must be a power of 2 and hence p is a Fermat Prime (otherwis, the prime factorisation of r would contain an odd prime q and

$$(a^q+1) = (a+1)(a^{q-1} - a^{g-2} + a^{q-3} - \dots + 1).$$

(We remark that whenever $m > n$, then

$$2^{2^n}+1|2^{2^m}-1.$$

This implies that the numbers $F_n = 2^{2^n}+1$ are pairwise relatively prime).

A prime number of the form $2^n - 1$ is called a **Mersenne prime.** In this case, n must also be prime. (For $m|n \Rightarrow 2^m - 1|2^n - 1$ according to the well-known scheme

$$(a^r - b^r) = (a-b)(a^{r-1} + \dots + b^{r-1}).$$

The Mersenne numbers (those of the form $2^p - 1$), are **the** candidates for large prime numbers (the record as of 1994 is $2 - 1$). We shall now show that not all Mersenne numbers are prime by proving that $47|2^{23}-1$. For $2^{23} = (2^5)^4.2^3$ and $2^5 = 32 = -15$ (mod47). Hence $2^{10} = 225 = -10$ (mod47) and so $2^{20} = 100 = 6$ (mod47). Thus $2^{23} = 6.8 = 48 = 1$ (mod47).

Equations We shall now consider polynomial equations of the form $P(x) = 0$ (modm) in $\mathbf{Z}_m$. The following examples show that the behaviour of such equations is completely different from the case of equations over $\mathbf{R}$ or $\mathbf{C}$.

Example The equation $2x = 3$ (mod4) has no solutions. The equation $x^2 = 1$ (mod8) has four solutions.

The case of linear equations in one unknown is dealt with the in following theorem

Proposition 8 *The equation $ax = b$ (modm) has a solution if and only if* $\gcd(a, m)|b$. *If this holds, then it has d solutions* ($d = \gcd(a, m)$).

PROOF. Necessity: Let x be a solution. Then

$$d|ax, \ d|m$$

and so $d|b$ (since $b = ax - cm$).

Suppose now that $d|b$. $d = ar + ms$ for suitable r, s so that

$$b = \frac{b}{d}(d) = a(\frac{b}{d})r + ms\frac{b}{d}$$

i.e. $x = \frac{b}{d}r$ is a solution.

We calculate the number of solutions as follows: If x_0 is a solution, then the numbers $\{t, t+\frac{m}{d}, \dots, t+(d-1)\frac{m}{d}\}$ form a complete list of the solutions. In particular, if a and m relatively prime, then the equation has exactly one solution.

$\blacksquare$

We can draw the following conclusions from these facts:
a) $\mathbf{Z}_m$ is a field if and only if m is prime.
b) for a general m we denote by $\mathbf{Z}_m^*$ the multiplicative group of the ring $\mathbf{Z}_m$ (i.e. the set of all invertible elements). $\mathbf{Z}_m^*$ consists of $[r_1], \dots, [r_s]$, where $r_1, \dots, r_s$ is a list of the numbers from $\{1, \dots, n-1\}$ which are relatively prime to m. We write $\phi(m)$ for the number of such elements i.e. $\phi(m) = |\mathbf{Z}_m^*|$.

$$n \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \tag{35}$$

$$\phi(n) \quad 1 \quad 1 \quad 2 \quad 2 \quad 4 \quad 2 \quad 6 \tag{36}$$

Proposition 9 *(Fermat's t Let p be a prime number. Then for any a with* $\gcd(a, p) = 1$ *we have $a^{p-1} = 1$ (modp). More generally $a^p = a$ (modp) for any a. Still more generally, for any $m \in \mathbf{N}$ and any a with* $\gcd(a, m) = 1$ *we have*

$$a^{\phi(m)} = 1.$$

PROOF. $\mathbf{Z}_m^*$ is a group of cardinality $\phi(m)$. Thus we have for any $a \in \mathbf{Z}_m^*$, $a^{\phi(m)} = 1$.

$\blacksquare$

Proposition 10 *Wilson's theorem Let p be a prime number. Then $(p-1)! = -1 \pmod{p}$.*

Proof. Suppose that p is odd. If $0 < a < p$, then there exists a unique number a' with $0 < a' < p$ and $aa' = 1$ (modp). a and a' are distinct, except for the case $a = 1$ or $a = p - 1$ ($\mathbf{Z}_p$ is a field). Then

$$1.2.\dots.(p-1) = 1.(p-1)(2.3.4.\dots.(p-2)) = p - 1 = -1 \pmod{p}$$

since we can pair each a from $\{2, 3, \dots, p-2\}$ with the corresponding a'.

■

Simultaneous Systems: We now consider systems of the form

$$x = c_i \pmod{m_i} \quad (i = 1, \dots, n)$$

whereby the m_i are pairwise relatively prime. Let $M = m_1 \dots m_n$, $M_i = \frac{M}{M_i}$ (i.e. $M_i = \prod_{j \neq i} m_j$). Then $\gcd(M_i, m_i) = 1$. Hence the equation $M_i y_i = 1 \pmod{m_i}$ is solvable. Let $\overline{y}_i$ be a solution and put $x = \sum_i c_i M_i \overline{y}_i$. This is clearly a solution of the above system. This proves the existence part of the following result:

Proposition 11 *(the Chinese remainder theorem) The system $x = c_i \pmod{m_i}$ $(i = 1, \dots, n)$ is always solvable and the solution is unique mod M i.e. two solutions x, y satisfy the equation $x = y \pmod{M}$).*

Proof. Uniqueness: Since $x = c = y \pmod{m_i}$, we know that $m_i | x - y$ for each i and so $M | x - y$. □

■

1.4 Polynomial equations:

We consider now equations of the form

$$P(x) = 0 \pmod{m}$$

whereby P is a polynomial.

Remark: If $m = m_1 \dots m_n$, where the m_i are pairwise relatively prime, then

$$P(x) = 0 \pmod{m}$$

has exactly one solution when for each i $P(x) = 0 \pmod{m_i}$ has a solution. For let x_i be a solution of this equation and choose x so that $x = x_i \pmod{m_i}$

for each i. x is then a solution of the original lequation. This argument also shows that

$$|\{x \in \mathbf{Z}_m : P(x) = 0\}| = \prod |\{x \in \mathbf{Z}_{m_i} : P(x) = 0\}|.$$

Hence we can confine our attention to the case $m = p^\alpha$.

We start with $n = p$, a prime number. Then we have

Proposition 12 *(The theorem of Lagrange) Suppose that P his a non-trivial polynomial of degree n. Then the equation*

$$P(x) = 0 \quad (mod\, p)$$

has at most n distinct solutions (mod p).

Proof. In this case $\mathbf{Z}_p$ is a field and the classical proof (for $\mathbf{R}$ or $\mathbf{C}$) of the corresponding theorem can be carried over.

■

As we saw above, the equation

$$x^2 = 1 \quad (\text{mod } 8),$$

shows that this result can fail in the casae where $\mathbf{Z}_m$ is not a field.

Applications of the theorem of Lagrange: We can reformulate the Proposition as follows: Suppose that P is a polynomial of degree n which has more than n solutions (mod p). Then p is a divisor of all coefficients of P.

As an example consider the polynomial

$$P(x) = (x-1)\dots(x-p+1) - (x^{p-1} - 1) = G(x) - H(x).$$

Both G and H have $p-1$ roots—$1, 2, \dots, p-1$. P is thus a polynomial of degree $p-2$ with $p-1$ roots. Hence p is a divisor op each coefficient of q. This provides a second proof of

Proposition 13 *(Wilson's theorem)* $(p-1)! = -1$ *(mod p).*

(It suffices to consider the constant coefficient).

Proposition 14 *Wostenholme's theorem For $p \geq 5$ we have*

$$\sum_{k=1}^{p-1} \frac{(p-1)!}{k} = 0 \qquad (mod\, p^2).$$

Proof. Exercise.

■

In order to solve the equation $P(x) = 0$ (mod p^α), we begin with the case $\alpha = 1$. We solve this equation by trial and error and then use the following method to obtain from solutions of $P(x) = 0$ (mod p^α) solutions for $P(x) = 0$ (mod $p^{\alpha+1}$).

Proposition 15 *Let x be a solution of the equation $P(x) = 0$ (mod (p^α)), $(0 \leq x < p^\alpha)$. Es gilt: Then with le$\overline{x} < p^{\alpha+1}$ and $\overline{x} = x$ (mod$(p^{\alpha-1})$), so that $q(\overline{x}) = 0$ (mod $(p^{\alpha+1})$).*

b) If $P'(x) = 0$ (mod p), then one of two things can occur. Either b1) $P(x) = 0$ (mod $(p^{\alpha+1})$). Then there are r the equation $P(\overline{x}) = 0$ (mod $p^{\alpha+1}$) mit $\overline{x} = x$ (mod p^α).

or

b2) $P(x) \neq 0$ mod $(p^{\alpha+1})$. Then $P(\overline{x}) = 0$ (mod $p^{\alpha+1}$) has no solutions $\overline{x}$ with $\overline{x} = x$ (mod p^α).

Proof. a) We substitute $\overline{x} = sp^\alpha + x$ in the polynomial and get

$$P(x+h) = P(x) + hP'(x) + \dots + \frac{P^{(n)}(x)}{n!}h^n$$

where the coefficient $\frac{P^{(k)}(x)}{k!}$ of h^k is a polynomial with whole-number coefficients. For $\overline{x} = sp^\alpha + x$ we have $P(\overline{x}) = P(x) + P'(x)sp^\alpha +$ terms with $p^{\alpha+1}$ as factor, i.e.

$$P(\overline{x}) = P(x) + P'(x)sp^\alpha \pmod{p^{\alpha+1}}.$$

Since $P(x) = 0$ (*mod* p^α) we have $P(x) = kp^\alpha$ $(k \in \mathbf{Z})$. Hence

$$\begin{aligned} P(\overline{x}) &= kp^\alpha + P'(x)sp^\alpha \pmod{p^{\alpha+1}} &(37)\\ &= p^\alpha(k + P'(x)s) \pmod{p^{\alpha+1}} &(38)\end{aligned}$$

Hence it suffices to take s to be a solution of the equation

$$P'(x)s + k = 0 \pmod{p}.$$

(The uniqueness of $\overline{x}$ follows from that of the solution of this equation).
Case b): This uses a similar argument.

■

1.5 Arithmetical functions:

In this section we study so-called **arithmetical functions**. These are functions from $\mathbf{N}$ (sometimes also $\mathbf{N}_0$ or $\mathbf{Z}$) into $\mathbf{C}$ (usually the values are also taken in $\mathbf{Z}$).

Example The following functions on $\mathbf{N}$ are $\mathbf{Z}$-valued arithmetical functions.

Table:

Such a function f is called **multiplicative**, if we have $f(m.n) = f(m)f(n)$ $(m, n \in \mathbf{N}, \gcd(m, n) = 1)$ resp. **strongly multiplicative**, if $f(m.n) = f(m)f(n)$ $(m, n \in \mathbf{N})$.

For example $f(n) = n$ resp. $f(n) = n^\alpha$ are strongly multiplicative. μ and ϕ are multiplicative, but not strongly multiplicative (for μ the multiplicativity is clear, for ϕ see below).

If f is multiplicative, then $f(m) = \prod f(p_i^{\alpha_i})$, with $m = \prod p_i^{\alpha_i}$.

If f is strongly multiplicative, then we even have $f(m) = \prod(f(p_i))^{\alpha_i}$.

The **sum funtion** S_f of an arithmetic function f is defined as follows:

$$S_f(n) = \sum_{d|n} f(\frac{n}{d}).$$

For example:

$$S_\phi(12) = \phi(12) + \phi(6) + \phi(4) + \phi(3) + \phi(2) + \phi(1) = 4 + 2 + 2 + 2 + 1 + 1 = 12 \tag{39}$$

$$S_\mu(12) = 0 + 1 + 0 - 1 - 1 + 1 = 0 \tag{40}$$

(in fact:

$$S_\phi(n) = n \quad (n \in \mathbf{N}) \tag{41}$$

$$S_\mu(n) = \delta(n) \quad (n \in \mathbf{N}) \tag{42}$$

as will be shown later).

Further examples of arithmetic functions:

$\tau(n) = \sum_{d|n} 1 = S_1(n)$—the number of divisors of n;

$\sigma(n) = \sum_{d|n} d = S_{Id}(n)$—the sum of the divisors;
$\sigma_k(n) = \sum_{d|n} d^k = S_{Id^k}(n)$ — the sum of the k-th powers of the divisors.

Lemma 2 *If f is multiplicative, then so is S_f.*

PROOF. We begin with the remark that each divisor d of a product mn of two relatively prime numbers has a unique factorisation $d_1 d_2$ where $d_1|m$, $d_2|n$.

For example, the divisors

$$\text{of } 12\text{: } 1, 2, 3, 4, 6, 12, \tag{43}$$

$$\text{of } 25\text{: } 1, 5, 25, \tag{44}$$

$$\text{of} 12 \times 25 = 300 : 1, 2, 3, 4, 6, 12, 5, 10, 15, 20, 30, 60, 25, 50, 75, 100, 150, 300. \tag{45}$$

Hence

$$S_f(mn) = \sum_{d|mn} f(d) = \sum_{d_1,d_2,d_1|m,d_2|n} f(d_1)f(d_2) \tag{46}$$

$$= \sum_{d_1|m} f(d_1) \sum_{d_2|n} f(d_2) \tag{47}$$

$$= S_f(m)S_f(n). \tag{48}$$

■

Example We calculate the values of τ and σ_k. Since these functions are multiplicative, it suffices to calculate their values for powers of primes. Since the divisors of p^α are $\{1, p, p^2, \ldots, p^\alpha\}$, we have

$$\phi(p^\alpha) = p^\alpha - p^{\alpha-1} = p^\alpha(1 - \frac{1}{p}) \tag{49}$$

$$\tau(p^\alpha) = (\alpha + 1) \tag{50}$$

$$\sigma_k(p^\alpha) = \sum_{i=0}^{\alpha} p^{ik} = \frac{p^{k(\alpha+1)} - 1}{p - 1} \quad (k \neq 0) \tag{51}$$

$$\tag{52}$$

Hence for $n = p_1^{\alpha_1} \ldots p_r^{\alpha_r}$

$$\sigma_k(n) = \prod_i \frac{p_i^{k(\alpha_i+1)} - 1}{p_i^k - 1} \tag{53}$$

$$\tau(n) = \prod_i (\alpha_i + 1) \tag{54}$$

$$\phi(n) = \prod p_i^{\alpha_i} (\frac{1}{1 - p_i}) = n \prod_i (1 - \frac{1}{p_i}) \tag{55}$$

(We have tacitly used the fact that ϕ is multiplicative. This will be proved below).

We now calculate the sum function S_μ of the μ-function as follows: since μ is multiplicative, the same is true for S_μ. For a power p^α of a prime, we have:

$$\begin{aligned} S_\mu(p^\alpha) &= \mu(1) + \mu(p) + \mu(p^2) + \dots && (56) \\ &= 1 - 1 + 0 \dots && (57) \\ &= 0 \text{ falls } \alpha > 0. && (58) \end{aligned}$$

This means that $S_\mu(n) = \delta(n)$ whenever n is a power of a p rime and hence for every n (since both functions are multiplicative).

The formula $S_\mu = \delta$ then follows from the following considerations:

If f is multiplicative, then

$$\sum_{d|n} f(d) = (1+f(p_1)+\dots+f(p_1^{\alpha_1}))(1+f(p_2)+\dots+f(p_2^{\alpha_2}))\dots(1+\dots+f(p_r^{\alpha_r}))$$

where $n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$. For example, when $f(n) = n^s$ we have:

$$\sum_{d|n} d^s = (1 + p_1^s + \dots + p_1^{\alpha_1 s}) \dots (1 + p_r + \dots + p_r^{\alpha_r s}) \tag{59}$$

$$(= \prod_i \frac{p_i^{\alpha_i} - 1}{p_i - 1} \text{ for } s = 1). \tag{60}$$

Corollar 2 $\sum_{d|n} \mu(d) f(d) = (1 - f(p_1) \dots (1 - f(p_r)).$

If we take $f = 1$, we get: resp.

Proposition 16 *Let f be an arithmetic function and put $g = S_f$. Then*

$$f(n) = \sum_{d|n} \mu(d) g(\frac{n}{d}).$$

Proof. The right hand side is

$$\sum_{d|n} \mu(d) \sum_{d'|\frac{n}{d}} f(d').$$

But the set of pairs d, d' so that $d|n$ resp. $d'|\frac{n}{d}$ are exactly the pairs d, d' with $dd'|n$. Hence

$$\sum_{d|n} \mu(d) \sum_{d'|\frac{n}{d}} f(d') = \sum_{d,d;dd'|n} \mu(d) f(d')$$

By symmetry, this is also the sum

$$\sum_{d'|n} f(d') \sum_{d|\frac{n}{d'}} \mu(d) = \sum_{d'|n} f(d') S_\mu(\frac{n}{d'}) = f(n)$$

(since $S_\mu(\frac{n}{d'}) = 0$, except when $\frac{n}{d'} = 1$, d.h. $d' = n$.)

■

A similar argument shows that a function g with the property

$$f(n) = \sum_{d|n} \mu(d) g(\frac{n}{d})$$

must be S_f.

Example Since $S_\phi = id$, we have:

$$\phi(n) = n \sum_{d|n} \frac{1}{d} \mu(d).$$

Example n is called **perfect**, if it is equal to the sum of its proper divisors i.e. $\sigma(n) = 2n$. For example, the numbers

$$6 = 1 + 2 + 3, \quad 28 = 1 + 2 + 4 + 7 + 14$$

are perfact. More generally, if $n = 2^p - 1$ is a Mersenne prime, then

$$m = 2^{p-1}(2^p - 1)$$

is perfect. For

$$
\begin{aligned}
\sigma(m) &= \sigma[(2^{p-1})(2^p-1)] && (61)\\
&= \sigma(2^{p-1})\sigma(2^p-1) && (62)\\
&= (2^p-1)(2^p-1+1) && (63)\\
&= 2.2^{p-1}(2^p-1) = 2m && (64)
\end{aligned}
$$

In fact, every even prime number is of this form. (for example $6 = 2.3$, $28 = 2^2.7$). (We remark that it is not known whether odd perfect numbers exist).

Proof. (that every even perfect number n is of the form $2^{p-1}(2^p-1)$, with 2^p-1 a Mersenne prime.) If $\sigma(n) = 2n$ where $n = 2^k m$, with m odd, then

$$\sigma(2^k m) = \sigma(2^k)\sigma(m) = (2^{k+1}-1)\sigma(m).$$

Since $\sigma(n) = 2n$, we have

$$2^{k+1}m = (2^{k+1}-1)\sigma(m).$$

Hence 2^{k+1} is a divisor of $\sigma(m)$. Write $\sigma(m) = 2^{k+1}\ell$ so that $m = (2^{k+1}-1)\ell$. If $\ell > 1$, then

$$\sigma(m) \geq \ell + m + 1$$

(since $1, \ell, m$ are divisors of m). But $\ell + m = 2^{k+1}\ell = \sigma(m)$ which is a contradiction. Hence $\ell = 1$, i.e. $\sigma(m) = 2^{k+1}$ and $m = 2^{k+1}-1$. Thus $\sigma(m) = m+1$, and so $m = (2^{k+1}-1)$ which is a prime number and so the Mersenne prime (2^k-1) ($p = k+1$ prime). Hence

$$n = 2^k m = 2^{p-1}(2^p-1) \qquad \square$$

■

The structure of $\mathbf{Z}_m^*$: Let $g \neq e$ be an element of a group G and consider its powers $1, g, g^2, \ldots$. There are two possibilities:

a) No power of g^r is equal to e. Then g generates the infinite cyclic group $(\mathbf{Z}, +)$. In this case we write $ord(g) = \infty$.

b) There is a $r > 0$ so that $g^r = e$. The smallest such r is called the **order** of g—written ord r. Then g generates a subgroup $Z(g)$ which is isomorphic to $\mathbf{Z}_r$. It follows from the structure of the latter that:

$$g^s = e \Leftrightarrow r|s \quad (s \in \mathbf{Z})$$

$Z(g)$ has $\phi(r)$ generators, namely those elements $g^{s_1}, \ldots, g^{s_{\phi(r)}}$, where $s_1, \ldots, s_{\phi(r)}$ is a listing of $\mathbf{Z}_m^*$ (one calls it a **reduced residue class** for $\mathbf{Z}_r$).

Of course, if the group is finite, the first case cannot occur and the order of g as well as b eing finite is a divisor of $|G|$. (This provides a new proof for the theorem of Fermat: $a^{\phi(m)} = 1$ (modm), if $m \in \mathbf{N}$, $m > 1$, and $a \in \mathbf{Z}$ with $\gcd(a, m) = 1$).

Remark 1) If ord $g = r = k\ell$ where $k, \ell > 0$, then $\text{ord}(g^k) = \ell$.
2) If ord g_1 and ord g_2 are relatively prime, and the group is commutative, then ord $g_1g_2 =$ (ord g_1)(ord g_2).
Proof. Put $r =$ ord g_1, $s =$ ord g_2. We have

$$(g_1g_2)^{rs} = g_1^{rs}g_2^{rs} = e.$$

Hence $t|rs$ where $t =$ ord g_1g_2. Since $(g_1g_2)^{rt} = e = g_1^{rt}$, then $s|rt$, and so $s|t$. Similarlyl, $r|t$ and so $rs|t$. This imlpies $t = rs$.)

■

We shall now show that the arithmetical function ϕ introduced above is multiplicative.

Proposition 17 *Let $\{r_1, \ldots, r_{\phi(m)}\}$ be a reduced system of residues (mod m), resp. $\{s_1, \ldots, s_{\phi(n)}\}$ a reduced system of residues (mod n), whereby m and n are relatively prime. Then the numbers*

$$\{nr_i + ms_j : i \in \{1, \ldots \phi(m)\},\ j \in \{1, \ldots, \phi(n)\}\}$$

are a reduced system (mod mn).

Corollar 3 *$\phi(mn) = \phi(m)\phi(n)$ provided that* $\gcd(m, n) = 1$.

We bring an alternative proof of the following result:

$$\sum_{d|n} \phi(d) = n \quad (n \in \mathbf{N})$$

Let $D(d)$ denote the number of y from $\{1, \ldots, n\}$, with $\gcd(n, y) = d$. Since every number is associated to a $D(d)$,

$$n = \sum_{d|n} |D(d)|.$$

But $|D(d)| = \phi(\frac{n}{d})$. For $x \in D(d) \Leftrightarrow x = qd$ where $1 \le q \le \frac{n}{d}$ and $\gcd(\frac{n}{d}, q) = 1$. There are exactly $\phi(\frac{n}{d})$ such q's.

We now consider the following problem: When is the group $\mathbf{Z}_n^*$ cyclic? This is equivalent to the question: Does there exist $g \in \mathbf{Z}_n^*$ with ord $g = \phi(n)$?

Remark Such an element is called a **primite root** (mod m). There are then exactly $\phi(\phi(n))$ primitive roots. (For cyclic group $\mathbf{Z}(r)$ has $\phi(r)$ generators).

Our main result is as folllows:

Proposition 18 *$\mathbf{Z}_m^*$ is cyclic if and only if $m = p^\alpha$ (p an odd prime), or $m = 2p^\alpha$ (p an even prime), or $m = 2$, or $m = 4$.*

Proof. We begin with the case where m is a power of 2.

$m = 2$: then $g = 1$ is a primitive root.

$m = 4$: then $g = 1, 3$ are primitive roots.

$m = 2^\alpha (\alpha > 2)$: In this case, there are no primitive roots. For

$$a^{2^{\alpha-2}} = 1 \pmod{2^\alpha} \quad (a \text{ odd })$$

This is proved by induction. We prove the case $\alpha = 3$. Let $a = (2k + 1)$. Then

$$a^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 4k(k + 1) + 1 = 1 \pmod 8.$$

The case where $\alpha > 3$ is similar.

We now show that if m is not on the above list, then it faisl to have a primitive root. It is clear that if m is not a power of 2, then it has a factorisation $m = m_1 m_2$ where $\gcd(m_1, m_2) = 1$ and $\phi(m_1)$ resp. $\phi(m_2)$ is even. Then if a and m are relatively prime,

$$a^{\frac{1}{2}\phi(m)} = (a^{\phi(m_1)})^{\frac{1}{2}\phi(m_i)} = 1 \text{ (mod } m_1)$$

resp.

$$a^{\frac{1}{2}\phi(m)} = 1 \text{ (mod } m_2).$$

Hence $a^{\frac{1}{2}\phi(m)} = 1$ (mod m) (by the uniqueness part of the Chinese remainder theorem), and so ord $a \le \frac{1}{2}\phi(m)$.

We now consider the case where m is an odd prime p.

■

Proposition 19 *The group $\mathbf{Z}_p$ is cyclic.*

Proof. For every $d|p-1$ let $\psi(d)$ be the number of elements of $\{1, 2, \ldots, p-1\}$, which have order d. We show $\bigwedge_{d|p-1} \psi(d) \neq 0$, in particular $\psi(p-1) \neq 0$. We have

a) $0 \le \psi(d) \le \phi(d)$ - (even: $\psi(d) = 0$ or $\psi(d) = \phi(d)$) (since the cyclic group of order d has $\phi(d)$ generators).

b) $\sum_{d|p-1} \psi(d) = p - 1$ $(= \sum_{d|p-1} \phi(d))$. For every element of $\{1, \ldots, p-1\}$ must have some order.

These two facts immediately imply that $\phi(d) = \psi(d)$.

■

The same proof demonstrates the following fact:

Proposition 20 *Let G be a group with $|G| = m$, so that $d|m \Rightarrow G$ has at most d elements with $x^d = e$. Then G is cyclic.*

Corollar 4 *Let K be a finite field. Then K^* is cyclic.*

For the equation $x^d - e = 0$ (which is of degree d) has at most d solutions.

Lemma 3 *Let $n = p^\alpha$. then n has a primite root.*

Proof. Let g be a primitive root (mod p). We show that there exists x, so that $h = g + px$ is a primitive root (mod p^α).

We know that $g^{p-1} = 1 + py$ for some $y \in \mathbf{Z}$. Hence

$$\begin{aligned} h^{p-1} &= (g+px)^{p-1} && (65)\\ &= g^{p-1} + p(p-1)xg^{p-2} + p^2\binom{p-1}{2}x^2g^{p-3} + \dots && (66)\\ &= 1 + py + p(p-1)xg^{p-2} + p^2\binom{p-1}{2}x^2pg^{p-3} + \dots && (67)\\ &= 1 + pz && (68)\end{aligned}$$

where $z = y + (p-1)xg^{p-2}$ (mod p). The coefficient of x is relatively prime to p and so we can choose x so that $z = 1$ (mod p). We claim that for this choice h is a primitive root. i.e. $\text{ord}(h) = \phi(p^\alpha) = p^{\alpha-1}(p-1)$. For put $d = \text{ord } h$, so that $d \mid p^{\alpha-1}(p-1)$. Since h is a primitive root (mod$p$), then $p - 1 \mid d$, and so $d = p^k(p-1)$, where $k < \alpha$. p is odd and so

$$(1+pz)^{p^k} = 1 + p^{k+1}\bar{z}$$

with $\bar{z}$ and p relatively prime. Hence

$$1 = h^d = h^{p^k(p-1)} = (1+pz)^{p^k} = 1 + p^{k+1}\bar{z}(\text{mod}p^\alpha).$$

This implies that $\alpha = k + 1$, as was to be proved.

■

Corollar 5 *$m = 2p^\alpha$ has a primitive root.*

For $\phi(2p^\alpha) = \phi(2)\phi(p^\alpha) = \phi(p^\alpha)$. Let g be a primitive root (mod p^α). One of the two, g or $g + p^\alpha$, is odd and so is an element of $\mathbf{Z}_m^*$ and thus a primitive root.

Higher order equations: Now let m be so that $\mathbf{Z}_m^*$ is cyclic and let g be a primitive root. Then we have that for each $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$ there exists a number k (which is uniquely determined mod $\phi(m)$) with $a = g^k \pmod{m}$. We write $k = \mathrm{ind}_m(a)$. Then

$$\mathrm{ind}_m(ab) = \mathrm{ind}_m a + \mathrm{ind}_m b \pmod{\phi(m)} \tag{69}$$

$$\mathrm{ind}_m(a^n) = n\ \mathrm{ind}_m a \pmod{\phi(m)} \tag{70}$$

With the help of this index function (which can be regarded as a sort of distcete logarithm), we can solve equations of higher order.

Example Solve the equation $x^5 = 2 \pmod{7}$.

3 is a primitive root (mod 7) and $2 = 3^2$, i.e. $\mathrm{ind}_7 2 = 2$. Then we have $5\ \mathrm{ind}_7(x) = 2 \pmod{6}$. This has solultion $\mathrm{ind}_7 x = 4$. Hence the solution of the original equation is

$$x = 3^4 = 4 \pmod{7}$$

1.6 Quadratic residues, quadratic reciprocity

We now consider the general quadratic equation

$$ax^2 + bx + c = 0 \pmod{n}$$

We first reduce to the special case

$$x^2 = r \pmod{n}$$

This is done by putting $d = b^2 - 4ac$, $y = 2ax + b$. Then a simple computation shows

$$y^2 = d \pmod{4n}$$

Thus in order to solve the original equation $ax^2 + bx + c = 0$ $(mod\ n)$ it suffices to solve the equation

$$y^2 = d \pmod{4an} \text{ bzw. } y = 2ax + b \pmod{n}$$

Definition Let $n \in \mathbf{N}$ and $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$. If the equation

$$x^2 = a \pmod{n}$$

is solvable, we say that a is a**quadratic residue** (mod n) herwise it is a **quadratic non-residue**.

The Legendre Symbol: If p is a prime number, and $a \in \mathbf{Z}$ is relatively prime to p then we put (of course, $a = a'$ (mod p) $\Rightarrow (\frac{a}{p}) = (\frac{a'}{p})$)

Proposition 21 $(\frac{a}{p}) = a^{\frac{1}{2}(p-1)}$ *(mod p).*

Proof. Since $a^{p-1} = 1$ (mod p), the right hand side is either $+1$ or -1. Let g be a primitive root (mod p). It is clear that the quadratic residues are the numbers $\{1, g^2, g^4, \dots, g^{2r}\}$ and the non-residues are $\{g, g^3, \dots, g^{2r+1}\}$ $(r = \frac{p-1}{2})$. It is equally clear that $a^r = 1$ for the elements from the first list resp. -1 for thos of the second.

■

Corollar 6 *The function* $a \mapsto (\frac{a}{p})$ *is strongly multiplicative. i.e.* $(\frac{ab}{b}) = (\frac{a}{p})(\frac{b}{p})$ $(a, b \in \mathbf{N})$.

Corollar 7 $(\frac{-1}{p}) = (-1)^{\frac{1}{2}}(p-1)$ *i.e.* (-1) *is a quadratic residue* $(mod\, p) \Leftrightarrow p = 1$ $(mod\ 4)$.

Our main result is the law of quadratic reciprocity a.

Proposition 22 *Let p, q be distinct prime numbers. Then*

$$(\frac{p}{q})(\frac{q}{p}) = (-1)^{\frac{1}{4}(p-1)(q-1)}$$

i.e.

$$(\frac{p}{q}) = (\frac{q}{p}) \textit{ if } p \neq 3 \textit{ and } q \neq 3 \ (mod\ 4)$$

resp.

$$(\frac{p}{q}) = -(\frac{q}{p}) \textit{ otherwise.}$$

Using this result one can calculate $(\frac{a}{p})$ in concrete cases with a routine manipulation as illustrated in the following example.

Example We calculate $(\frac{15}{71})$

$$(\frac{15}{71}) = (\frac{3}{71})(\frac{5}{71}) = -(\frac{71}{3})(\frac{71}{5}) = -(\frac{2}{3})(\frac{1}{5}) = 1.$$

Example We calculate $(\frac{-3}{p})$ $(p \geq 5)$.

$$(\frac{-3}{p}) = (\frac{-1}{p})(\frac{3}{p}) = (-1)^{\frac{1}{2}(p-1)}(\frac{3}{p}) = (\frac{p}{3}).$$

Hence -3 is a quadratic residue (mod p) $\Leftrightarrow p = 1$ (mod 6).

In order to prove the main result we use the following notion: If $a \in \mathbf{Z}$, the **absolutely smallest remainder** of a (mod n) is that a' from the interval $-\frac{n}{2} < a' \leq \frac{n}{2}$, with $a = a'$ (mod n).

Example (mod 13)

1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	3	4	5	6	-6	-5	-4	-3	-2	-1	0

If $\gcd(a,p) = 1$, we put a_j equal to the absolutely smallest remainder of $a_j (\text{mod } p)$ $(j = 1, 2, \dots, r = \frac{p-1}{2})$. Then $(\frac{a}{p}) = (-1)^{\ell}$, where ℓ is the number of a_j $a_j < 0$. For example, one can calculate $(\frac{2}{13})$ as follows:

a_j	2	4	6	8	1 0	12
a_j	2	4	6	-5	-3	-1

Since $\ell = 3$, we have $(\frac{2}{13}) = -1$.

Proof. Proof of the formula $(\frac{a}{p}) = (-1)^{\ell}$ We claim that $\{|a_j| : 1 \leq j \leq r\}$ is a permutation of $\{1, \dots, r\}$. For $a_i \neq a_j$ (since $ai = aj$) (mod$p$) $\Rightarrow i = j$ (mod$p$) $\Rightarrow i = j$ resp. $a_j \neq -a_j$ (da $a_i + a_j = 0 \Rightarrow a(i+j) = 0$ (mod$p$) which is impossible. since $0 < i + j < p$).

Thus we have

$$a_1 a_2 \dots a_r = (-1)^{\ell} r!$$

Further $a_1 \dots a_r = r! a^r$ (mod p). Hence $r! a^r = (-1)^{\ell} r!$ (mod p) or $a^r = (-1)^{\ell}$ (mod p).

■

Corollar 8 *For every odd prime we have*

$$(\frac{2}{p}) = (-1)^{\frac{1}{8}(p^2-1)}$$

i.e.. 2 is a quadratic residue (mod p) $\Leftrightarrow p = \pm 1$ mod 8.

Proof. Proof of the law of quadratic recprocity

$$(\frac{p}{q}) = (-1)^{\ell}$$

where ℓ is the number of points (x, y) in $\mathbf{Z} \times \mathbf{Z}$ with $0 < x < \frac{1}{2} q$ and $-\frac{1}{2} q < px - qy < 0$. Hence $0 < x < \frac{p}{2}$. We decompose the lattice points (x, y) of the rectangle $0 < x < \frac{q}{2}$, $0 < y < \frac{p}{2}$ into four regions I,II,III,IV, whereby I $= \{(x,y) : px - qy \leq -\frac{p}{2}\}$, II $= \{(x,y) : -\frac{p}{2} < qy - px < 0\}$, III $= \{(x,y) : -\frac{q}{2} < px - qy < 0\}$, IV $= \{(x,y) : px - qy \leq -\frac{p}{2}\}$.

The total number of lattice points is $\frac{1}{4}(p-1)(q-1)$ and so is even, when $p \neq 3$ and $q \neq 3$ (mod 4), odd otherwise.

We have

a) $(\frac{p}{q}) = (-1)^{\ell}$ where ℓ is the number of lalttice points in I;

b) $(\frac{q}{p}) = (-1)^{m}$ where m is the number of lattice points in II;

c) the number of lattice points in II = the number of lattice point in IV (since II can be mapped onto III by means of an affine mapping which

leaves the number of lattice points invariant). This implies the result.

The Legendre symbol can be generalised as follows: Put $n = p_1 \dots p_k$, (where the prime numbers p_i are not necessarily distinct) and $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$. We define the **Jacobi symbol** $(\frac{a}{n})$ as follows:

$$(\frac{a}{n}) = \prod_i (\frac{a}{p_i}).$$

Then if $(\frac{a}{n}) = -1$, a is a quadratic non-residue. with respect to n However, it is not true in general that $(\frac{a}{n}) = 1 \Rightarrow a$ is a quadratic residue with respect to n).

Example

$$(\frac{6}{35}) = (\frac{6}{5})(\frac{6}{7}) = (\frac{1}{5})(\frac{-1}{7}) = -1.$$

The calculation of the Legendre symbol can be simplified by employing the Jacobi symbol as the following example illustrates:

Example We calculate $(\frac{335}{2999})$ (Note that 2999 is a prime number). We have

$$\begin{array}{cccccc} (\frac{335}{2999}) = & -(\frac{2999}{335}) = & -(\frac{-16}{335}) = & -(\frac{-1}{335}) & (\frac{16}{335}) = & -(\frac{-1}{335}) = \;\; 1 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\ \text{Legendre-S.} & \text{Jacobi-S.} & \text{Jacobi-S.} & \text{Jacobi-S.} & 1 & \end{array}$$

The Jacobi symbol has the following properties:

$$(\frac{ab}{n}) = (\frac{a}{n})(\frac{b}{n}) \tag{71}$$

$$(\frac{-1}{n}) = (-1)^{\frac{1}{2}(n-1)} \tag{72}$$

$$(\frac{2}{n}) = (-1)^{\frac{1}{8}(n^2-1)} \tag{73}$$

$$(\frac{m}{n})(\frac{n}{m}) = (-1)^{\frac{1}{4}(m-1)(n-1)} \text{ (if } n, m \text{ are odd numbers with } \gcd(m, n) = 1). \tag{74}$$

We have thus established a complete theory for the equation

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

(where $\gcd(a, p) = 1$).

In order to deal with the general equation

$$x^2 = a \pmod{n}$$

it suffices (by the Chinese remainder theorem) to consider the cases

$$x^2 = a \pmod{p^\alpha}$$

i.e. to calculate the zeros of the polynomial

$$P(x) = x^2 - a.$$

We distinguish between two cases:
a) p is odd. Since $P'(x) = 2x$, then $P'(x) \neq 0$, if $\gcd(x,p) = 1$. Hence $x^2 = a \pmod{p^\alpha}$ is solvable if and only if a is a quadratic residue $\pmod{p}$ ist. Each solution of the equation $x^2 = a \pmod{p}$ provides exactly one solution for the equation $x^2 = a \pmod{p^\alpha}$ by the method developed above.
b) $p = 2$. Here we have

1.7 Quadratic forms, sums of squares

We now consider the problem of representing natural numbers as the sum of squares. We will found that it is useful to consider this as a special case of a more general problem, that of calculating the range of a quadratic form.

Quadratic Forms : These are mappings of the form

$$\begin{aligned} f(x,y) &= ax^2 + bxy + cy^2 && (75)\\ &= \frac{1}{2}X^t AX && (76) \end{aligned}$$

where

$$A = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \end{bmatrix}$$

A is called the matrix of the form. The discrimant of f is the number $d = b^2 - 4ac = -det\ A$. Then

$$\begin{aligned} d &= 0 \pmod 4 \text{ if } b \text{ is even} && (77)\\ d &= 1 \pmod 4 \text{ if } b \text{ is odd} && (78) \end{aligned}$$

Conversely, if $d = 1$ or $0 \pmod 4$, then d is the discrimant of a form, in fact the form mit matrix

$$a\begin{bmatrix} 2 & 0 \\ 0 & -\frac{d}{2} \end{bmatrix}$$

if $d = 0 \pmod 4$ and

$$\begin{bmatrix} 2 & 1 \\ 1 & \frac{1}{2}(1-d) \end{bmatrix}$$

if $d = 1 \pmod 4$.

If

$$P = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

is a 2×2 matrix over $\mathbf{Z}$ with det $P = 1$ (so that P^{-1} is also a matrix over $\mathbf{Z}$),

$$\begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x' \\ y' \end{bmatrix}$$

transforms f in x und y into one with matrix $A' = P^tAP$, i.e.

$$A' = \begin{bmatrix} 2a' & b' \\ b' & 2c' \end{bmatrix}$$

where

$$\begin{aligned} a' &= f(p, r) && (79) \\ b' &= 2apq + b(ps + qr) + 2crs && (80) \\ c' &= f(q, s) && (81) \end{aligned}$$

Two forms f and g are **equivalent** (written $f \sim g$), if we can transform one into the other by means of such a change of variables. If $f \sim g$, then f and g have the same range.

From now on we shall only deal with positive definite forms. A necessary and sufficient condition for this is that

$$a > 0,\ 4ac - b^2 > 0 \quad \text{(and hence also } c > 0\text{)}.$$

For we can write the form f as

$$4af(x, y) = (2ax + by)^2 - dy^2.$$

Using substitutions as above, we can bring each form into a so-called **reduced** form i.e. one with

$$\begin{aligned} -a < b \le a < c \text{ or} && (82) \\ 0 \le b \le a = c. && (83) \end{aligned}$$

Such forms satisfy the relation

$$-d = 4ac - b^2 \geq 3ac$$

and as a result

$$a \leq \frac{1}{3}|d|, \quad c \leq \frac{1}{3}|d|, \quad |b| \leq \frac{1}{3}|d|.$$

This implies that there are at most finitely many reduced forms with a given diskriminant d. We write $h(d)$ (the **class number** of d) for the number of reduced forms with diskriminant d.

Example For $d = -4$ we have $3ac \leq 4$ and so $a = c = 1$ and $b = 0$. Hence $h(-4) = 1$.

The importance of the function h lies in the fact that ir counts the number of equivalent forms. Indeed we have

Proposition 23 *If f, g are reduced and $d(f) = d(g)$, then $f \sim g$.*

PROOF. We shall show that two distinct reduced form f, g are not equivalent. We begin with the remark that $|x| \geq |y|$ implies:

$$f(x,y) \geq |x|(a|x| - |by|) + c|y|^2 \geq |x|^2(a - |b|) + c|y|^2 \geq a - |b| + c.$$

resp.

$$f(x,y) \geq a - |b| + c,$$

if $|y| \geq |x|$. Hence the three smallest values of f are a (for $(1,0)$), c (for $(0,1)$) and $a - |b| + c$ (for $(1,1)$ or $(1,-1)$). Hence $a = a'$, $c = c'$ and $b = \pm b'$, where a', b', c' are the coefficients of g.

We now show that $b = -b'$ implies $b = 0$. We can assume that $-a < b < a < c$. (For $-a < -b$ and $a = c$ implies that $b \geq 0$ and $-b \geq 0$, i.e. $b = 0$). In this case we have

$$f(x,y) \geq a - |b| + c > c > a.$$

If we transform f into g by means of the matrix P above, then $a = f(p,r)$. Hence $p = \pm 1$ and $r = 0$. The condition $ps - qr = 1$ then implies that $s = \pm 1$. Further $c = f(q,s)$ and so $q = 0$. This easily implies that $b = 0$.

■

Definition n is represented by f if $x, y \in \mathbf{Z}$ exists so that

$$\gcd(x,y) = 1, \quad f(x,y) = n.$$

Proposition 24 *n is representable by a form f with discrimant d $\Leftrightarrow$ the equation $x^2 = d(mod\, 4n)$ has a solution.*

Proof. $\Leftarrow$: Choose a solution x and put $b = x$. There is a c, sothat $b^2 - 4nc = d$. The form f with matrix

$$\begin{bmatrix} 2n & b \\ b & 2c \end{bmatrix}$$

has discriminant d and satisfies the condition $f(1,0) = n$.

$\Rightarrow$: Put $n = f(x,y)$ and choose r, s with $xr - ys = 1$. We then use the substition with

$$P = \begin{bmatrix} x & y \\ s & r \end{bmatrix}$$

. This transform f into a form f' with $a' = n$. f and f' have the sameie discriminant i.e. $b'^2 - 4nc' = d$. Hence $x = b'$ is a solution of the equation $x^2 = d \pmod{4n}$.

■

Using this result, we can solve the problem of the sum of squares.

Proposition 25 *A whole number n has a representation $n = x^2 + y^2$ ($x, y \in \mathbf{Z}$) $\Leftrightarrow$ for every $p = 3 \ (mod\ 4)$ $w(p)$ is even. ($w(p)$ is the index of p in the prime representation of n).*

Proof. $\Rightarrow$: If $n = x^2 + y^2$, then $x^2 = -y^2 \pmod{p}$ (whereby p is a prime divisor of n with $p = 3 \pmod{4}$). But (-1) is a quadratic non-residue (mod p) and hence so is $-y^2$. From this we can deduce that $p|x$ and so $p|y$, $p^2|n$. Now we have

$$(\frac{x}{p})^2 + (\frac{y}{p})^2 = \frac{n}{p^2}$$

and we repeat the argument for each prime divisor p' of n with $p' = 3 \pmod{4}$.

$\Leftarrow$: We remark firstly that products of sums of squares are also sums of squares. (For $(x^2 + y^2)(\overline{x}^2 + \overline{y}^2) = (x\overline{x} + y\overline{y})^2 + (x\overline{y} - y\overline{x})^2$.) It suffices to show that the theorem holds for $n = p$, where $p = 1 \pmod{4}$. But the form $x^2 + y^2$ is reduced with $d = -4$. Since $h(-4) = 1$, n is properly represntable $\Leftrightarrow$ the equation $x^2 = -4 \pmod{4p}$ is solvable. But there exists y with $y^2 = -1 \pmod{p}$ ankd so $(2y)^2 = -4 \pmod{4p}$.

■

Sums of four squares In this case, we have the result

Proposition 26 *Every number $n \in \mathbf{N}$ is the sum of four squares.*

Proof.

1) It suffices to consider the case of an odd prime. This follows as in the case of sums of two square from the identity

$$
\begin{aligned}
(x^2+y^2+z^2+w^2)(x'^2+y'^2+z'^2+w'^2) = \qquad (84)\\
= (xx'+yy'+zz'+ww')^2 + (xy'-yx'+wz'-zw')^2 \qquad (85)\\
+ (xz'-zx'+yw'-wy')^2 + (xw'-wx'+zy'-yz')^2 \qquad (86)
\end{aligned}
$$

Now let $n = p$ (an odd prime). The numbers

$$0, 1^2, 2^2, \ldots, (\frac{1}{2}(p-1))^2$$

are pairwise non-congruent (mod p). Hence there exist x, y with $0 < x, y \leq \frac{1}{2}(p-1)$, so that $x^2 = -1 - y^2$. We also have $x^2 + y^2 + 1 < p^2$. Thus there exists m with $0 < m < p$, so that $mp = x^2 + y^2 + 1$.

Let now ℓp be the smallest number with the property that $\ell\ p$ is a sum $x^2 + y^2 + x^2 + w^2$ of four squares. Then $\ell \leq m < p$.

Further, ℓ is odd. For otherwise either 0,. 2 or 4 of the numbers x, y, z, w would be odd. We could then rename x, y, z, w in such a way that $x+y, x-y, z+w, z-w$ are all even. Then

$$\frac{1}{2}\ell p = (\frac{1}{2}(x+y))^2 + (\frac{1}{2}(x-y))^2 + (\frac{1}{2}(z+w))^2 + (\frac{1}{2}(z-w))^2$$

and this is a contradiction.

We show that $\ell = 1$. Suppose that $\ell > 1$. We show that this leads to a contradiction. Let x', y', z', w' be the absolutely smallest residues of x, y, z, w (mod ℓ) and $n = x'^2 + y'^2 + z'^2 + w'^2$.

Then $n = 0 \pmod{\ell}$ resp. $n > 0$ (otherwise ℓ would be a divisor of x, y, z, w and hence of p). Since ℓ is odd, we have the inequality $n < 4(\frac{1}{2}\ell^2) = \ell^2$.

Hence $n = kl$, where $0 < k < l$. Since both kl and lp are sums of four squares, the same is true for $(kl)(lp)$. It follows form the representation that these squares are all divisible by ℓ^2. This implies that kp is a sum of four squares. This is a contradiction

■

Sums of three squares In this case we quaote the foolwoing theorem without proof:

Proposition 27 *A natural number n is the sum of three squares $\Leftrightarrow$ n is of the form*

$$4^j(8k+7) \quad (\textit{ wobei } j, k \in \mathbf{N}_0).$$

In this connection we mention Waring's conjecture (which was confirmed by Hilbert in 1909): For every $k \geq 2$ there exists a number s (which depends of course on k), so that every $n \in \mathbf{N}$ has a representation

$$n_1^k + \ldots + n_s^k$$

as the sum of s k-powers. (for example for $k = 2$ we can take $s = 4$).

1.8 Repeated fractions:

These are fractions of the form where $a_0 \in \mathbf{Z}$, $a_i \in \mathbf{N}$ $(i \geq 1)$, $c_i \in \mathbf{N}$ $(i \geq 1)$ and $c_i < a_i$. We write

$$a_0 + \frac{c_1|}{|a_1} + \frac{c_2|}{|a_2} + \ldots + \frac{c_n|}{|a_n}$$

for this expresiion. We can calculate its value algorithmically by defining finite sequences (p_k) and (q_k) recursively as follows:

$$p_k = a_k p_{k-1} + c_k p_{k-1} \qquad (87)$$

$$q_k = a_k q_{k-1} + c_k q_{k-1} \qquad (88)$$

(with initial valules : $p_{-2} = 0$, $q_{-2} = 1$, $p_{-1} = 1$, $q_{-1} = 0$, $c_0 = 1$). Then we have

$$\frac{p_k}{q_k} = a_0 + \frac{c_1|}{|a_1} + \ldots + \frac{c_k|}{|a_k}$$

In particular, $\frac{p_n}{q_n}$ is the value of the fraction.

($\frac{p_k}{q_k}$ is called k-**approximant** of the fraction).

Proof. We prove this by induction: The cases $k = 0$, $k = 1$ are trivial

$k \to k+1$. Put $a'_k = a_k + \frac{c_{k+1}}{a_{k+1}}$. Then

$$a_0 + \frac{c_1|}{|a_1} + \ldots + \frac{c_{k+1}|}{|a_{k+1}} = a_0 + \frac{c_1|}{|a_1} + \ldots + \frac{c_k|}{|a'_k} \qquad (89)$$

$$= \frac{p'_k}{q'_k} \qquad (90)$$

where

$$
\begin{aligned}
p'_k &= a'_k p_{k-1} + c_k p_{k-2} && (91)\\
q'_k &= a'_k q_{k-1} + c_k q_{k-2} && (92)\\
& && (93)
\end{aligned}
$$

(by the induction hypothesis) and hence

$$\frac{p'_k}{q'_k} = \frac{p_{k-1}(a_k + \frac{c_{k+1}}{a_{k+1}}) + c_k p_{k+2}}{q_{k-1}(a_k + \frac{c_{k+1}}{a_{k+1}}) + c_k q_{k+2}} = \frac{p_{k+1}}{q_{k+1}}$$

■

Proposition 28 $p_k q_{k-1} - q_k p_{k-1} = (-1)^k c_0 \ldots\ c_k$.

Proof. This is again an induction proof. The case $k = 0$ is trivial.

$k \to k+1$:

$$
\begin{aligned}
p_{k+1}q_k - q_{k+1}p_k &= (a_{k+1}q_k + c_{k+1}p_{k-1})q_k - (a_{k+1}q_k + c_{k+1}q_{k-1})p'_k && (94)\\
&= c_{k+1}(p_{k-1}q_k - q_{k-1}p_k) && (95)\\
&= c_{k+1}(-1)^k c_0 \ldots c'_k. && (96)
\end{aligned}
$$

■

We append here an alternative proof of this fact:

Put

$$P_k = \begin{bmatrix} p_{k+1} & p_k \\ q_{k+1} & q_k \end{bmatrix}$$

so that det $P_k = p_{k+1}q_k - p_k q_{k+1}$. Then

$$P_k = \begin{bmatrix} a_k & c_k \\ 1 & 0 \end{bmatrix} P_{k-1}$$

i.e. det $P_k = (-c_k)$ det P_{k-i}.

A continued fraction is called **regular**, if $c_i = 1$ for each i. It then has the form

$$a_0 + \frac{1|}{|a_1} + \ldots + \frac{1|}{|a_n} \text{ (geschr. } [a_0; a_1, \ldots, a_n].)$$

In this case we have

$$
\begin{aligned}
& p_k = a_k p_{k-1} + p_{k-2} && (97)\\
& q_k = a_k q_{k-1} + q_{k-2} && (98)\\
& p_k q_{k-1} - q_k p_{k-1} = (-1)^{k-1} && (99)\\
\text{Further } & p_k q_{k-2} - p_{k-2} q_k = (-1)^k a_k && (100)
\end{aligned}
$$

For

$$
\begin{aligned}
p_{k+1}q_{k-1} - p_{k-1}q_{k+1} &= (a_{k+1}p_k + p_{k-1})q_{k-1} - p_{k-1}(a_{k+1}q_k + q_{k+1}) && (101)\\
&= a_{k+1}(p_k q_{k-1} - p_{k-1}q_k) && (102)
\end{aligned}
$$

Thus $q_{k+1} \geq q_k$ and so $q_k > k$. Further

$$\frac{p_{2k}}{q_{2k}} < \frac{p_{2k+2}}{q_{2k+2}} \text{ resp. } \frac{p_{2k+1}}{q_{2k+1}} < \frac{p_{2k-1}}{q_{2k-1}}$$

and

$$\frac{p_k}{q_k} - \frac{p_{k-1}}{q_{k-1}} = \frac{(-1)^k}{q_k q_{k-2}}.$$

As we shall now show every rational number is representable as a continued fraction.

Proof. Let $r = \frac{a}{b} = a_0 + \frac{r_1}{b}$ with $0 \leq r_1 < b$. We determine a sequence $r_1, r_2, \ldots$ by means of the division allgorithm as follows

$$r_{i-1} = a_i r_i + r_{i+1},$$

(where $0 \leq r_{i+1} < r_i$.) There exists an n with $r_{n+1} = 0$. Hence $r = [a_0; a_1, \ldots, a_n]$.

■

Example We calculate the representation of $\frac{37}{49}$

$$
\begin{aligned}
37 &= 0.49 + 37 && (103)\\
49 &= 1.37 + 12 && (104)\\
31 &= 3.12 + 1 && (105)\\
12 &= 12.1 + 0 && (106)
\end{aligned}
$$

Hence $\frac{37}{49} = [0; 1, 3, 12]$.

Application We show how to use continued fractions to solve the folllowing equation:

$$ax + by = c$$

It suffices to consider the case where $\gcd(a, b) = 1$, $c = 1$. (Otherwise we can go over to the equation $\frac{a}{d}x + \frac{b}{d}y = \frac{c}{d}$, where $d = \gcd(a, b)$). Let $[a_0; a_1, \ldots, a_n]$ be the continued fraction representation of $\frac{a}{b}$ and $\frac{p_{n-1}}{q_{n-1}}$ the $(n-1)$-th approximant. The pair $(q_{n-1}(-1)^{n-1}, p_{n-1}(-1)^n)$ is a solution of the equation.

Example $37x + 13y = 3$. We first solve the equation $37x + 13y = 1$.

$$\frac{37}{3} = [12; 1, 5], \quad n = 3,\ p_{n-1} = 17,\ q_{n-1} = 6.$$

Hence a solution is $(6, -17)$. A solution of the original equation is thus $(18, -51)$. The set of solutions is then $\{(18 + 13t, -51 + 37t) : t \in \mathbf{Z}\}$.

Example The fact that

$$[2; 2, 3] = [2; 2, 2, 1] \quad (\text{i.e. } 2 + \frac{1}{2 + \frac{1}{3}} = 2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{1}}})$$

show that the above representation of a rational number as a continued fraction is not unique. In fact,

$$[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1] \text{ (if } a_n \geq 2) \qquad (107)$$

$$\text{resp. } [a_0; a_1, \ldots, a_{n-1}, 1] = [a_0; \ldots, a_{n-1} + 1] \qquad (108)$$

However, if

$$[a_0; a_1, \ldots, a_n] = [b_0; b_1, \ldots, b_m]$$

with $a_n > 1$, $b_m > 1$, then $m = n$ and $a_i = b_i$ for each i.

Infinite continued fractions: These are expressions of the form

$$a_0 + \frac{1|}{|a_1} + \frac{1|}{|a_2} + \ldots \quad ([a_0; a_1, \ldots])$$

where $(a_n)_{n=1}^{\infty}$ is a sequence from $\mathbf{N}_0$ (and $a_n > 0$ for $n > 1$). We define sequences (p_n) and (q_n) as in the finite case and remark firstly that $(\frac{p_n}{q_n})$ converges. If $\vartheta = \lim_{n\to\infty} \frac{p_n}{q_n}$, then we write

$$\vartheta = [a_0; a_1, \ldots].$$

The convergence follows from the fact that

a) $|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}| = \frac{1}{q_n q_{n+1}} \leq \frac{1}{n^2}$.

b) every $\frac{p_m}{q_m}$ $(m > n+1)$ lies in the interval with endpoints $\frac{p_n}{q_n}$ bzw. $\frac{p_{n+1}}{q_{n+1}}$.

Conversely, every real number x has a representation as an infinite continued fraction. This can be determined as follows: Put $x = [x] + \vartheta$ with $\vartheta \in]0, 1[$ and

$$\frac{1}{\vartheta} = a_1', \quad a_1 = [a_1'], \quad \vartheta_1 = a_1' - a_1 \tag{109}$$

$$\frac{1}{\vartheta_1} = a_2', \quad a_2 = [a_2'], \quad \vartheta_2 = a_2' - a_2 \text{ resp.} \tag{110}$$

We have

$$x = [a_0; a_1'] = [a_0; a_1 + \frac{1}{a_2'}] = [a_0; a_1, a_2'] \tag{111}$$

$$= [a_0; a_1, a_2, a_3'] \ldots \tag{112}$$

x lies between $\frac{p_n}{q_n}$ and $\frac{p_{n+1}}{q_{n+1}}$ and so we have

$$|x - \frac{p_n}{q_n}| \leq \frac{1}{q_n q_{n+1}} \leq \frac{1}{n^2}.$$

Further, for every n

$$x = \frac{a_{n+1}' p_n + p_{n-1}}{a_{n+1}' q_n + q_{n-1}}$$

(and so

$$x - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(a_n' q_n + q_{n-1})}.)$$

The rational numbers $\frac{p_n}{q_n}$ are called the **convergents** of x.

Example The contined fraction of π is

$$3 + \frac{1|}{|17} + \frac{1|}{|15} + \frac{1|}{|1} + \frac{1|}{|292} + \frac{1|}{|1} + \frac{1|}{|1} + \frac{1|}{|2} + \ldots$$

(with first convergents 3, $\frac{22}{7}$, $\frac{333}{106}$, $\frac{355}{113}$ ...).

$\sqrt{2}$ has the representation $[1; 2, 2, 2, \ldots] = [1; \overline{2}]$. (For

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

Similarly, one shows that

$$\sqrt{3} = 1 + \frac{1|}{|1} + \frac{1|}{|2} + \frac{1|}{|1} + \frac{1|}{|2} + \cdots = [1; \overline{1, 2}] \tag{113}$$

$$\sqrt{5} = [2; \overline{4}] \tag{114}$$

$$\sqrt{7} = [2; \overline{1, 1, 1, 4}] \tag{115}$$

$$\tag{116}$$

More generally, consider the fraction

$$x = [\overline{b; a}] = b + \frac{1|}{|a} + \frac{1|}{|b} + \frac{1|}{|a} + \dots$$

where $a, b \in \mathbf{N}$ and $a|b$ (say $b = a.c$). x is a solution of the equation

$$x = b + \frac{1|}{|a} + \frac{1|}{|x} = \frac{(ab+1)x + b}{ax + 1}.$$

Hence $x^2 - bx - c = 0$, i.e. $x = \frac{1}{2}(b + \sqrt{b^2 + 4c})$.

The latter are examples of **periodical** continued fractions i.e. those of the form

$$[a_0; a_1, \dots, a_n, \overline{a_{n+1}, \dots, a_m}].$$

Proposition 29 *ϑ has a periodical representation $\Leftrightarrow$ ϑ is a quadratic irrational number i.e. a solution of an equation*

$$a\vartheta^2 + b\vartheta + c = 0$$

where $a, b, c \in \mathbf{Z}$ and $d = b^2 - 4ac > 0$ is not a quadratic number.

Lemma 4 *Let $x, y \in \mathbf{R}$, with $y > 1$ and let*

$$x = \frac{py + r}{qy + s}$$

where $p, q, r, s \in \mathbf{Z}$ with $ps - qr = \pm 1$. Then if $q > s > 0$, there exists an n, so that

$$\frac{p}{q} = \frac{p_n}{q_n}, \qquad \textit{resp. } \frac{r}{s} = \frac{p_{n-1}}{q_{n-1}}$$

where $(\frac{p_n}{q_n})$ is the sequence of convergents of the continued fraction representation of x.

PROOF. Let $\frac{p}{q} = [a_0; a_1, \dots, a_n]$ be the representationa of $\frac{p}{q}$. We can suppose that

$$ps - qr = (-1)^{n-1}$$

(since we can alwaysfind a representation with n even or odd —see above). Since p and q are relatively prime, we have $p = p_n$, $q = q_n$. Hence

$$p_n s - q_n r = ps - qr = (-1)^{n-1} = p_n q_{n-1} - p_{n-1} q_n$$

and so

$$p_n(s - q_{n-1}) = q_n(r - p_{n-1}).$$

Since $\gcd(p_n, a_n) = 1$,

$$q_n | s - q_{n-1}.$$

Hence $q_n = q > s > 0$ and $q_n \geq q_{n-1} > 0$. Hence $|s - q_{n-1}| < q_n$ and so $s = q_{n-1}$ and $r = p_{n-1}$.

■

Let $y = [a_{n+1}; a_{n+2}, \dots]$ be the representation of y as a continued fraction. We deduce from the equation

$$x = \frac{p_n y + p_{n-1}}{q_n y + q_{n-1}}$$

thet

$$x = [a_0; a_1, \dots, a_n, y] = [a_0; a_1, \dots, a_n, a_{n+1}, \dots].$$

Definition $x, y \in \mathbf{R}$ are said to be **equivalent**, if there exist $a, b, c, d \in \mathbf{Z}$ so that $ad - bc =^{+}_{-} 1$ and $x = \frac{ay+b}{cy+d}$.

Q is an equivalence class with respect to $\sim$ For it is clear that $r \in Q$ can only be equivalent to a rational number. On the other hand, every $r \in Q$ is equivalent ot 0 (and so to each $s \in Q$). For if $r = \frac{p}{q}$ with $\gcd(p, q) = 1$, the there are $s, t \in \mathbf{Z}$ with $ps - qt = 1$. Hence $\frac{p}{q} = \frac{t.0+p}{s.0+q}$. Q.E.D.

Proposition 30 *Two irrational numbers x, y are equivalent if and only if they have representations of the form*

$$x = [a_0; a_1, \dots, a_n, c_0, c_1, \dots], \tag{117}$$
$$y = [b_0; b_1, \dots, b_p, c_0, c_1, \dots] \tag{118}$$

PROOF. ⇐: Let $u = [c_0, c_1, \ldots]$. Then

$$x = \frac{p_n u + p_{n-1}}{q_n u + q_{n-1}}$$

and so $x \sim u$. Similarly, $y \sim u$.

⇒: If $y = \frac{ax+b}{cx+d}$, where $ad - bc = \pm 1$, then we can assume without loss of generality that $cx + d > 0$. Suppose that x has the representation

$$[a_0; a_1, \ldots, a_k, a_{k+1}, \ldots] = [a_0; a_1, \ldots, a_k, a'_k] = \frac{p_{k-1}a'_k + p_{k-2}}{q_{k-1}a'_k + q_{k-2}}.$$

Then $y = \frac{pa'_k + r}{qa'_k + s}$, with $ps - qr = \pm 1$. (In fact, we have

$$p = ap_{k-1} + bq_{k-1} \tag{119}$$

$$q = cp_{k-1} + dq_{k-1} \tag{120}$$

$$r = ap_{k-2} + bq_{k-2} \tag{121}$$

$$s = cp_{k-2} + dq_{k-2}.) \tag{122}$$

We know that

$$p_{k-1} = xq_{k-1} + \frac{\delta}{q_{k-1}} \tag{123}$$

$$p_{k-2} = xq_{k-2} + \frac{\delta'}{q_{k-2}} \tag{124}$$

where $|\delta| < 1,\ |\delta'| < 1$. Hence

$$q = (cx + d)q_{k-1} + \frac{c\delta}{q_{k-1}} \tag{125}$$

$$s = (cx + d)q_{k-2} + \frac{c\delta'}{q_{k-2}} \tag{126}$$

Since $cx + d > 0$, $q_{k-1} > q_{k-2} > 0$, $q_{k-1} \to \infty$, $q_{k-2} \to \infty$,

$$q > s \text{ for } k \text{ large.}$$

We can now deduce the result from the Lemma

■

Proof. Proof of the main result ⇒: Put

$$x = [a_0; a_1, \dots, a_n, \overline{a_{n+1}, \dots, a_{n+p}}] \tag{127}$$
$$y = [\overline{a_{n+1}, \dots, a_p}] \tag{128}$$

Then $y = [a_{n+1}, \dots, a_p, y]$, and so $y = \frac{\overline{p}_r y + \overline{p}_{r-1}}{\overline{q}_r y + \overline{q}_{r-1}}$ for suitable $\overline{p}_r, \overline{p}_{r-1}, \overline{q}_r, \overline{q}_{r-1}$. This is a quadratic equation. Hence y is a quadratic irrational number. But $x = \frac{p_n y + p_{n-2}}{q_n y + q_{n-1}}$ and so $x \sim y$. The result now follows form the simple fact that a number which is equivalent to a quadratic irrational number is itself quadratic irrational.

⇐: Let x be a solution of the equation $ax^2 + bx + c$, with representation $x = [a_0; a_1, a_2, \dots]$. Then for each n,

$$x = [a_0; a_1, \dots, a'_n, \dots] = \frac{p_{n-1}a'_n + q_{n-2}}{q_{n-1}a'_n + q_{n-2}}.$$

a'_n is thus a solution of the quadratic equation

$$A_n x^2 + B_n x + C_n = 0$$

where

$$A_n = ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2 \tag{129}$$
$$B_n = 2ap_{n-1}p_{n-2} + b(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + 2cq_{n-1}q_{n-2} \tag{130}$$
$$C_n = ap_{n-2}^2 + bp_{n-2}q_{n-2} + cq_{n-2}^2. \tag{131}$$

Then $A_n \neq 0$ (otherwise $\frac{p_{n-1}}{q_{n-1}}$ would be a rational solution of the equation $ax^2 + bx + c = 0$)) and

$$B_n^2 - 4A_nC_n = b^2 - 4ac.$$

Since $p_{n-1} = xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}}$ with $|\delta_{n-1}| < 1$,

$$A_n = a(xq_{n-1} + \frac{\delta_{n-1}}{q_{n-1}})^2 + bq_{n-1}(x + \frac{\delta_{n-1}}{q_{n-1}}) + cq_{n-1}^2 \tag{132}$$
$$= 2ax\delta_{n-1} + a\frac{\delta_{n-1}^2}{q_{n-1}^2} + b\delta_n^{-1} \tag{133}$$

and so $|A_n| < 2|ax| + |a| + |b|$. Similarly, we have $|C_n| \leq 2|ax| + |a| + |b|$ and so $B_n^2 \leq 4|A_nC_n| + |b^2 - 4ac|$. Hence the family of quadratic equstions which are satisfied by (a'_n) is finite . Hence there exist n_1 and n_2 with $a'_{n_1} = a'_{n_2}$. Q.E.D.

■

We shall now show that the approximant $\frac{p_n}{q_n}$ is in a certain sense the best rational approximation of x.

Proposition 31 *For $n > 1$, $0 < q \leq q_n$ and $\frac{p}{q}$ a rational number ($\neq \frac{p_n}{q_n}$),*

$$|p_n - q_n x| < |p - qx|$$

and so

$$|\frac{p_n}{q_n} - x| < |\frac{p}{q} - x|.$$

PROOF. We begin with the remark that

$$\frac{1}{q_{n+2}} < |p_n - q_n x| < \frac{1}{q_{n+1}}.$$

(For $|x - \frac{p_n}{q_n}| = \frac{1}{q_n q'_{n+1}}$ and so $|q_n x - p_n| = \frac{1}{q'_{n+1}}$, where $q'_{n+1} = a'_{n+1} q_n + q_{n-1}$ and so

$$q'_{n+1} > a_{n+1} q_n + q_{n-1} = q_{n+1}$$

resp.

$$q'_{n+1} < a_{n+1} q_n + q_{n-1} + q_n = q_{n+1} + q_n \leq a_{n+2} q_{n+1} + q_n = q_{n+2})$$

$\frac{1}{q_{n+2}} < |q_n x - p_n| < \frac{1}{q_{n+1}}$ as claimed. Hence it suffice to show that the statement holds for $q_{n-1} < q \leq q_n$.

Case 1: $q = q_n$. Then $|\frac{p_n}{q_n} - \frac{p}{q_n}| \geq \frac{1}{q_n}$ bzw. $|\frac{p_n}{q_n} - x| \leq \frac{1}{q_n q_{n+1}} < \frac{1}{2q_n}$. Q.E.D.

Case 2: $q_{n-1} < q < q_n$. In this case, neither $\frac{p}{q}$ not $\frac{p_{n-1}}{q_{n-1}}$ nor $\frac{p_n}{q_n}$.

Since the matirx

$$\begin{bmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{bmatrix}$$

is invertible over $\mathbf{Z}$, There exist uniquely determined $\lambda_1, \lambda_2 \in \mathbf{Z}$, so that

$$\lambda_1 p_n + \lambda_2 p_{n-1} = p \tag{134}$$

$$\lambda_1 q_n + \lambda_2 q_{n-1} = q \tag{135}$$

Since $q = \lambda_1 q_n + \lambda_2 q_{n-1} < q_n$, $\lambda_1 \lambda_2 < 0$. But $(p_n - q_n x)(p_{n-1} - q_{n-1} x) < 0$ and so

$$\lambda_1 (p_n - q_n x) . \lambda_2 (p_{n-1} - q_{n-1} x) > 0.$$

But $p - qx = \lambda_1 (p_n - q_n x) + \lambda_2 (p_{n-1} - q_{n-1} x)$ an o $|p - qx| > |p_{n-1} - q_{n-1} x| > |p_n - q_n x|$. Q.E.D.

■

The Farey sequence $\mathcal{F}_\backslash$ is the finite sequence of all fractions $\frac{p}{q} \in [0,1]$, so that $q \leq n$.

$$\begin{array}{llllllllllll}
\mathcal{F}_\infty & 0 & 1 \\
\mathcal{F}_\in & 0 & \frac{1}{2} & 1 \\
\mathcal{F}_\ni & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{3} & 1 \\
\mathcal{F}_\triangle & 0 & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} & \frac{4}{5} & 1 \\
\mathcal{F}_\bigtriangledown & 0 & \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{2}{5} & \frac{1}{2} & \frac{1}{2} & \frac{3}{5} & \frac{2}{3} & \frac{3}{4} & \frac{4}{5} & 1
\end{array} .$$

Proposition 32 *If $\frac{l}{m}$ is the successor of $\frac{h}{k}$ $\mathcal{F}_\backslash$, then $kl - hm = 1$.*

Proof. This is proved by induction: It follows from the above table that the theroem holds for $\mathcal{F}_\backslash(\backslash \leq \bigtriangledown)$.

Suppose now that $\mathcal{F}_\infty, \ldots, \mathcal{F}_{\backslash+\infty}$ is true and consider the case $\mathcal{F}_{\backslash+\infty}$. Let $\frac{a}{b} \in \mathcal{F}_{\backslash+\infty} \setminus \mathcal{F}_\backslash$. If $\frac{h}{k}$ resp. $\frac{l}{m}$ is the predecessor resp. the succesoor of $\frac{a}{b}$ in $\mathcal{F}_{\backslash+\infty}$, then $\frac{h}{k}$ is the predecessor of $\frac{l}{m}$ in $\mathcal{F}_\backslash$. Hence $kl - hm = 1$ (induction hypothesis). Let $\frac{p}{q}$ be a rational number with $\frac{h}{k} < \frac{p}{q} < \frac{l}{m}$. The system

$$\lambda l + \mu h = p \qquad \lambda m + \mu k = q$$

then has a unique whole number solution $\lambda = kp - hq$, $\mu = ql - pm$ (where $\lambda, \mu > 0$). Conversely, for each $\lambda, \mu \in \mathbf{N}$ the corresponding rational number $\frac{}{q}$ between $\frac{h}{k}$. The characterisation of $\frac{a}{b}$ that the latter is that fraction for which q is minimal. But this means that $\lambda = \mu = 1$ and so $\frac{a}{b} = \frac{h+l}{k+m}$. It then follow easily that $ak - bh = 1$.

■

Corollar 9 *Let $\frac{a}{b} \in \mathcal{F}_\backslash$ with predecessor $\frac{h}{k}$ and successor $\frac{l}{m}$. Then $\frac{a}{b} = \frac{hn+l}{k+m}$.*

Example We can use the Farey series to solve the diophantine equation $ax + by = 1$. Without loss of generality, we can assume that $0 < a < b$. Choose n, so that $\frac{a}{b} \in \mathcal{F}_\backslash$. Let $\frac{h}{k}$ be the predecessor of $\frac{a}{b}$ $\mathcal{F}_\backslash$. Then $ak - bh = 1$, i.e. $(k, -h)$ is a solution.

Rational Approximation

Proposition 33 *Let $\xi \in]0,1[$ be irrational, $N \in \mathbf{N}$. Then there exist $\frac{p}{q} \in \mathbf{Q}$ with $q \leq N$, so that $\left|\xi - \frac{p}{q}\right| < \frac{1}{qN}$ (and so $\leq \frac{1}{N^2}$).*

Proof. Consider the intervals $\Big]0,\frac{1}{N}\Big[,\dots\Big]\frac{N-1}{N},1\Big[$ and the numbers $\big\{n\xi - [n\xi]\colon n=1,\dots,N\big\}$. By the pigeon hole principle we have.: either there is an n so that $n\xi-[n\xi]\in\Big]0,\frac{1}{N}\Big[$ or there is an $m\neq n$, so that $n\xi-[n\xi]$ uand $m\xi-[n\xi]$ are in the same interval. Im ersten Fall gilt: In the first case we have $1\frac{}{N}$ in the second

$$\big|(n-m)\xi-\big([n\xi]-[m\xi]\big)\big| < frac1N$$

Henc we can put $p=[n\xi], q=n$ resp. $p=[n,\xi]-[m\xi], q=n-m$.

■

Alternative proof We use the theory of Farey series. There is a $\frac{a}{b}\in\mathcal{F}_\mathcal{N}$ with $\frac{a}{b}<\xi<\frac{c}{d}$, where $\frac{c}{d}$ is the successor of $\frac{a}{b}$ in $\mathcal{F}_\backslash$. Then $\xi\in\Big]\frac{a}{b},\frac{a+c}{b+d}\Big[$ or $\xi\in\Big]\frac{a+c}{b+d},\frac{c}{d}\Big[$. In the first case, we have

$$0<\xi-\frac{a}{b}<\frac{a+c}{b+d}-\frac{a}{b}\leq\frac{1}{b(N+1}$$

since $b+d\geq N+1$.

■

Corollar 10 *Let $\xi\in[0,1]$ be irrational. Then there exist infinitely many rational numbers $\frac{p}{q}$, so that $\Big|\xi-\frac{p}{q}\Big|<\frac{1}{q^2}$.*

Using continued fractions, this result can be improved as follows: Consider the convergents $\frac{p_n}{q_n}$, $\frac{p_{n+1}}{q_{n+1}}$. Then $\Big|\xi-\frac{p}{q}\Big|<\frac{1}{2q^2}$ for some $\frac{p}{q}\in\Big\{\frac{p_n}{q_n},\frac{p_{n+1}}{q_{n+1}}\Big\}$. For

$$q_nq_{n+1}<\frac{1}{2q_n^2}+\frac{1}{2q_{n+1}^2}$$

(The last step employs the inequality $\alpha\beta\leq\frac{1}{2}\big(\alpha^2+\beta^2\big)\big)$. One can improve the result still further by using the convergents

$$\frac{p_n}{q_n},\frac{p_{n+1}}{q_{n+1}},\frac{p_{n+2}}{q_{n+2}}$$

. We claim that

$$\Big|\xi-\frac{p}{q}\Big|<\frac{1}{\sqrt{5}q^2}$$

for an

$$\frac{p}{q}\in\{\frac{p_n}{q_n},\frac{p_{n+1}}{q_{n+1}},\frac{p_{n+2}}{q_{n+2}}\}$$

. If this is not the case, then

$$\frac{1}{\sqrt{5}q_n^2} + \frac{1}{\sqrt{5}q_{n+1}^2} \leq \frac{1}{q_n q_{n+1}},$$

and so $\lambda + \frac{1}{\lambda} < \sqrt{5}$, where $\lambda = \frac{q_{n+1}}{q_n}$ (λ is rational).

By elementary manipulations, we get the inequality $\lambda < \frac{1}{2}(1+55)$ for λ. Similarly, $\mu < \frac{1}{2}(1+55)$, where $\mu = \frac{q_{n+2}}{q_{n+1}}$.
But from the relationship

$$q_{n+2} = a_{n+2}q_{n+1} + q_n$$

we can deduce that $\mu \geq 1 + \frac{1}{\lambda}$.
This is a contradiction since $\lambda < \frac{1}{2}(1+\sqrt{5})$ implies that

$$\frac{1}{\lambda} > \frac{1}{2}(-1+\sqrt{5})$$

.

Remark This result is sharp as the example

$$\tau = [1, 1, 1, 1, \dots]$$

shows.

Now let ξ be a quadratic irrational number. From the general relationship

$$\left|\xi - \frac{p_n}{q_n}\right| \geq \frac{1}{(a_{n+1}+2)q_n^2}$$

and the fact that the sequence (a_n) is bounded,l we deduce that there exists $c\big(= c(\xi)\big) > 0$, so that

$$\left|\xi - \frac{p}{q}\right| > \frac{c}{q^2}$$

for each $\frac{p}{q} \in \mathbf{Q}$.

More generally

Proposition 34 *Let ξ be an algebraic number of degree n. Then there exists $c > 0$, so that*

$$\left|\xi - \frac{p}{q}\right| > \frac{c}{q^n}$$

for each $\frac{p}{q} \in \mathbf{Q}$.

PROOF. Let P be a minimal polynomial over over $\mathbf{Z}$, so that $P(\xi) = 0$ (P is then irreducible over $\mathbf{Q}$). For $p, q \in \mathbf{Z}$ and $q > 0$ we have

$$P(\xi) - P\Big(\frac{p}{q}\Big) = \Big(\xi - \frac{p}{q}\Big) P'(\xi_0)$$

where $\xi_0 \in \Big]\xi, \frac{p}{q}\Big[$ (resp. $\Big]\frac{p}{q}, \xi\Big[$). Then $P\Big(\frac{p}{q}\Big) \neq 0$ and $q^n P\Big(\frac{p}{q}\Big) \in \mathbf{Z}$. We thus have the estimate $\Big|P\Big(\frac{p}{q}\Big)\Big| \geq \frac{1}{q^n}$. We now choose c so that $\Big|P'(\xi_0)\Big| < \frac{1}{c}$ if $|\xi_0 - \xi| \leq 1$. Then $\Big|\xi - \frac{p}{q}\Big| > \frac{c}{q^n}$ as claimed.

■

2 Geometry

2.1 Triangles:

We begin with one of the simplest, but richest of geometrical figures—the triangle. A triangle is determined by its three vertices A, B and C. (Figure 1). It is then denoted by ABC. Normally we shall assume that it is non-degenerate i.e. that A, B and C are not collinear. This can be expressed analytically as the statement that the vectors $x_B - x_B$ and $x_C - x_A$ are linearly independent i.e. that there are no non-trivial pairs (λ, μ) of scalars so that

$$\lambda(x_B - x_A) + \mu(x_C - x_A) = 0.$$

(non-trivial means that either $\lambda \neq 0$ or $\mu \neq 0$).

We can rewrite this equation in the form

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C = 0$$

where $\lambda_1 = -(\lambda + \mu)$, $\lambda_2 = \lambda, \lambda_3 = \mu$.

This leads to the following more symmetric description of the non-degeneracy of ABC: the triangle is non-degenerate if and only if there is no triple $(\lambda_1, \lambda_2, \lambda_3)$ of scalars so that $\lambda_1 + \lambda_2 + \lambda_3 = 0$ and

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C = 0.$$

The vectors x_A, x_B and x_C are then said to be **affinely independent**. In this case, any point x in $\mathbf{R}^2$ can be written as

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Before proving this we first note that these λ are uniquely determined by x. for if we have a second such representation, say with coefficients μ_1, μ_2, μ_3, then substracting leads to the equation

$$(\lambda_1 - \mu_1)x_A + (\lambda_2 - \mu_2)x_B + (\lambda_3 - \mu_3)x_C = 0$$

where the sum of the coefficients is zero and the λ's nand the μ's coincide. We are therefore justified in calling the λ's the **barycentric coordinates** of x with respect to A, B and C.

We now turn to the existence: suppose that x is the coordinate vector of the point P and produce AP to meet BC at the point Q (see figure 2). We are tacitly assuming that P is distinct from A and that AP is not parallel to BC - these two special cases are simpler. There are scalars λ, μ so that

$$x_Q = \lambda x_B + (1 - \lambda)x_C$$

and

$$x_P = \mu x_A + (1 - \mu)x_Q.$$

Then we have

$$x_P = \mu x_A + (1 - \mu)\lambda x_B + (1 - \mu)(1 - \lambda)x_C$$

and the sum of the coefficients is

$$\mu + (1 - \mu)[\lambda + (1 - \lambda)] = 1.$$

From this we see that the barycentric coordinates have the following geometrical interpretation. For simplicity we shall assume that the point P lies within the triangle (i.e. that its barycentric coordinates are all positive). If we refer to figure 2, we see that the coefficient λ_1 of x_A is the ratio $\frac{|PQ|}{|AQ|}$ of the lengths of PQ and AQ and we see from figure 3 that this is the ratio of the distance of P resp. of A to the line BC.

The point x with barycentric coordinates $(\lambda_1, \lambda_2, \lambda_3)$ can be regarded as the centroid of a physical system consisting f three particles A, B and C with masses λ_1, λ_2 and λ_3 respectively. The important case of equal masses produces the **centroid** i.e. the point M with $x_M + \frac{1}{3}(x_A + x_B + x_C)$. We see from the above that, as expected, M lies two thirds of the way along the line from A to the midpoint of BC (figure 4) (for $x_M = \frac{1}{3}x_A + \frac{3}{2}(\frac{x_B+x_C}{2})$.

The existence of barycentric coordinates can be derived in a more analytic fashion as follows: since $(x_B - x_A)$ and $(x_C - x_A)$ are linearly independent, they span $\mathbf{R}^2$. Hence there exist λ, μ so that

$$x_P - x_A = \lambda(x_B - x_A) + \mu(x_C - x_A)$$

and so

$$x_P = (1 - \lambda - \mu)x_A + \lambda x_B + \mu x_C.$$

We can now use this apparatus to prove a famous result on the collinearity of points on the sides of a triangle, known as the theorem of Menelaus. We suppose that ABC is a non-degenerate triangle and that P, Q and R are points on BC resp. CA resp. AB. (figure 5). Let the barycentric coordinates of P,Q and R be as follows

$$x_P = (1 - t_1)x_B + t_1 x_C \tag{136}$$
$$x_Q = (1 - t_2)x_C + t_2 x_A \tag{137}$$
$$x_R = (1 - t_3)x_A + t_3 x_B. \tag{138}$$

Then the points P,Q and R are collinear if and only if

$$t_1 t_2 t_3 = -(1 - t_1)(1 - t_2)(1 - t_3).$$

(The classical statement is that collinearity occurs when

$$\frac{|AP|}{|PB|} \cdot \frac{|BQ|}{|QC|} \cdot \frac{|CR|}{|RA|} = -1$$

where the negative sign is to be interpreted as stating that if two of the points P, Q and R are internal, then the third is external, resp. if two are external, then so is the third). In order to prove this, we introduce the notation $D_{\frac{\pi}{2}}(x)$ for the vector $(-\xi_2, \xi_1)$ i.e. x rotated through 90^0 anti-clockwise (see figure 6). We then introduce the notation $x \wedge y$ for the scalar product

$$-(x|D_{\frac{\pi}{2}}y) = (D_{\frac{\pi}{2}}x|y)$$

i.e. the expression $(\xi_1\eta_2 - \xi_2\eta_1)$ which the reader will recognise as twice the signed area of the triangle with vertices at O, x and y. (It is also the determinant of the 2×2 matrix with the coordinates of x and y as columns).

The expression $x \wedge y$ satisfies the equations

$$(x + x_1) \wedge y = x \wedge y + x_1 \wedge y \tag{139}$$
$$x \wedge y = -y \wedge x \tag{140}$$
$$x \wedge x = 0. \tag{141}$$

The equation $x \wedge y = 0$ is equivalent to the fact that x and y are proportional i.e. that O, x and y are collinear. if we apply this to the vectors x_{RP} and x_{RQ}, we see that P, Q and R are collinear if and only if

$$(x_P - x_R) \wedge (x_Q - x_R) = 0.$$

Now

$$x_P - x_R = -(1-t_3)x_A + (1-t_1-t_3)x_B + t_1 x_C$$

and

$$x_Q - x_R = -(1-t_2-t_3)x_A - t_3 x_B + (1-t_3)x_C.$$

At this stage we note that we can assume without loss of generality that C is at the origin i.e. that $x_C = 0$. (We have avoided making such an assumption at an earlier stage in order not to artificially destroy the symmetry between A, B and C). Then the above expression is the numerical quantity

$$(1-t_3)t_3 + (1-t_1-t_3)(1-t_2-t_3)$$

times $x_A \wedge x_B$ and so the required condition is

$$(1-t_3)t_3 + (1-t_1-t_3)(1-t_2-t_3) = 0$$

which can be reduced by an easy calculation to the desired one.

A result which is in a certain sense dual to the theorem of Menelaus is Ceva's theorem which we shall now proceed to state and prove. let ABC be a triangle and let A', B' and C' be points on the opposite sides as in figure 7. Then the lines AA', BB' and CC' are concurrent if and only if

$$\frac{|BA'|}{|A'C|} \cdot \frac{|CB'|}{|B'A|} \cdot \frac{|AC'|}{|C'B|} = 1.$$

Once again, this means that if

$$x_{A'} = \lambda_1 x_B + (1-\lambda_1)x_C \qquad (142)$$

$$x_{B'} = \lambda_2 x_C + (1-\lambda_2)x_A \qquad (143)$$

$$x_{C'} = \lambda_3 x_A + (1-\lambda_3)x_B \qquad (144)$$

then

$$\lambda_1\lambda_2\lambda_3 = (1-\lambda_1)(1-\lambda_2)(1-\lambda_3).$$

We prove this as follows: suppose that P is a point of intersection, with coordinates (μ_1, μ_2, μ_3) i.e.

$$x_P = \mu_1 x_A + \mu_2 x_B + \mu_3 x_C.$$

We calculate the coordinates of A' as the intersection of AP and BC i.e. we must find a suitable combination of x_A and $\mu_1 x_A + \mu_2 x_B + \mu_3 x_C$ which does not contain the term x_A. This must clearly be m_1 times x_A minus the second term and this leads to the result

$$x_{A'} = \frac{1}{\mu_2 + \mu_3}(\mu_2 x_B + \mu_3 x_C)$$

with corresponding expressions for $x_{B'}$ and $x_{C'}$. Hence

$$\lambda_1 = \frac{\mu_2}{\mu_2 + \mu_3} \qquad 1 - \lambda_1 = \frac{\mu_3}{\mu_2 + \mu_3} \tag{145}$$

$$\lambda_2 = \frac{\mu_3}{\mu_3 + \mu_1} \qquad 1 - \lambda_2 = \frac{\mu_1}{\mu_3 + \mu_1} \tag{146}$$

$$\lambda_3 = \frac{\mu_1}{\mu_1 + \mu_2} \qquad 1 - \lambda_3 = \frac{\mu_2}{\mu_1 + \mu_2}. \tag{147}$$

from which the above equation follows immediately.

The converse direction can be deduced from this as follows. Suppose that this relationship between the barycentric coordinates holds. Let P be the point of intersection of BB' and CC' and let AP meet BC in A''. Then the barycentric coordinates of A'' satisfy the same equation as those of A' by the above result and so A' and A'' coincide.

As an application of Ceva's theorem, we show that the bisectors of the angles of a triangle are concurrent. We use the standard notation

$$|BC| = a \quad |CA| = b \quad |AB| = c$$

and let A' denote the point on BC where the bisector of the angle at A meets BC (figure 8). Then we claim that

$$\frac{|BA'|}{|A'B|} = \frac{c}{b}$$

i.e. that

$$x_{A'} = \frac{1}{b+c}(bx_B + cx_C)$$

and the result then follows easily from Ceva's theorem. In order to do this we introduce the unit normal vectors

$$\mathbf{n}_1 = \frac{1}{a} D_{\frac{\pi}{2}} x_{BC} \quad \mathbf{n}_2 = \frac{1}{b} D_{\frac{\pi}{2}} x_{CA} \quad \mathbf{n}_3 = \frac{1}{c} D_{\frac{\pi}{2}} x_{AB}$$

to the three sides (figure 9).

Then the bisector of the angle at A consists of those points P so that

$$(x_{AP}|\mathbf{n}_2) = -(x_{AP}|\mathbf{n}_3).$$

This it suffices to substitute $\frac{1}{b+c}(bx_B + cx_C)$ for P in this equation and verify that it is valid. If we make this substitution and assume (as we may without loss of generality) that $x_A = 0$, then the left hand side reduces to $x_B \wedge x_C$ and the right hand to $-x_C \wedge x_B$.

It follows easily from the above that the incentre of the triangle has barycentric representation

$$\frac{1}{s}(ax_A + bx_B + cx_C)$$

where $s = a + b + c$ is the perimeter of the triangle. For one sees at a glance that the above point is a suitable combination of x_A and $x_{A'}$ (in fact,

$$\frac{a}{s}x_A + \frac{b+c}{s}x_{A'})$$

and so lies on each of the bisectors of the angles of ABC by symmetry.

2.2 Complex numbers and geometry

As the reader knows, the points of the plane can be identified with the set of complex numbers and we shall now use this fact to give a simple approach to various special points, lines and circles associated with a triangle. Let ABC be such a triangle and let z_1, z_2, z_3 denote the corresponding complex numbers. If we take the centre of the circumscribed circles to be the origin and suppose that its radius is 1, then

$$|z_1| = |z_2| = |z_3| = 1.$$

(figure 10). The centroid of the triangle is the point M with complex coordinate $m = \frac{1}{3}(z_1 + z_2 + z_3)$. Let H_3 be the point with coordinate $z_1 + z_2$ (figure 11). Then $AOBH_3$ is a rhombus whose centre is the midpoint of AB as can be checked as follows: it is a parallelogram since

$$(z_1 + z_2) - z_1 = z_2 - 0.$$

Also the sides OB and OA are equal in length. Then centre of this rhombus is

$$\frac{1}{4}(z_1 + z_2 + z_3 + z_4) = \frac{1}{2}(z_1 + z_2)$$

i.e. the midpoint of AB. Consider now the point H with coordinate $h = z_1 + z_2 + z_3$. OH_3HC is a parallelogram since $h_3 - 0 = h - z_3$. This means that CH is parallel to OH and so is perpendicular to AB. Hence H is the orthocentre of ABC i.e. the intersection of the perpendiculars from the vertices to the opposite sides, since, by symmetry, it lies on the other perpendiculars.

Since $m = \frac{1}{3}h$, the centroid M, the circumcentre and the orthocentre lie on a straight line which is called the **Euler line**. In fact, M lies a third of the way along the segment OH.

We now introduce a third point E on this line (figure 11), the midpoint of OH. Thus E has coordinate $e = \frac{1}{2}(z_1 + z_2 + z_3)$. The circle with centre E and radius $\frac{1}{2}$ is called the **nine-point circle** since it passes through nine significant points of the circle as we shall now show. First note that E is the centre of the parallelogram $OCHH_3$ and sop

$$|EM_3| = \frac{1}{2}|HH_3| = \frac{1}{2}|OC| = \frac{1}{2}$$

where M_3 is the midpoint of AB. Hence M_3 (and so the midpoints of all three sides) are on this circle.

Now if K denotes the foot of the perpendicular CH from C to AB, then it is clear from diagram 12 that

$$|EK| = |EM_3| \quad (= \frac{1}{2})$$

(since $HK \perp OM$ and E is the midpoint of OH). Hence our circle passes through the three feed of the perpendiculars.

Finally, if L is the midpoint of the segment AH (so that L has coordinate

$$\frac{1}{2}(z_1 + (z_1 + z_2 + z_3)) = z_1 + \frac{1}{2}(z_2 + z_3)$$

then, since $|EH| = |EO|$, $|EL| = \frac{1}{2}$ i.e. L (and the two corresponding points for B and C) also lie on the circle.

2.3 Quadrilaterals

We now turn to figure which are determined by four vertices—the quadrilaterals. The quadrilateral with vertices A, B, C, and D is denoted by $ABCD$ (figure 1). Note that, in contrast to the case of triangles, the order of the vertices is important. Thus $ABDC$ is the quadrilateral of figure 2. In general, we shall assume that the quadrilateral is not self-intersecting (i.e. the above case is excluded) and convex (i.e. the ;quadrilateral of figure 2 is excluded). In fact, this is more a question of convenience since most of what we shall do is independent of such restrictions.

In contrast to the case of a triangle, the vertices of a quadrilateral cannot be affinely independent i.e. there exist non-zero scalars λ_1, λ_2, λ_3, λ_4 so that $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 0$ and

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_D = 0.$$

Some interesting special types of quadrilateral can be characterised by the type of relationship which holds for the vertices. For example, a **trapezoid** is a quadrilateral with two opposite sides (say, AB and DC parallel). In this case, we have a relation of the form

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_D = 0$$

where $\lambda_1 + \lambda_2 = 0 = \lambda_3 + \lambda_4$. For there is a scalar μ so that $x_{AB} = \mu x x_{DC}$ and this reduces to an equation of the above form (figure 3).

Parallelograms are special types of trapezoids with the relation

$$x_A - x_B + x_C - x_D = 0.$$

If AB and DC are not parallel, then in the equation

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_D = 0$$

we have, by the same reasoning, that $\lambda_1 + \lambda_2 \neq 0$ and $\lambda_3 + \lambda_4 \neq 0$. Hence the point P with coordinates

$$\xi_P = \frac{\lambda_1 x_A + \lambda_2 x_B}{\lambda_1 + \lambda_2} = \frac{\lambda_3 x_C + \lambda_4 x_D}{\lambda_3 + \lambda_4}$$

lies on both lines and so is their intersection (figure 4).

We use the above characterisation of parallelograms to verify three simple results on them:
I. The midpoints of the sides of a quadrilateral $ABCD$ form the vertices of a parallelogram. For the coordinates of the midpoints are

$$x_P = \frac{1}{2}(x_A + x_B) \quad x_Q = \frac{1}{2}(x_B + x_C)$$

etc. and it follows easily that

$$x_P - x_Q + x_R - x_S = 0$$

as required. (figure 5).
II. Suppose that $ABCD$ is a parallelogram and that P and Q are such that AP and QC are equal and parallel (figure 6). Then we claim that $BQDP$ is also a parallelogram. For the fact that $ABCD$ and $APCQ$ are parallelograms can be expressed in the equations

$$x_A - x_B + x_C - x_D = 0$$

and

$$x_A - x_P + x_C - x_Q = 0.$$

Substracting leads to the equation

$$x_B - x_P + x_D - x_Q = 0$$

which is the required result.

III. Consider a quadrilateral $ABCD$ and let B_1 and D_1 be such that ABB_1C and CD_1D are parallelograms. Then BB_1D_1D is a parallelogram (figure 7). For the hypotheses can be expressed in the equations

$$x_A - x_B + x_{B_1} - x_C = 0$$

and

$$x_A - x_C + x_{D_1} - x_D = 0.$$

Substracting, we get the equation

$$x_B - x_{B_1} + x_{D_1} - x_D = 0$$

as required.

The above result is interesting in that it shows how to construct for a given quadrilateral $ABCD$ a parallelogram and a point C so that the lines from C to the vertices of the parallelogram are equal and parallel to te vertices of the given quadrilateral. Also the four angles subtended at C by the sides of the parallelogram are equal to the angles of the original quadrilateral. We remark that the area of the constructed parallelogram is twice that of the original quadrilateral.

Cyclic quadrilaterals In general, a quadrilateral cannot be inscribed in a circle, in contrast to the case of triangles. Those which *can* (the so-called **cyclic quadrilaterals**) form the setting for some of the most elegant results of elementary geometry. We shall deduce some these with the methods of complex numbers used to discuss the nine-point circle of a triangle. We choose the coordinate system so that the quadrilateral $ABCD$ is inscribed in the unit circle i.e. that the complex coordinates z_1, z_2, z_3, z_4 all have absolute values 1. (figure 8). We can associate to the quadrilateral in a natural way four triangles BCD, CDA, DAB and ABC. These all have circumcentre O and their orthocentres are H_1, H_2, H_3 and H_4 with coordinates $z_2 + z_3 + z_4$ etc. Their centroids will be denoted by M_1, M_2, M_3 M_4 and have coordinates $\frac{1}{3}(z_2 + z_3 + z_4)$ etc. We introduce the points H and M with coordinates

$$z_1 + z_2 + z_3 + z_4 \quad \text{and} \quad \frac{1}{4}(z_1 + z_2 + z_3 + z_4)$$

respectively and call them the **orthocentre** resp. the **centroid** of $ABCD$. Now the vector HH_4 is represented by the complex number z_4 and so is equal and parallel to OD, in particular has length one. This implies that the circles of radius 1 with centres at the orthocentres of the four triangles intersect at the point H, a fact which gives a geometrical interpretation of the latter which was initially introduced for purely formal reasons.

This encourages us to introduce the point E with coordinate $\frac{(}{z\,1} + z_2 + z_3 + z_4)$ and to seek a geometrical description of it. if we denote by E_1, E_2, E_3 and E_4 the centres of the Euler circles of the four triangles, then the vector EE_4 is represented by the complex number

$$\frac{(}{z\,1} + z_2 + z_3 + z_4) - \frac{1}{2}(z_1 + z_2 + z_3) = \frac{z_4}{2}.$$

In particular, it has length $\frac{1}{2}$. Thus the Euler circles of the four triangles intersect at the point E which is thus provided with the sought geometrical interpretation. Another way of looking at this result is as follows: the circle with centre E and radius $\frac{1}{2}$ passes through the Euler centres of the four triangles. It is called the **Euler circle** of the quadrilateral.

We now proceed to give a geometrical interpretation to the point M. Consider the lie through A and M_4, the centroid of BCD. We calculate the ratio

$$\frac{m - z_1}{m_4 - z_1} = \frac{3(z_1 + z_2 + z_3 + z_4 - 4z_1)}{4(z_2 + z_3 + z_4 - 3z_1)} = \frac{3}{4}.$$

This means that M lies on the segment AM_4 (and in fact is three quarters of the way along it). Hence M is the intersection of the lines joining the vertices of $ABCD$ with the centroids of the opposite triangles.

As a last remark on the points O, H, M and M introduced here, we note that they all lie on the line joining O and H (since their coordinates are all real multiples of $(z_1 + z_2 + z_3 + z_4)$). This line is naturally called the **Euler line** of the quadrilateral. (figure ??).

The complete quadrilateral This is the figure obtained from a quadrilateral by producing the opposite sides, respectively the diagonals, until they intersect (figure ???). Suppose that we have the relationship

$$\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_D = 0$$

where $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 0$ but $\lambda_1 + \lambda_2 \neq 0$, $\lambda_1 + \lambda_3 \neq 0$ and $\lambda_1 + \lambda_4 \neq 0$ (i.e. no sum of two of the l's is zero). This condition ensures that none of the three points P, Q and R in the figure is "at infinity". (i.e. the corresponding lines are not parallel). Using the methods employed above, we can calculate the coordinates of the points of intersection as follows:

$$\xi_P = \frac{\lambda_1 x_A + \lambda_2 x_B}{\lambda_1 + \lambda_2} = \frac{\lambda_3 x_C + \lambda_4 x_D}{\lambda_3 + \lambda_4} \tag{148}$$

$$\xi_Q = \frac{\lambda_1 x_A + \lambda_4 x_D}{\lambda_1 + \lambda_4} = \frac{\lambda_2 x_B + \lambda_3 x_C}{\lambda_2 + \lambda_3} \tag{149}$$

$$\xi_P = \frac{\lambda_1 x_A + \lambda_3 x_C}{\lambda_1 + \lambda_3} = \frac{\lambda_2 x_B + \lambda_4 x_D}{\lambda_2 + \lambda_4}. \tag{150}$$

We denote by E the point of intersection of AB and QR and calculate its barycentric coordinates with respect to A and B as follows: we must find a suitable combination of x_Q and x_R which is also a combination of x_A and x_B. it is clear that this must be

$$\frac{\lambda_2 + \lambda_3}{\lambda_2 - \lambda_1} x_Q - \frac{\lambda_1 + \lambda_3}{\lambda_2 - \lambda_1} x_R = \frac{\lambda_2}{\lambda_2 - \lambda_1} x_B - \frac{\lambda_1}{\lambda_2 - \lambda_1} x_A.$$

(we are tacitly assuming that $\lambda_1 \neq \lambda_2$ – the reader should consider what happens in the case that these two values coincide). If we compare the coordinates

$$x_P = \frac{\lambda_1}{\lambda_1 + \lambda_2} x_A + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_B$$

of P with those of E, we see that the ratios $\frac{|AE|}{|EB|}$ and $\frac{|AP|}{|PB|}$ are numerically equal but opposite in sign (they are $-\frac{\lambda_1}{\lambda_2}$ and $\frac{\lambda_1}{\lambda_2}$ respectively). Points E and P with this property are said to be **harmonic conjugates** with respect to A and B.

The above result suggests the following method for constructing the harmonic conjugate of a given point E on a segment AB with respect to A and B. let Q be a point which is not on AB and join AQ, EQ and BQ. Let C be a point on BQ and join AC. Let AC meet EQ at T and produce BR to meet AQ at D. Then the required point is P, the intersection of DC and AB.

The concept of harmonic conjugates can be characterised by using the so-called **cross-ratio** which can be most conveniently defined via complex numbers. Suppose that A, B, C are distinct points in the plane with complex coordinates z_1, z_2 and z_3. Then the number $\dfrac{z_3 - z_1}{z_3 - z_2}$ is real if and only if the points are collinear and then it is just the signed ratio $\frac{|AC|}{|CB|}$. If D is a fourth point with coordinate z_4, then the cross-ratio $(A, B; C, D)$ of the four points is the complex number

$$\frac{z_3 - z_1}{z_3 - z_2} \div \frac{z_4 - z_1}{z_4 - z_1}.$$

We shall show below that this is real if and only if the quadrilateral $ABCD$ is cyclic or the points A, B, C and D are collinear. If A, B and C are collinear, then it is real if and only if D lies on this line and C and D are harmonically conjugate if and only if its value is -1.

2.4 Polygon

Having examined triangles and quadrilaterals, we now turn to pentagons, hexagons etc. The generic name for such figures is "polygon". A polygon with n-vertices is described by listing the latter as $A_1 A_2 \dots A_n$. If it is desirable to indicate the number of vertices specifically, we call it an n-gon. It is often convenient to assume that the polygon is not self-intersecting i.e. the segments $[A_i \; A_{i+1}]$ are mutually disjoint (except for the trivial exception that adjacent ones have a common endpoint). More particularly, one often demands that the polygon be convex. If it is convex and the vertices A_i have coordinates (ξ_1^i, ξ_2^i), then as one sees from figure 1, its area is given by the formula

$$A = \frac{1}{2} \sum_{i=1}^{n} (\xi_1^i \xi_2^{i+1} - \xi_2^i \xi_1^{i+1})$$

where we put $A_{n+1} = A_1$.

In fact, this formula holds for all polygons but the geometrical interpretation of the area is not quite so simple since one has to take the orientation of the associated triangles into consideration (figure 2). Things become particularly non-transparent in the case of self-intersecting polygons (figure 3).

In complex notation, the regular n-gon is that one with vertices

$$(1, \omega, \omega^2, \omega^3, \dots, \omega^{n-1})$$

where ω is the primitive n-th root of unity i.e. the complex number $e^{2\pi i/n}P$.

However, in several situations, the non-convex, self-intersecting polygons which are obtained by choosing other roots of unity are of interest. hence for any $d \in \{1, \dots, n\}$, we write $\{n|d\}$ for the polygon with vertices

$$(1, \omega^d, \omega^{2d}, \dots, \omega^{d(n-1)}).$$

In fact, the only interesting case is where d and n are relatively prime (otherwise we get an m-gon where m is the quotient of n by the greatest common denominator of n and d and some of the vertices are repeated).

Probably the most famous example of such a polygon is the pentagram $\{5|2\}$ (figure 4).

Transformations of polygons We saw above that the midpoints of the sides of a quadrilateral always span a parallelogram. There are a number of related results which can be easily proved by using the arithmetic of complex numbers. We bring a few of thee:

I. Let $A_1, \dots, A_{2n}$ be the vertices of the $2n$-gon $\mathbf{P}$ and let $\mathbf{P}'$ be the polygon whose vertices are the midpoints of the sides of $\mathbf{P}$. Then the two sequences formed by taking every second side of $\mathbf{P}'$ form the edges of closed polygons. (This condition means that the sum of the vectors corresponding to the sides is zero so that they "join up" when placed end to end). (figure 5).

For suppose that the coordinates of the vertices of $\mathbf{P}$ are $z_1, \dots, z_{2n}$. Then those of $\mathbf{P}'$ are

$$\frac{1}{2}(z_1 + z_2) \dots \frac{1}{2}(z_{2n} + z_1)$$

and the complex numbers which represent the sides form the two series

$$\frac{1}{2}(z_3 - x_1), \frac{1}{2}(z_5 - z_3), \dots, \frac{1}{2}(z_1 - z_{2n-1})$$

resp.

$$\frac{1}{2}(z_4 - z_2), \dots, \frac{1}{2}(z_2 - z_{2n})$$

from which the result can be read off at a glance.

II. We define a $2n$-**parallelogram** to be a $2n$-gon so that the opposite sides are equal and parallel (figure 6). *in fact, if we number the vertices in the usual way, then the vector describing the side A_1A_2 say will point in the direction opposite to that of $A_{n+1}A_{n+2}$ as one can see from figure ??).

The result which we shall prove is the following:
Let $\mathbf{P}$ be a hexagon. Then the polygon $\mathbf{P}'$ whose vertices are the centroids of the triangles $A_1A_2A_3$, $A_2A_3A_4$ etc. is a 6-parallelogram.
PROOF. The vertices of $\mathbf{P}'$ are

$$\frac{1}{3}(z_1+z_2+z_3), \quad \frac{1}{3}(z_2+z_3+z_4), \ldots, \quad \frac{1}{3}(z_6+z_1+z_2)$$

and $A_1'A_2'$ is represented by the complex number $\frac{1}{3}(z_4-z_1)$ whereas $a_4'A_5'$ is represented by $\frac{1}{3}(z_1-z_4)$.

These results can be fitted into the following general framework. If the n-gon is determined by the vertices $A_1, \ldots, A_n$ with complex coordinates $z_1, \ldots, z_n$, then it can be described by the complex n-vector

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}.$$

If A is an $n \times n$ matrix which can be complex but in our applications will always be real, then it induces a transformation of polygons as follows: the column vector AZ can be regarded as the sequence of vertices of a new polygon (which we denote by $A(\mathbf{P})$ where $\mathbf{P}$ is the original polygon). For example, the transformations considered above are induced by the matrices $\operatorname{circ}(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$ resp. $\operatorname{circ}(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0)$.

This observation allows one to employ the spectral theorem for normal matrices to give some perhaps surprising proofs of geometrical results. We illustrate this method with the following example.

Let $\mathbf{P}$ be a polygon and consider the sequence

$$A\mathbf{P}, A^2\mathbf{P}, A^3\mathbf{P}, \ldots$$

where $A\mathbf{P}$ is the polygon with the midpoints of the sides of $\mathbf{P}$ as vertices. We call this the **successor** of $\mathbf{P}$. Our problem is to determine the nature of this sequence. A little experimentation on the part of the reader will persuade him that there is a tendency for the polygons to converge to a convex one.

however, there are exceptions as we shall see below. Before proceeding to a general analysis, we dispose of the cases $n = 2, 3, 4$ which are exceptional.

A 2-gon is a line and its successor is its midpoint. The sequence is stable thereafter.
In the case of a triangle, we get a succession of similar triangles as in figure ??.
In the case of a quadrilateral, we know that the first successor is a parallelogram. After this, the series oscillates between two series of similar parallelograms (figure ??).

In order to treat the general case, we consider the spectral analysis of A. We know that $A = p(C)$ where C is the standard circular matrix $\operatorname{circ}(0, 1, 0, \dots, 0)$ and p is the polynomial

$$t \mapsto \frac{1}{2}(1 + t).$$

hence the eigenvalues of A are the column matrices

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \begin{bmatrix} \omega \\ \omega^2 \\ \vdots \\ \omega^n \end{bmatrix} \quad \begin{bmatrix} \omega^2 \\ \omega^4 \\ \vdots \\ \omega^{2n} \end{bmatrix} \dots \begin{bmatrix} \omega^{n-1} \\ \omega^{2(n-1)} \\ \vdots \\ \omega^{n(n-1)} \end{bmatrix}$$

where $\omega = e^{2\pi i/n}$ is the primitive n-th root of unity.

We shall denote the above vectors by $Z_0, Z_1, \dots, Z_{n-1}$. The corresponding eigenvalues are $\lambda_0, \lambda_1, \dots, \lambda_n - 1$ where

$$\lambda_i = \frac{1}{2} + \frac{\omega^i}{2}.$$

If we interpret the eigenvectors as polygons, we see that they represent a sequence of "regular" polygons which are stable (up to similarity) with respect to A. We use the quotation marks to indicate that these polygons are not all convex and include the degenerate case and self-intersecting polygons mentioned above. For example, for $n = 6$, we have the following cases.

We now use spectral theory to reduce the general case to a suitable combination of the above regular polygons. let Z be the column matrix representing the given polygon. Then Z has an eigenfunction expansion

$$\sum_{k=0}^{n-1} a_k Z_k$$

where the Z_k's are the eigenvectors of A described explicitly above and a_k is the scalar product

$$\frac{1}{n}(Z|Z_k) = \frac{1}{n}\sum_{i=1}^{n} z_i\omega^{k(i-1)}.$$

These coefficients have the following geometrical interpretation:
For $k = 0$,

$$a_0 = \frac{1}{n}\sum_{k=1}^{n} z_k\omega^{-k+1}$$

which is just the centroid of the polygon with vertices

$$z_1,$$

i.e. the polygon obtained by "unwinding" $\mathbf{P}$ i.e. rotating the i-th vertix through an angle of $2\pi i/n$ in the clockwise direction.

The higher coefficients have a similar interpretation.

Note that in the expansion $Z = \sum_{k=0}^{n-1} a_k Z_k$, multiplication by a_k transforms the polygon with vertices $(z_1, \dots, z_n)$ into the one with vertices $(a_k z_1, \dots, a_k z_n)$.

We remark that the sum of two such polygons with vertices $(z_1, \dots, z_n)$ and $(w_1, \dots, w_n)$ is simply the polygon whose vertices are the sums $z_i + w_i$ of the corresponding vertices of the original ones.

Now if we apply A to Z in its eigenvector expansion, we get

$$AZ = \sum_{n=0}^{n-1} a_k\lambda_k Z_k$$

and, more generally,

$$A^p Z = \sum_{k=0}^{n-1} a_k\lambda_k^p Z_k.$$

This allows us to make the following deductions concerning the nature of the successors of $\mathbf{P}$, whereby we assume for the sake of simplicity that $a_0 = 0$ i.e. the centre of gravity of $\mathbf{P}$ is at the origin.

Note first that the two eigenvalues λ_1 and λ_{n-1} are equal in absolute value and larger than the other ones., hence if we normalise the successive polygons by multiplying them by $|\lambda_1|^{-1}$, then we get the sequence

$$a_1 e^{\pi i/n} Z_1 + a_{n-1}e^{-\pi i/n} Z_{n-1} + \sum_{k=2}^{n-2} a_k \frac{\lambda_k^p}{\lambda_1^p} Z_k.$$

All of the terms except the first two tend to zero and so in the limit the dominating term are

$$a_1 e^{\pi i/n} Z_1 + a_{n-1} e^{-\pi i/n} Z_{n-1}.$$

This is the image of the standard regular n-gon under the mapping

$$z \mapsto a_1 e^{\pi i/n} z + a_{n-1} e^{-o\pi i/n} z.$$

This is an example of an affine mapping. Mappings of this type will be examined in detail in the next section. Here we shall require only the following simple facts which can be checked by simple computations. The mapping

$$z \mapsto az + b\bar{z}$$

of the complex plane, when reduced to real form, is the mapping of multiplication by a matrix whose determinant is $|a|^2 - |b|^2$. if we apply this to the above mapping, we see that the dominating terms of $A^p Z$ are the images of the standard regular n-gon under an affine mapping and so lie on an ellipse if $|a_1| =\neq |a_{n-1}$.

Another natural question on successors which can be solved by the above method is the following: when is every polygon a successor resp. which polygons are successors? The most interesting case is that of a matrix of the form

$$A = \operatorname{circ} \frac{1}{d}(1, \ldots, 1, 0, \ldots, 0)$$

where we have d "ones".

For example, if $n = 4$ and $d = 2$, we have seen that only parallelograms are successors. it is not hard to show that every parallelogram is indeed a successor. if we return to the general case, the question whether each polygon is a successor is the question whether A is invertible. In the above case, $A = p(C)$ with p the polynomial

$$t \mapsto \frac{1}{d}(1 + t + \ldots t^{d-1})$$

and so the eigenvalues of A are

$$\lambda_k = \frac{1}{d}(1 + \omega^k + \cdots + \omega^{k(d-1)}) \qquad (k = 0, \ldots, n-1).$$

This leads to the following result:
If d and n are mutually prime, then every polygon is a successor. In the case where A is not invertible, we are interested in a characterisation of its image. Since A is normal (in the above special cases), its image is the orthogonal complement of its kernel and this can often be used to give an explicit description of the successors. We illustrate this in the case $d = 2$. Then we see that ???????

2.5 Circles

The circle with centre x_0 and radius r has equation

$$\|x - x_0\|^2 = r^2$$

which can be rewritten in the form

$$\xi_1^2 + \xi_2^2 - 2a\xi_1 - 2b\xi_2 + c = 0$$

where $x_0 = (a, b)$ and $r^2 = a^2 - b^2 - c$.

We denote this circle by C (or $C_{a,b,c}$ if we want to specify the parameters). It is convenient to use the notation S_C for the function

$$x \mapsto \xi_1^2 + \xi_2^2 - 2a\xi_1 - 2b\xi_2 + c.$$

If x lies outside of the circle, this is positive and can be interpreted as the square of the length of the tangent from x to C, or more generally as the product $|PQ||PR|$ with Q and R as in figure 1.

If x is inside the circle, it is once again the product $|PQ||PR|$, this time with negative sign which reflects the fact that PQ and PR point in opposite directions (figure 2).

These facts can be proved algebraically as follows. Consider the line $\{x + tn : t \in \mathbf{R}\}$ where n is a direction i.e. a unit vector. This cuts the circle in two points (in general) which correspond to the roots of the quadratic equation $S_C(x + tn) = 0$ in t. The product of these two roots is the value of the quadratic at 0 and this is just $S_C(x)$. But this product is the product of the (signed) lengths of the segments from the point x to the two points of intersection with the circle. (Note that we have proved that it is independent of the direction n, in particular that $|PT|^2 = |PQ||PR|$ with P, Q, R and T as in the figures).

$S_C(x)$ is called the **power** of x with respect to C. If C and C_1 are non-concentric circles, then the locus of points whose powers with respect to the two circles coincide i.e. the set

$$\{x : S_C(x) = S_{C_1}(x)\}$$

is a straight line, in fact the line

$$2(a - a_1)\xi_1 + 2(b - b_1)\xi_2 + (c - c_1) = 0.$$

If C and C_1 are external to each other, this line is the locus of the set of points for which the lengths of the tangents to the two circles coincide. If C and C_1 intersect at two points, it is the line through these two points. If C and C_1 touch each other (either externally or internally), it is their mutual tangent. In any of these cases, it is called the **radical axis** of the two circles.

If C_1, C_2 and C_3 are circles, no two of which are concentric, then the three radical axes are either parallel or concurrent. In the former case, the point of intersection is called the **power point** of the triple. To see this note that if (a_1, b_1, c_1), (a_2, b_2, c_2) and (a_3, b_3, c_3) are the parameters of the circles, then the equations of the axes are

$$2(a_1 - a_2)\xi_1 + 2(b_1 - b_2)\xi_2 + c_1 - c_2 = 0 \tag{151}$$
$$2(a_2 - a_3)\xi_1 + 2(b_2 - b_3)\xi_2 + c_2 - c_3 = 0 \tag{152}$$
$$2(a_3 - a_1)\xi_1 + 2(b_3 - b_1)\xi_2 + c_3 - c_1 = 0. \tag{153}$$

We now use the simple fact (which will be proved later), that three lines

$$A_1\xi_1 + B_1\xi_2 + c_1 = 0 \tag{154}$$
$$A_2\xi_1 + B_2\xi_2 + C_2 = 0 \tag{155}$$
$$A_3\xi_1 - B_3\xi_2 + C_3 = 0 \tag{156}$$

in $\mathbf{R}^2$ are concurrent or parallel if and only if the corresponding vectors

$$(A_1, C_1, C_1), \quad (A_2, B_2, C_2), \quad (A_3, B_3, C_3)$$

are linearly dependent in $\mathbf{R}^3$. This is clearly the case for the three radical axes (the syum of the three vectors is zero).

Suppose that C and C_1 are circles with parameters (a, b, c) resp. (a_1, b_1, c_1). Then for each real l we define a function

$$S_\lambda(x) = \lambda S_C(x) + (1 - \lambda)S_{C_1}(x).$$

of course, S_λ is the power function of a circle, in fact the circle C_λ is parametrised by by

$$\lambda(a, b, c) + (1 - \lambda)(a_1, b_1, c_1).$$

The system $\{C_\lambda\}$ is called a **pencil of circles.** It is also sometimes called a coaxial system for the following reason. If C_{λ_1} and C_{λ_2} are two distinct circles from the pencil, then their radical axis has equation

$$\lambda_1 S_C(x) + (1 - \lambda_1)S_{C_1}(x) = \lambda_2 S_C(x) + (1 - \lambda_2)S_{C_1}(x)$$

which simplifies to

$$(\lambda_1 - \lambda_2)S_C(x) = (\lambda_1 - \lambda_2)S_{C_1}(x)$$

i.e. is the radical axis of C and C_1. In other words, any two pairs of the pencil have the same radical axis (figure ??).

Note that if C and C_1 intersect at two points, then each C also passes through these two points and so the pencil consists of the family of all circles which pass through these points.

We can give the following more geometrical interpretation of the family $\{C_\lambda\}$. We have

$$C_\lambda = \{x : S_\lambda(x) = 0\} \tag{157}$$

$$= \{x : \lambda S_C(x) + (1-\lambda)S_{C_1}(x) = 0\} \tag{158}$$

$$= \{x : \frac{S_C(x)}{S_{C_1}(x)} = \frac{\lambda - 1}{\lambda}\}. \tag{159}$$

Hence C_λ can be regarded as the locus of points for which the ratio of their powers with respect to C and C_1 is constant (the value of this constant being $\frac{\lambda-1}{\lambda}$).

If C_1 and C_2 are two circles which intersect at the point x, then the angle θ between the circles at this point is given by the formula

$$2r_1r_2\cos\theta = 2a_1a_2 + 2b_1b_2 - c_1 - c_2.$$

In particular, the circles are orthogonal to each other if and only if

$$2a_1a_2 + 2b_1b_2 - c_1 - c_2 = 0.$$

From this it follows easily that if a circle C is orthogonal to C_1 and C_2, then it is also orthogonal to any circle in the coaxial system that they generate.,

Conversely if C_1 and C_2 are non-intersecting circles, then there are two points through which each circle which is orthogonal to C_1 and C_2 passes. In other words, the family of circles orthogonal to C_1 and C_2 is the coaxial system determined by these two points. We prove this fact analytically. In order to simplify the notation, we assume that the two circles have centres on the x-axis and that their radical axis is the y-axis. The reader can check that this means that their equations take on the special forms

$$\xi_1^2 - \xi_2^2 + 2a_1\xi_1 + c = 0$$

and

$$\xi_1^2 + \xi_2^2 + 2a_2\xi_1 + c = 0$$

whereby $c > 0$. (see figure ?).

Now suppose that the circles C_3, with parameters (a_3, b_3, c_3) is orthogonal to both of the above. This leads to the equations

$$2a_1a_3 = c + c_3 \tag{160}$$

$$2a_2a_3 = c + c_3. \tag{161}$$

Since $a_1 \neq a_2$ (we are tacitly assuming that the circles are not concentric), this means that $a_3 = 0$ and $c_3 = -c$ i.e. the circle has the form

$$\xi_1^2 + \xi_2^2 + 2b_3\xi_2 - c = 0.$$

Then the centre lies on the y-axis (i.e. on the radical axis of C_1 and C_2) and the circle cuts the x-axis in the two points $(\sqrt{c}, 0)$ and $(-\sqrt{c}, 0)$. These are the two points we are looking for.

2.6 Affine mappings

We now consider topics in geometry which involve transformations of space. Given the privileged position of straight lines in elementary geometry, it is natural to begin with a discussion of those mappings which take lines into lines. if we consider the nature of the analytic description

$$\{x : a\xi_1 + b\xi_2 + c = 0\}$$

of a line (where $a^2 + b^2 \neq 0$), then it is not surprising that mappings of the form

$$x = (\xi_1, \xi_2) \mapsto (a_{11} + a_{12} + c_1, a_{21}\xi_2 + a_{22}\xi_2 + c_2)$$

for suitable constants a_{11}, a_{12}, a_{21}, a_{22}, c_1, c_2 possess this property, whereby we assume that $a_{11}a_{22} - a_{12}a_{21} \neq 0$ to ensure that the mapping is a bijection.

In the language of matrices, this transformation can be written as

$$X \mapsto AX + C$$

where A is the 2×2 matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and C and X are the column matrices

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

and

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}.$$

The condition imposed on the a's means that A is invertible. The inverse of the mapping is then

$$Y \mapsto A^{-1}(Y - C) = A^{-1}Y - A^{-1}C$$

which is one of the same type.

Mappings of the above form are called **affine** and A is called **the matrix of the mapping**. The simplest affine mappings are those whose matrices are the identity i.e. those of the form

$$x \mapsto x + u$$

where $u = (c_1, c_2)$. Of course this is just a translation by u and we denote it by T_u. The other extreme is represented by those mappings which only have a matrix part i.e. are of the form

$$X \mapsto AX.$$

Of course these are just the *linear* mappings i.e. those mappings $f : \mathbf{R}^2 \to \mathbf{R}^2$ which are such that

$$f(x + y) = f(x) + f(y) \quad f(\lambda x) = \lambda f(x)$$

$(x, y \in \mathbf{R}^2, \lambda \in \mathbf{R})$.

An affine mapping f is linear if and only if $f(0) = 0$ and an arbitrary affine mapping f can be expressed as a product of a linear one and a translation. In fact $f = T_u \circ \tilde{f}$ where $u = f(0)$ and $\tilde{f}$ is the linear mapping with the same matrix as f.

In working with affine mappings, it is useful to have an expression for their composition. If the mappings f and g are linear, then of course their composition $g \circ f$ is the linear mapping with matrix BA where A is the matrix of f and B that of g. This can be computed directly or one can simply note that in matrix form f is left multiplication by A, which is then followed by left multiplication by B. For general affine mappings f and g, we write them in the form $T_u \circ \tilde{f}$ and $T_v \circ \tilde{g}$ as above. Then $g \circ f$ is the mapping $T_{g(u)+v} \circ \tilde{g} \circ \tilde{f}$ as can be checked easily.

The following types of affine mapping are among the most important in elementary geometry:

D_θ: This is the mapping of rotation through an angle θ about the origin as axis and is defined to be the linear mapping with matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Using D_θ we can define the operator $D_{x,\theta}$ of rotation through an angle θ around the point x by the formula

$$D_{x,\theta} = T_x \circ D_\theta \circ T_{-x} = T_{x-D_\theta(x)} \circ D_\theta.$$

R_L: This is the operation of reflection in the line L which passes through the origin. It is the linear mapping with matrix

$$\begin{bmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{bmatrix}$$

where ϕ is the angle between L and the x-axis. More generally, if L is a line in the plane, then the operator R_L of reflection in L is defined to be the composition $T_x \circ R_{L_0} \circ T_{-x}$ where L_0 is the line through the origin parallel to L and x is any point on L (the reader can check that this mapping is independent of x).

Glide-reflections: These are mappings of the form $T_u \circ R_L$ where L is a line and u is a (non-zero) vector, parallel to L.

The above mappings (together with the translations) are all examples of **congruences** (or **isometries** as they are sometimes called). These terms simply means that the mappings preserve distance i.e. are such that $\|f(x)-f(y)\| = \|x-y\|$ for each $x, y \in \mathbf{R}^2$.

In fact, the isometries of the plane are precisely those listed above i.e. translations, rotations, glide reflections and reflections. In addition, we have the following rules for the composition of these isometries:

- rotation plus rotation = rotation or translation (the latter if the sum of the angles of rotation is a multiple of 2ϕ);
- translation plus rotation = rotation;
- reflection plus reflection = rotation or translation (the latter if the two lines of reflection are parallel);
- translation plus reflection = glide reflection.

- rotation plus reflection = glide reflection or reflection (if the axis of rotation lies on the line of reflection).

We now proceed to discuss a further class of mappings—the so-called **similarities**. These are characterised by the condition

$$\|f(x) - f(y)\| = \lambda \|x - y\|$$

where λ is a positive scalar (more explicitly, such an operator is called a λ-similarity).

Examples: Dilations are mappings of the form

$$\mathrm{Dil}_{x,\lambda} = T_x \circ \lambda \mathrm{Id} \circ T_{-x}.$$

In particular, linear dilations are mappings of the form $x \mapsto \lambda x$ $(\lambda \neq 0)$ (i.e $\lambda \mathrm{Id}$).

Spiral rotations: These are mappings of the form

$$SR_{x,\theta,\lambda} = T_x \circ \lambda D_\theta \circ T_{-x}.$$

Dilatory reflections: These are mappings of the form

$$DR_{L,\lambda,x} = T_x \circ \lambda R_{L_0} \circ T_{-x}$$

where x is a point on L and L_0 is the line through 0, parallel to L.

Using the classification of isometries mentioned above it is easy to describe all k-similarities (whereby we shall assume that $k \neq 1$ i.e. we are excluding the isometries). We claim firstly that any such operator has a fixed point i.e. an x so that $f(x) = x$. For in this case, $\frac{1}{k} f$ is an isometry and so has the form $T_u \circ g$ where g is a linear isometry. Then the fixed point equation has the form

$$kg(x) + ku = x$$

i.e.

$$(\mathrm{Id} - kf)x = ku.$$

Hence it suffices to show that $(I - kA)$ is invertible, where A is the matrix of g. But this follows easily form the fact that the eigenvalues of A (as an orthonormal matrix) all have absolute value one. In particular, $\frac{1}{k}$ is not an eigenvalue.,

From this it follows that if x is the fixed point of f and we consider the mapping $\tilde{f} = \frac{1}{k}(T_{-x} \circ f \circ T_x)$, then $\tilde{f}$ is a linear isometry and so either a rotation or a reflection. Then $f = T_x \circ k\tilde{f} \circ T_{-x}$ is either a spiral symmetry or a dilatory reflection. The dilations occur as special cases of spiral symmetries with $\theta = 0$ or $\theta = \pi$. The former correspond to dilations with positive factor λ (the **direct** dilations), the latter to negative λ (the **indirect ones**).

We now consider the possible forms of the composition of two dilations. Firstly it is clear that the composition

$$\mathrm{Dil}_{x,\lambda} \circ \mathrm{Dil}_{x,\mu}$$

of two dilations with the same centre is the dilation $\mathrm{Dil}_{x,\lambda\mu}$.

In the case where the centres are distinct i.e. a product of the form $\mathrm{Dil}_{x,\lambda} \circ \mathrm{Dil}_{y,\mu}$ where $x \neq y$, we distinguish two cases:
Case 1: $\lambda\mu \neq 1$. We shall assume that $x = 0$ to simplify the calculations. Then the operator reduces to

$$\lambda \mathrm{Id} \circ T_y \circ \mu \mathrm{Id} \circ T_{-y}$$

which is clearly a dilation. Its centre is the fixed point of the mapping i.e. the solution x_0 of the equation

$$\lambda\mu(x_0 - y) + y = x_0$$

which is $x_0 = \frac{\lambda(1-\mu)}{1-\lambda\mu} y$. Hence the composition is a dilation and the centre of dilation lies on the segment between x and y.

As an example of a simple result involving dilations, consider the following fact which can be proved elegantly using dilations:
let $ABCD$ be a quadrilateral, A_1 the centroid of BCD, B_1 that of CDA etc. (figure ??). Then the quadrilateral $A_1B_1C_1D_2$ is similar to $ABCD$. For we can clearly suppose that the centroid of the original quadrilateral is at the origin i.e.

$$x_A + x_B + x_C + x_D = 0.$$

Then we have

$$x_{A_1} = \frac{1}{3}(x_B + x_C + x_D) = -\frac{1}{3} x_P A$$

etc. and so $\mathrm{Dil}_{0,-\frac{1}{3}}$ maps $ABCD$ onto $A_1B_1C_1D_1$.

If C_1 and C_2 are circles with distinct centres O_1 and O_2 and distinct radii r_1 and r_2, then there are two dilations (a direct and an indirect one) which map C_1 onto C_2. These can be found as follows. On the line O_1O_2 we locate points P and Q with barycentric coordinates

$$x_Q = \frac{-r_2}{r_1 - r_2} x_{O_1} + \frac{r_1}{r_1 - r_2} x_{O_2}$$

$$x_P = \frac{r_2}{r_1 + r_2} x_{O_1} + \frac{r_1}{r_1 + r_2} x_{O_2}$$

(we are assuming that $r_1 > r_2$) (see figure ??).

Then it is clear that Dilx_Q, $\frac{r_1}{r_2}$ resp. Dil$_{x_P,-\frac{r_1}{r_2}}$ map C_2 onto C_1. The points P and Q are called the **centres of similitude** (internal and external) of C_1 and C_2. Note that they are harmonic conjugates with respect to the two centres.

Perhaps the most famous example of this situation is the following: if ABC is a triangle, then the orthocentre is the centre of external similitude between the nine-point circle and the circumcircle while the centroid is the centre of internal similitude. In fact, the mapping Dil$_{x_H,\frac{1}{2}}$ and Dil$_{x_M,-\frac{1}{2}}$ map the former onto the latter (figure ??).

We close our remarks on dilations with some simple problems which can be solved by their use:

I. A fixed point P on a circle is given. Determine the locus of the midpoints of the chords PQ as Q varies on the circle (figure ??). It is clear that the locus is the image of the circle C under the dilation Dil$_{x_P,\frac{1}{2}}$ and so is itself a circle. This circle is tangential to the original one at the point P.

II. An acute angled triangle ABC is given. We are required to construct a square $PQRS$ with base PQ on BC and vertices R on AC and S on AB (figure ??). We begin by constructing the square on BC as in the figure and note that the required square is its image under a dilation with centre at the foot of the perpendicular from A to BC.

The theorem of Desargues: The algebra of dilations can be used to prove one of the most famous results of elementary geometry—the theorem mentioned in the paragraph heading. Consider figure ???. here ABC and $A_1B_1C_1$ are triangles which are such that the lines AA_1, BB_1 and CC_1 are concurrent. Then the theorem states that the points of intersection of AB and A_1B_1 resp. of BC and B_1C_1 resp. of CA and C_1A_1 are collinear. We prove this as follows:

Spiral similarities
We bring an application of mappings of this type to an elementary geometrical proposition. Two circles C_1 and C_2 touch externally at P and a line through P meets C_1 at A and C_2 at B. Show that the tangent to C_1 at A is parallel to the tangent at B to C_2.

To prove this we need only remark that if the radius of C_1 is k times that of C_2, then

2.7 Applications of isometries:

In the following pages, we would like to try to demonstrate the fundamental importance of congruences in elementary geometry by showing how they can be used to define basic concepts and solve elegantly classical problems on constructions and loci

Translations We begin with translations. Among the concepts which can be defined using translations are parallelism and parallelograms. Two lines are parallel if and only if they can be mapped onto each other by means of a translation. Similarly, the quadrilateral $ABCD$ is a parallelogram if and only if there is a translation which maps A onto D and B onto C (figure 1).

We now turn to two constructions which can be carried out by using translations:

I. Given a triangle ABC and a line segment L. We are required to construct a line which is parallel to the given segment and which cuts the triangle as in figure 2 so that the length of the segment cut off is equal to that of L.

We solve this as follows: the segment L is first translated so that one endpoint is at A. (figure 3). We now translate it so that this endpoint moves along AB. Then the other endpoint traces a line parallel to AB which we can construct. The point where this line intersects AC is one of the endpoints of the required segment—the other is obtained by completing the parallelogram as in figure 4.

II. Two circles C_1 and C_2 are given, together with a line segment L. We are required to find points A (on C_1) and B (on C_2) so that AB is equal and parallel to L (figure 5).

In order to do this, we translate C_1 along the vector determined by L. One of the points of intersection of C_2 with the translated circle is an endpoint of the required segment.

The following diagram illustrates a simple proof of the theorem of Pythagoras, using translations.

We now bring an example of a problem on loci which can be solved easily by a judicious use of translations:

]We are given two intersecting lines L and L_1 and a number a. We are then required to determine the locus of those points for which the sum of the distances fro L and L_1 is a. before solving this problem, we shall make a little more precise. We choose unit normal vectors n and n_1 to L and L_1 so that the equations of these lines are

$$(x - x_0|n) = 0 \quad \text{resp.} \quad (x - x_0|n_1) = 0$$

where x_0 is the point of intersection. Then the distances $d(x, L)$ and $d(x, L_1)$ of a point x to these lines are given by the equations

$$d(x, L) = (x - x_0|n) \quad d(x, L_1) = (x - x_0|n_1).$$

Notice that there are *directed* distances i.e. they are positive or negative according as x lies on the same side of the line as the normal or on the opposite side. We now turn to our original problem and begin with the special case $a = 0$. Then the solution to this case is the bisector of one of the angles between the two lines. Which of the two angles is to be bisected is determined by the choice of normals (see figure ??).

For the general case, we construct the line $\{x : d(xL) = a\}$. This is parallel to L and is at a distance a from it. It lies on the same side as the normal if a is positive, otherwise it is on the opposite side (figure ??).

Let this line cut L_1 at P. Then the solution to our problem is the line through P, parallel to the bisector of the angle between L and L_1 as the reader can easily verify (figure ??).

Rotations We now discuss congruences of this type. many concepts can be described in terms of these mappings. For example, if A is the midpoint of the segment PR, then

$$\begin{aligned} D_{x_Q,\pi} \circ D_{x_P,\pi} &= T_{x_Q} \circ D_\pi \circ T_{-x_Q} \circ T_{x_P} \circ D_\pi \circ T_{-x_P} && (162) \\ &= T_{x_Q + x_Q - x_P - x_P} && (163) \\ &= T_{2x_{PQ}} = T_{x_{PR}}. && (164) \end{aligned}$$

This reasoning is reversible, so that we can characterise the fact that A is the midpoint of PR by the equation

$$D_{x_Q,\pi} \circ D_{x_P,\pi} = T_{x_{PR}}.$$

As a further example, consider the product

$$D_{x_R,\pi} \circ D_{x_Q,\pi} \circ D_{x_P,\pi}$$

of three half-turns. Of course, this is again a half-turn, the axis being the fixed point. If we simplify the above product to

$$T_{2(x_P - x_Q + x_R)} \circ D_\pi$$

then we can calculate the fixed point x_S as the solution of the equation

$$2(x_P - x_Q + x_R) - x_S = x_S$$

which just means that S is that point so that $PQRS$ is a parallelogram.

The fact that a quadrilateral $ABCD$ is a square can be expressed in the equation

$$x_{AD} = D_{\pm\frac{\pi}{2}} x_{AB}.$$

(see figure ??). Similarly, a triangle ABC is equilateral if and only if

$$x_{BA} = D_{\pm\frac{\pi}{3}} x_{BC}.$$

(figure ??).

A rather more important example is that of the angle between lines. If L and L_1 are two lines in the plane, then one can define the angle between L and L_1 to be θ where $\cos\theta = (n|n_1)$, n and n_1 being the unit normals to L and L_1. However, this fails to distinguish between the interior and exterior angles and for many purposes one requires the more subtle concept of a **directed angle** which is defined using rotations as follows: to begin with one defines a **ray**. if O and A are distinct points in the plane, the ray AA (written R_{AA}) is defined to be the set of points $\{x_) + tx_{AA} : t > 0\{$. (figure ??). The following simple property of rays can easily be verifies if

B is a point on the ray AA, then $r_{AA} = r_{OB}$;   two rays with the same initial points are either equal or disjoint.

Now if r_{AA} is a ray, then the set $D_{x_O,\theta} r_{AA}$ is also a ray (more generally, the image or a ray under any isometry is a ray). On the other hand, if R_{AA} and r_{OB} are two rays with the same initial points, then there is an angle θ so that $D_{x_O,\theta} r_{AA} = r_{OB}$. This θ is uniquely determined up to whole number multiples of 2π. it is called the (directed) angle from AA to OB, written $\angle AOB$.

As an exercise, the reader can try his hand at proving now that the sum of the angles of a triangle is 180^o. This means that if A, B and C are distinct points in the plane, then

$$\angle ABC + \angle BCA + \angle CAB = \pm\pi$$

(the sign being chosen according to the orientation of the triangle).

As an example of a construction which can be carried out using rotation consider the following:
Two lines L_1 and L_2 are given, together with a point A. We are required to construct an equilateral triangle ABC whereby B is on L_1 and C is on L_2 (figure ??).

To do this, we rotate L_2 through an angle of 60^o around the point A. Then the intersection of the rotated line with L_1 is the required point B.

There are a number of results related to a famous theorem ass associated with name of Napoleon which can be proved elegantly with the help of suitable rotations. We begin with the so-called Napoleon's theorem:

Let ABC be a triangle and construct on each side an equilateral triangle external to ABC as in figure ??. Let P, Q and R be the respective centroids of these triangles. Then PQR is equilateral.
PROOF. We prove this as follows. Consider the mapping $\sqrt{3}D_{x_C,\frac{\pi}{6}}$ and $\frac{1}{\sqrt{3}}D_{x_B,\frac{\pi}{6}}$. Then the first maps A into A and the second maps Q onto R. Hence their composition maps Q onto R. From its form, it is a rotation through an angle of $\frac{\pi}{3}$. The axis of rotation is its fixed point. But the reader can check that P is fixed by the mapping. This implies that PQR is equilateral. (The factor $\sqrt{3}$ in the above argument is the ratio of the length of a side of an equilateral triangle to the line from a vertex to the centre).

■

Consider now a quadrilateral $ABCD$ with equilateral triangles constructed on the sides, this time alternatively internal and external as in figure ??. Then

the quadrilateral $PQRS$ is a parallelogram, where P, Q, R and S are the centroids of the triangles.,
Proof. For consider the mapping

It is clear from its form that this is a translation and one can check that it maps P onto Q and S onto R.

We now consider two constructions which can be carried out by using rotations:
I. We are given a triangle ABC and a point R on AB and are required to find points S and T as in figure ?? so that the triangle RST is equilateral.

This is done by rotating AB around R through an angle of 60^o. We denote the intersection of the rotated line with BC by T. S is the pre-image of T under this rotation (figure ??).
II. We are given a circle C, a point P in its interior and a segment AB and are required to find a chord through P which has the sam length as AB (figure ??).

We begin by constructing a chord of the circle with the same length as AB (figure ??). We now rotate P around the centre of the circle until it lies on the chord, then the chord constructed above through the same angle, but in the opposite direction (figure ??).

Pseudo-squares A square has the characteristic property that its diagonals are of the same length and cross each other at right angles in their respective midpoints. if we drop the very last condition in this statement, we arrive at the concept of a **pseudo-square** i.e. a quadrilateral $ABCD$ so that $x_{BD} = D_{\pm\frac{\pi}{2}}]x_{AB}$. In order to avoid irritating ambiguities we shall assume that pseudo-squares are labelled in such a way that we have the plus sign in the above formula (see figure ??).

We shall show that this definition is equivalent to the fact that there exists a point F so that AFD and BFC are right-angled isosceles triangles (then, by symmetry, there exists a G so that BGA and CGD have the same properties (figure ??).

For suppose that $x_{BD} = D_{\frac{\pi}{2}} x_{AC}$ and let F be the point so that $D_{x_F,\frac{\pi}{2}}$ maps B onto C. Then this mapping sends the segment BD onto a segment parallel to CA. Since the endpoint B is mapped onto C, it follows that D is mapped onto A i.e. AFD is a right-angled isosceles triangle. On the other hand, if there is an F so that $D_{x_f,\frac{\pi}{2}}$ maps B onto C and D onto A, then this rotation maps BD onto CA and so $ACBD$ is a pseudo-square.

We shall use these concepts to prove some simple results on pseudo-squares:

Proposition 35 *If $ABCD$ is a quadrilateral and squares are erected externally on its sides, with centres P, Q, QR and S, then $PQRS$ is a pseudo-square.*

Proof. We begin with a preliminary result: ABC is a triangle and on the sides AB and AC we construct squares external to ABC, with centres P and Q. Then PMQ is a right-angled, isosceles triangle, where M is the midpoint of BC (figure ??).

For consider the operator

$$D_{x_m,\pi} \circ D_{x_Q,\frac{\pi}{2}} \circ D_{x_P,\frac{\pi}{2}}.$$

As a product of three rotations with the sum of the angles equal to 2π, this is either a translation or the identity. Consider the successive images of P under these mappings. Under the first one it is invariant. Second one sends it to the point P' whereby PQP' is a right-angled isosceles triangle. Now $D_{x_M,\pi}$ sends P' to P (since the composition is the identity) and this means that M is the midpoint of PP'. The result is an immediate consequence.

We now turn to the result that $PQRS$ is a pseudo-square i.e. that PR and QS are equal in length and perpendicular. Let M be the midpoint of AC. Then, by the above, the mapping $D_{x_M,\frac{\pi}{2}}$ maps P onto Q and R onto S. ∎

Reflections We now turn to applications of reflections. It may be rather surprising that the operation of reflection is one of the most basic in geometry. The reason for this is the fact that the reflections in the plane generate the family of isometries in the sense that an arbitrary isometry can be expressed as a product of reflections. For example, the reader will be aware of the fact that the product of two reflections in parallel lines is a translation and it is clear that every translation can be obtained in this way. Similarly, the product of two reflections in intersecting lines is a rotation with the point of intersection as axis. Likewise, every rotation can be generated in this way. Finally, a glide reflection is a product of a translation and a reflection and so of three reflections. This fact implies that every geometrical notion which can be expressed in terms of isometries can also be described by reflections so that the latter can be taken as the basis of euclidean geometry. We illustrate this with some simple examples:

A parallelogram $ABCD$ is a rhombus if and only if it is mapped onto itself by the isometry R_L where L is the line through A and C (figure ??).

A triangle ABC is isosceles if it is mapped onto itself by the operator R_L where L is the straight line through A and midpoint of BC (figure ??).

As an example of the use of reflections to give instant solutions to simple geometrical problems, consider the following: a straight line L is given, together with two points A and B on the same side of L. We are required to determine the shortest path (not necessarily composed of straight line segments) from A to B which also passes through a point on the line L. We can solve this problem as follows. Let B' be the reflection of B in L. Clearly, any solution of the above problem corresponds to a path from A to B' (we simply reflect that part of the path which is on the other side of L). Of course the shortest path from A to B' is a straight line and this leads to the solution of the original problem (figure ??).

A similar solution to a well-known problem is as follows; A, B and L are given as above. The problem is to construct a point P on L so that the angles which AP and BP make with the line L are equal. (figure ??). Once again, we reflect B in L to the point B' and take for P the point of intersection of L with the line AB. (figure ??).

Further problems of construction which can be solved with the aid of reflections are as follows:
I. Given are three lines L_1, L_2 and L_3 which intersect at the point O. Construct a triangle with these lines as bisectors (figure ??). [This can be done as follows: choose a point A on L_1 and let A' and A" be the refections of A in L_2 and L_3 respectively. Then A' lies on the line BC and so A" lies on AB. We now have two points on the line AB and this allows us to construct the latter. The rest is easy (figure ??).
II. Tow circles C_1 and C_2, on the opposite sides of a line L are given. We are required to construct a square $ABCD$ with diagonal AC on the line and with vertices B and D on C_1 and C_2 respectively (figure ??).

To do this, we reflect C_1 in L and let D be a point of intersection of the reflected circle with C_2 resp. B be the reflection of D in L. Then the square on BD as diagonal clearly has the required properties. (We remark here that the above construction will not always be possible. This will be displayed by the fact that the reflection of C_1 will fail to intersect C_2. Similar remarks apply to other constructions considered in this chapter).

We consider one last example of a classical problem which can be solved using reflections. let ABC be a triangle and let P, Q and R be the feet of the perpendiculars (figure ??). We begin with the remark that the lines AP, BQ and CR are the bisectors of the angles of PQR (and so H is the incentre of the latter triangle). This can be seen from the the figure where the marked angles are equal due to the fact that the quadrilaterals $ARPC$ and $ARHQ$ are cyclic. Note that this means that the perimeter of PQR is a possible path of a ray of light which is reflected from the sides of the original triangle. This makes the following result quite plausible:

Proposition 36 *The triangle PQR is that triangle of minimal perimeter amongst all triangles $P'Q'R'$ with P' on BC, Q' on BC and R' on AB.*

Proof. We prove this by use of reflections as follows: consider figure ?? which is obtained by reflecting ABC successively in its sides (each side being taken twice). Then the side of PQR are reflected onto six segments of the straight line from P to P_1, whereas the sides of any other inscribed triangle form a broken line which necessarily is longer.

2.8 Further transformations

There are also examples of interesting transformations which are not affine. The simplest of these is that of inversion in a circle which is defined as follows: the inversion of the point P in the circle with centre O and radius r is that point on the ray OP so that $|OP||OQ| = |OA|^2$ (A is the point where the ray OP cuts the circle–see figure 1).

In coordinates we have

$$x_Q = x_O + \frac{r^2}{\|x_{OP}\|^2} x_{OP}.$$

For the unit circle, with centre at the origin, this takes on the simple form $x_Q = \frac{x_P}{\|x_P\|^2}$ or, in coordinates

$$(\frac{\xi_1}{\xi_1^2 + \xi_2^2}, \frac{\xi_1}{\xi_1^2 + \xi_2^2}).$$

In terms of complex coordinates, inversion in the unit circle has the form $z \mapsto \frac{z}{|z|^2}$.

The inversion of a point P in a circle C can be constructed with compasses as follows (see figure 2). We draw the circle with centre P which passes through O, the centre of the original circle. Let it cut the latter in the points

A and A'. We then draw circles with centres A and A' which also pass through O. Their second point of intersection A is the required inversion as the reader can verify.

At this point, we note that the inversion of a circle in a second circle is a circle or a straight line (depending on whether the first circle passes through the centre of the inverting one or not). We shall prove this shortly. A circle C_1 is invariant under inversion in a second one C if and only if it is orthogonal to C. For suppose that C_1 inverts into itself. Then it follows from the fact that inversion is angle-preserving (see below) that the angles PTR and RTQ (figure 3) coincide and so both are right angles.

On the other hand, suppose that C is orthogonal to C_1. Let the line OO_1 joining their centres meet C_1 at P and Q and let T be one of the points of intersection of the two circles (figure 4). Then by the orthogonality condition OT is tangential to C_1 and so $|OP||OQ| = |OT|^2$. This just means that P is the inversion of Q in C. Now the image of C_1 under inversion is the circle through P and the two points of intersection, T and its opposite., which of course are mapped onto themselves by inversion. In other words, the image of C_1 is C_1 itself (since a circle is determined by three points).

Rather than pursue the special transformation of inversion we shall consider a wider class–the so-called **Möbius transformations** which include all of the transformations which we have considered un till now (more precisely, all of the *orientation-preserving* transformations. Möbius transformations are, by definition, mappings of the complex plane of the form

$$z \mapsto w = \frac{az+b}{cz+d}$$

where a,b,c,d are complex numbers with $ab - cd \neq 0$. In order to avoid difficulties for the case where the denominator vanishes, it is convenient to regard there as mappings on the extended complex plane $\bar{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$ (i.e. $\mathbf{C}$ with an ideal point at infinity added). Then the condition $ab - cd \neq 0$ means that the mapping is a bijection. In fact, its inverse is the Möbius transformation

$$z \mapsto \frac{dz - b}{-cz + a}$$

as the reader can easily check.

We note here that, due to the fact that the expression for the image of z is homogeneous, a Möbius transformation is unaffected if we multiply all of its parameters by the same (non-zero) complex number.

Among the transformations which can be accomodated into this scheme

are:
Translations. This is the case $a = 1$, $c = 0$, $d = 1$;
Dilations: Here a is real, $b = 0$, $c = 0$, $d = 1$;
Rotations: a is a complex number with $|a| = 1$, $b = 0$, $c = 0$ and $d = 1$.
Spiral rotations: a is non-zero, $b = c = 0$, $d = 1$.
Inversions: The mapping $z \mapsto \frac{1}{z}$ i.e. the Möbius transformation with $a = 0$, $b = 1$, $c = 1$ and $d = 0$ is (up to a reflection in the x-axis) inversion in the unit circle (figure 5).

In fact, any Möbius transformation can be broken down into a product of such simple ones as we shall now show. Firstly assume that $c = 0$. The d is non-zero and the transformation has the form

$$w = \frac{a}{d}z + \frac{b}{d}$$

which is the composition of the transformations $z \mapsto \frac{a}{d}z$ (a spiral rotation) and $z \mapsto z + \frac{b}{d}$ (a translation)).

If we suppose that c is non-zero, then we can write the transformation in the form

$$w = \frac{a}{c} + \frac{bc - ad}{c^2}\frac{1}{z + \frac{d}{c}}$$

which displays it is the following composition

$$z \mapsto z + \frac{d}{c} \quad \text{(a translation);} \tag{165}$$

$$z \mapsto \frac{1}{z} \quad \text{(inversion in a circle plus reflection in the x-axis);} \tag{166}$$

$$z \mapsto \frac{bc - ad}{c^2}z \quad \text{(a spiral rotation);} \tag{167}$$

$$z \mapsto z + \frac{a}{c} \quad \text{(a translation).} \tag{168}$$

There is an intimate connection between Möbius transformations and 2×2 complex matrices which we now describe. We denote by

$$T_{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$$

the standard Möbius transformation. Then a simple calculation shows that if A and B are invertible, 2×2 matrices, the composition of the corresponding Möbius transformations corresponds to matrix multiplication ie. $T_A \circ T_B =$

T_{AB}. In particular, the inverse of T_A is the transformation with matrix A^{-1} i.e.

$$\frac{1}{ad-bc}\left[\begin{array}{cc} d & -b \\ -c & a \end{array}\right]$$

which confirms a result mentioned above (of course we can drop the factor $\frac{1}{ad-bc}$ by the homogeneity).

The special types of transformation listed above are induced by the matrices:

$$\left[\begin{array}{cc} 1 & b \\ 0 & 1 \end{array}\right] - \text{ translation;} \tag{169}$$

$$\left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] - \text{ inversion and reflection;} \tag{170}$$

$$\left[\begin{array}{cc} a & 0 \\ 0 & 1 \end{array}\right] - \text{ spiral rotations.} \tag{171}$$

We note that Möbius transformations preserve angle i.e. if C and C_1 are curves in the plane which cross at the point x at an angle θ, then the same holds for the images of C and C_1 under a Möbius transformation T at Tx. The easiest (if least elementary) way to see this is to note that Möbius transformation are complex analytic (except for a single pole) and analytic functions are angle-preserving. However, we can prove it more elementarily by noting that it suffices to show that the result is true for the three basic types of operation—translation, spiral rotation and taking of inverses. The first two trivially have the required property and the case of inversion can be dealt with by elementary calculus—calculating the Jacobi matrix of the mapping $z \mapsto \frac{1}{z}$).

Note that Möbius transformations preserve cross-ratios i.e. we have the equality

$$(Tz_1, Tz_2; Tz_3, Tz_4) = (z_1, z_2; z_3, z_4)$$

for z_1, z_2, z_3, $z_4 \in \mathbf{C}$. This can be verified by substituting the values of Tz_1 etc. in the left-hand side and simplifying. (Even simpler, note that it suffices to verify the equations for the special type of transformation discussed above. Only the case of inversion is not completely trivial).

We can deduce some interesting consequence from this fact. For example, a Möbius transformation is determined by the images of three distinct points. More precisely, if z_1,, z_2 and z_3 resp. w_1, w_2 and w_3 are distinct points in the plane, then the Möbius transformation which maps each z_i onto w_i is given by the equation

$$(w,w_1;w_2,w_3) = (z,z_1;z_2,z_3).$$

For it follows from the above property that if w is the image of z under T, then the above equation must hold. On the other hand, this equation, when solved for w as a function of z, provides a Möbius transformation.

The following special case is often useful. If z_1 and z_2 are distinct points in the plane, then the only Möbius transformations which leave z_1 and z_2 fixed are those of the form

$$\frac{w-z_1}{w-z_2} = a \cdot \frac{z-z_1}{z-z_2}$$

where a is a non-zero complex number.

This implies the result above that inversion in a circle has this property.

Another important property of Möbius transformations is that they map circles (or straight line) onto circles (or straight lines). More precisely, a circle or straight line is mapped onto a circle, unless it contains that point which is mapped onto the point at infinity, in which case its image is a straight line. In order to prove this, it suffices to consider the three special types of Möbius transformation. Once again, the first two trivially have the required property and so we can confine our attention to transformations of the form $z \mapsto \frac{1}{z}$. Then we note that the equation

$$\xi_1^2 + \xi_2^2 - 2a\xi_1 - 2b\xi_2 + c = 0$$

for the circle can be written in the complex form as

$$z\bar{z} + \bar{d}z + d\bar{z} + c = 0$$

where $d + a + ib$. Then if we substitute $\frac{1}{z}$ for z, we get the equation

$$cz\bar{z} + \bar{d}\bar{z} + dz + 1 = 0$$

which is also the equation of a circle (or a straight line if $c = 0$ i.e. if 0 lies on the circle).

From this fact we can easily deduce that if z_1, z_2 and z_3 are distinct points in the plane, then the complex equation of the circle through them (or of the straight line, if they are collinear) is

$$\Im(z,z_1;z_2,z_3) = 0.$$

For there exists a Möbius transformation T so that tz_1, Tz_2 and Tz_3 all lie on the real axis. Then a point w lies on the real axis if and only if $(w, Tz_1; Tz_2, Tz_3)$ is real. Now the pre-image of the real axis under T is a circle (or straight line) which contains z_1, z_2 and z_3 and a point z lies on this line if and only if $w = Tz$ satisfies this condition. Thus we see that z lies on the circle through z_1, z_2 and z_3 if and only if $\Im(z, z_1; z_2, z_3) = 0$.

The same reasoning shows that four points z_1, z_2 z_3 and z_4 are cyclic or collinear if and only if their cross-ratio is real. As an application of this last fact, suppose that C_1, C_2 C_3 and C_4 are circles, each of which meets two others in two points which are labelled as in figure ??. Then if A_1, A_2, A_3 and A_4 are cyclic, then so are B_1, B_2, B_3 and B_4. This is proved as follows: the conditions on the circles mean that the following cross-ratios are all real:

$$(A_1, A_2; A_3, A_4) \quad (A_2, B_3; A_3, B_2) \quad (A_3, B_4; A_4, B_3) \quad (A_4, B_1; A_1, B_4).$$

hence their product is real. if we multiply these out and cancel, this reduces to the product

$$(A_1, A_3; A_2, A_4) \cdot (B_1, B_3; B_2, B_4)$$

and so the second term is real if the first one is.

2.9 Three dimensional geometry

The model for three dimensional geometry is the vector space $\mathbf{R}^3$. In addition to its vector space structure and scalar product which are completely analogous to those of $\mathbf{R}^2$, it possesses two special ones, the vector product $x \times y$ and the dot product $[x, y, z]$ which are defined as follows:

$$x \times y = (\xi_2\eta_3 - \xi_3\eta_2, \xi_3\eta_1 - \xi_1\eta_3, \xi_1\eta_2 - \xi_2\eta_1)$$

and

$$[x, y, z] = (x|y \times z).$$

$[x, y, z]$ is the signed area of the parallelopiped spanned by x, y and z, while $x \times y$ is a vector which is perpendicular to the plane spanned by x and y (assuming that they are not proportional) and whose length is the area of the parallelogram spanned by x and y (figure 1).

Planes in three dimensional space are determined by four parameters. More precisely, if a, b, c and d are real numbers with $a^2 + b^2 + c^2 \neq 0$, then

$$M_{a,b,c,d} = \{x : a\xi_1 + b\xi_2 + c\xi_3 + d = 0\}$$

is a plane (which, of course, remains unchanged if we replace the quadruple (a, b, c, d) by some non-zero multiple). The vector

$$\mathbf{n} = \frac{(a, b, c)}{\sqrt(a^2 + b^2 + c^2)}$$

is the **unit normal** to the plane and we can write its equation in Hessean form

$$\{x : (x|\mathbf{n}) = (x_0|\mathbf{n})\}$$

where x_0 is an arbitrary point in the plane. Then if $x \in \mathbf{R}^3$, $d(x, M) = (x - x_0|\mathbf{n})$ is the directed distance from x to the plane.

Lines in $\mathbf{R}^3$ can be described as the intersection of two non-parallel planes. hence they have Hessean form

$$\{x : (x - x_0|\mathbf{n}) = 0 = (x - x_0|\mathbf{n}_1)\}$$

where $\mathbf{n}$ and $\mathbf{n}_1$ are normals to the two planes (and so $\mathbf{n} \neq \mathbf{n}_1$ and $\mathbf{n} \neq -\mathbf{n}_1$) and x_0 is an arbitrary point on the line (figure 2). In coordinates, this means that the line is the solution of a system of two linear equations in three unknowns.

We remark here that three lines

$$L_{b,c}, \quad L_{a_1,b_1,c_1} \quad \text{and} \quad L_{a_2,b_2,c_2}$$

in the plane are concurrent (or parallel) if and only if the vectors (a, b, c), (a_1, b_1, c_1) and $a_2, b_2, c_2)$ are linearly dependent (we have already used this result in I.4). This is a simple exercise in linear algebra. The fact that the point (ξ_1, ξ_2) lies on all three lines means that the homogeneous equation with matrix

$$\begin{bmatrix} a & b & c \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix}$$

has the non-trivial solution $(\xi_1, \xi_2, 1)$. hence the rank of the matrix is less than 3 i.e. the three vectors are linearly dependent. If the three lines are parallel, then the matrix

$$\begin{bmatrix} a & b \\ a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

has rank 1 and so the above 3×3 matrix has tank at most 2. On the other hand, if the rank is less than 2, the above homogeneous equation has a non-trivial solution (η_1, η_2, η_3). If $\eta_3 \neq 0$, then we can divide through to get a solution of the form $(\xi_1, \xi_2, 1)$ which just means that $x = (\xi_1, \xi_2)$ is in the intersection of the lines. If $\eta_3 = 0$, then a similar argument shows that the lines are parallel.

The natural three dimensional analogue of a triangle is a tetrahedron, which is determined by four vertices A, B, C, D (figure 3). The tetrahedron is **non-degenerate** if the corresponding vectors x_A, x_B, x_C and x_D are affinely independent i.e. if whenever λ_1, λ_2, λ_3 and λ_4 are scalars whose sum is zero so that $\lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_d = 0$, then $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$. If this is the case, then every point P has a unique representation

$$x_P = \lambda_1 x_A + \lambda_2 x_B + \lambda_3 x_C + \lambda_4 x_D$$

with $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$. The λ's are called the **barycentric coordinates** of P with respect to A, B, C and D.

We shall now prove some simple facts about tetrahedra:

Proposition 37 *Let A,B,C, D be the vertices of a tetrahedron for which the sides AB and CD (resp. AC and BD) are perpendicular. Then so are the sides AD and BC.*

PROOF. For the first two conditions can be expressed analytically in the form

$$(x_B - x_A | x_D - x_C) = 0 \text{ i.e. } (x_B|x_D) - (x_A|x_D) - (x_B|x_C) + (x_A|x_C) = 0$$

resp.

$$(x_C - x_A | x_B - x_D) = 0 \text{ i.e. } (x_C|x_B) - (x_A|x_B) - (x_C|x_D) + (x_A|x_D) = 0.$$

If we add the two equations, we get

$$(x_B|x_D) - (x_A|x_B) - (x_C|x_D) + (x_A|x_C) = 0.$$

i.e. $(x_D - x_A | x_C - x_B) = 0$.

■

As a second result, recall that a median of the tetrahedron is a line from a vertex to the centroid of the opposite triangle–for example the line from x_A to $\frac{1}{3}(x_B + x_C + x_C)$ (figure ??). It is clear that the point

$$x_M = \frac{1}{4}(x_A + x_B + x_C + x_D$$

is on this line from which we deduce the following result:

Proposition 38 *If $ABCD$ is a tetrahedron, then the medians are concurrent and intersect each other in the ratio* 3 *to* 1*.*

We conclude this section with the proofs of two three-dimensional analogues of the triangle inequality which follow very simply from the properties of the various products in $\mathbf{R}^3$.

Suppose that x,y and z are points in space. Then the following inequalities hold:

$$(y|y)(x|z) \geq (y|z)(x|y) - \|y \times z\| \|x \times y\|$$

and

$$\|x \times y\| + \|y \times z\| + \|z \times x\| \geq \|(x-y) \times (y-z)\|.$$

The first follows immediately from the identity

$$\begin{aligned} (y|z)(x|y) - (y|y)(x|z) &= (y|((x|y)z - (x|z)y)) && (172)\\ &= -(y|x \times (y \times z)) && (173)\\ &= [x, y, y \times z] && (174)\\ &= [y \times z, x, y] && (175)\\ &= (y \times z | x \times y) && (176)\\ &\geq \|y \times z\| \|x \times y\|. && (177) \end{aligned}$$

(Here we have use the identity

$$(x \times (y \times z) = (x|z)y - (x|y)z$$

which is proved in ??)

For the second inequality, we multiply out the right hand side to get

$$\|x \times y + y \times z + z \times x\|$$

and use the triangle inequality.

These inequalities can be interpreted geometrically as follows:

Consider a tetrahedron $OABC$. If we assume that AO, OB and OC are unit vectors and set $x = x_{OC}$, $y = x_{OB}$, $z = x_{OA}$ in the above inequality, we see that

$$\begin{aligned} \cos(\angle AOC) &\geq \cos(\angle AOB)\cos(\angle BOC) - \sin(\angle AOB)\sin(\angle BOC) && (178)\\ &= \cos(\angle AOB + \angle BOC). && (179) \end{aligned}$$

Since the cosine decreases with angle, this means that the angle $\angle AOC$ is less than the sum of the angles $\angle AOB$ and $\angle BOC$.

In order to interpret the second inequality, we return to the above tetrahedron but drop the condition that OA, OB and OC have unit length. If we set

$$x = x_{OA} \quad y = x_{OB} \quad z = x_{OC}$$

and interpret the norm of the vector product of two vectors as twice the area of the triangle that they span, we see that we have shown that the area of the triangle ABC is less than or equal to the sum of the areas of OAB, OBC and OCA.

of course, both of these results can be regarded as spacial versions of the triangle inequality.

Isometries We now turn to the topic of isometries in three dimensions. These can be analysed as in the two dimensional case, with the added complications that one would expect from the increase in dimension. In fact, one can reduce the classification to the two dimensional case by using the following simple fact: if f is a linear isometry of $\mathbf{R}^3$, then there is a unit vector x_1 so that $f(x_1) = \pm x_1$. if we then choose vectors x_2 and x_3 so that the three form an orthonormal basis, then the matrix of f with respect to this basis has the form

$$\begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & a_{11} & \alpha_{12} \\ 0 & a_{21} & a_{22} \end{bmatrix}$$

where A is the matrix of an isometry of the plane. By investigating the form of the latter, one produces the following list of possible linear isometries in space:

- $f(x_1) = x_1$ and A is the matrix of a rotation. Then f is a rotation around the axis x_1;
- $f(x_1) = f(x_1$ and A is the matrix of a reflection in the line L. Then f is a reflection in the plane spanned by L and x_1.
- $F(x_1) = -x_1$ and A is the matrix of a rotation. Then f is a rotary reflection i.e. a reflection in the plane spanned by x_1 and x_2, followed by a rotation around the axis perpendicular to this plane.

- $f(x_1) = -x_1$ ad A is the matrix of a reflection. In this case, f is rotation through 180^o about the line of reflection.

We now consider non-linear isometries i.e. those of the form $f = T_u \circ \tilde{f}$ where $\tilde{f}$ is the linear isometry with the same matrix. Again we examine various possibilities:

- $\tilde{f}$ is a rotation. Then we split u into the sum $u_1 + u_2$ where u_1 is parallel to the axis x_1 and u_2 is perpendicular to it. We claim the $T_{u_2} \circ \tilde{f}$ is a rotation about a line parallel to x_1. This is because the action of the mappings in the above composition take place in the plane orthogonal to x_1. But in the plane, a rotation followed by a translation is a rotation. Thus f is a rotation, followed by a translation in the direction of the axis of rotation. Such mappings are called **screw displacements.**

- f is a reflection in a plane M. We now write u as $u_1 + u_2$ where u_1 is parallel to M and u_2 is perpendicular. Then we claim that $T_{u_2} \circ R_M$ is a reflection in a plane parallel to M. Once again, this is proved by reducing to the dimensional fact that a reflection in a line, followed by a translation, is a reflection in a line parallel to the original one. Hence f is what we call a **glide reflection** i.e. a reflection in a plane, followed by a translation parallel to this plane (figure ??).

- f is a rotary reflection, say reflection in the plane M, followed by rotation around a vector x_1 which is perpendicular to M. Then f is itself a rotary reflection. Perhaps the easiest way to see this is to show that such an f must have a fixed point x. This is not difficult to do. Then if $g = T_{-x} \circ f \circ T_x$, g is a linear isometry and, in fact, a rotary reflection also. Hence so is $f + T_x \circ g \circ T_{-x}$.

(We remark that the cases of a translation, rotation or reflection are contained in the above as degenerate cases. Thus a translation is a screw-displacement where the rotational component is trivial).

Summing up, we have the following list of possible isometries of space:

- a translation;

- a rotation;

- a reflection;

- a screw-displacement;
- a glide reflection;
- a rotary reflection.

The Platonic solids: The culmination of Euclid's text is the topic of the Platonic bodies i.e. the regular polytopes in space. In contrast to the case of the plane where there are infinitely many regular polygons, there only five of the former. These are the tetrahedron, the cube or hexahedron, the octahedron, the dodecahedron and the icosahedron. We shall prove their existence by describing them with the methods of analytic geometry. We begin, however, with the proof that there are, in fact, only five. For suppose that we have a regular polyhedron in three-dimensional space and let each of its faces be a regular n-gon. We suppose that r faces meet at each vertex. Of course, r and n are both at least three. The sum of the angles at a vertex is strictly less than 2π (otherwise the polyhedron would be flat or non-convex). This fact leads to the inequality $\dfrac{r(n-2)}{n} < 2$. Trial and error shows that the only pairs (r, n) which satisfy this inequality are

$$(3,3) \quad (3,4) \quad (3,5) \quad (4,3) \quad (5,4).$$

These correspond precisely to the five solids mentioned above. We now discuss the individual cases in more detail.

The tetrahedron: Consider the points

$$A = (1,0,0) \quad B = (0,1,0) \quad C = (0,0,1).$$

Then there are two points O_1 and O_2 so that

$$|O_1A| = |O_1B| - |)_1C| = \sqrt{2}$$

and

$$|)_2A| = |O_2B| = |O_2C| = \sqrt{2}.$$

(see figure ??). For reasons of symmetry, a suitable O must have coordinates of the form (ξ, ξ, ξ) and the equation $|OA|^2 = 2$ then becomes

$$(\xi_1)^2 + \xi^2 + \xi^2 = 2 \text{ or } 3\xi^2 - 2\xi - 1 = 0$$

with solutions $\xi = \frac{1}{3}$ or $\xi = 1$. we choose one of these points and denote it by O. Then $OABC$ is a regular tetrahedron since each of its faces is an equilateral triangle. If can be more conveniently represented as being inscribed in a cube as in figure 9 below).

The cube: The eight points

$$A \quad (1,-1,-1) \qquad B \quad (1,1,-1) \qquad C \quad (-1,1,1) \qquad (-1,-1,-1)$$
$$E \quad (1,-1,1) \qquad F \quad (1,-1,1) \qquad G \quad (-1,1,1) \qquad H \quad (-1,-1,1)$$

form the vertices of a cube (figure ??). Note that then A, C, F and G are the vertices of a regular tetrahedron as remarked above.

The midpoints of the sides AB, BF, FG, GH, HD and DA lie in a plane and form the vertices of a regular hexagon as the reader can verify.

The octahedron: The centroids of the faces of the above cube form the vertices of a regular octahedron (as do the midpoints of the sides of the regular tetrahedron). Using the first representation, wee obtain the vertices

$$(1,0,0) \quad (-1,0,0) \quad (0,1,0) \quad (0,-1,0) \quad (0,0,1) \quad (0,0,-1)$$

for the octahedron (figure ??).

The icosahedron: If $b > 0$, then the twelve points

$$(0,\pm 1,\pm b) \qquad (\pm b,0,\pm 1) \qquad (\pm 1,\pm b,0)$$

forms the vertices of an icosahedron (figure ??). These twelve points are the vertices of three triangles, one in each of the coordinate planes. A simple calculation shows that this icosahedron is regular if and only if b is the *golden mean* $(= \frac{1}{2}(\sqrt{5}+1))$ and this provides a coordinate representation of the regular icosahedron.

The dodecahedron: This is the polytope whose vertices are the centroids of the faces of the regular icosahedron.

Using coordinate representations, it is now a routine if tedious matter to calculate the various parameters associated with the regular polyhedra such as the dihedral angles (i.e. the angles between adjacent faces), their volumes and surface areas.

3 Algebra

3.1 Groups

The concept of a group is an abstraction of that of a transformation group. One simply takes the defining characteristics of the latter as axioms. Hence a group is a set G together with a multiplication i.e. a mapping $(g, h) \mapsto gh$ from $G \times G$ to G so that

- "a)" $g_1(g_2g_3) = (g_1g_2)g_3(g_1, g_2, g_3 \in G)$;
- "b)" there is an $e \in G$ so that $ge = eg = g$ for each $g \in G$;
- "c)" if $g \in G$ there is an $h \in G$ so that $gh = hg = e$.

The element e of G with property b) is uniquely determined and is called the **unit** of the group. Similarly, the element h in c) is uniquely determined by g and is called the **inverse** of g–written g^{-1}. If $g \in G$, $r \in \mathbf{N}$, g^r denotes the product of r copies of g. If $r \in \mathbf{Z}$ is negative, g^r is defined to be $(g^-1)^{-r}$. The reader can check that

- "d)" if $gh = gh_1$, then $h = h_1$ (multiply both sides on the left by the inverse of g);
- "e)" $(gh)^-1 = h^{-1}g^{-1} \quad (g, h \in G)$;
- "f)" $g^rg^s = g^{r+s} \quad (g^r)^s = g^{rs} \quad (g \in G, r, s \in \mathbf{Z})$.

A **subgroup** of a group is a subset G_1 so that

- "g)" G_1 is closed under multiplication i.e. $g, h \in G_1$ implies that $gh \in G_1$;
- "h)" $e \in G_1$;
- "i)" if $g \in G_1$, then $G^{-1} \in G_1$.

Then G_1 is itself a group. if G is a non-empty subset of G, then these conditions can be subsumed in the single one that gh^{-1} be in G_1 whenever g and h are.

If g_1,...,g_k are elements of a group G, there is a *smallest* subgroup of G which contains them. it consists of all products of elements of this set, together with their inverses (with repetitions i.e. an element can occur more than once in such a product). In order to ensure that the unit is always in this set we introduce the convention that e is the product of an empty set of elements. This subgroup is called the **subgroup generated** by $g_1, \dots, g_k$, written $\langle g_1, \dots, g_k \rangle$. The elements $g_1, \dots, g_k$ are called **generators** of this subgroup. Particularly important are subgroups which are generated by *one* element i.e. subgroups of the form $\langle g \rangle$ $(g \in G)$. The elements of the latter consist of all powers (positive or negative) of g. There are two possibilities:

- there is an $n \in \mathbf{N}$ so that $G^n = e$. g is then said to be of **finite order** an smallest such n is called the **order** of g. $\langle g \rangle$ then consists of the n elements $e, g, g^2, \dots, g^{n-1}$;
- there is no such n i.e. $g^n \neq e$ for each $n \in \mathbf{N}$. Then the subgroup is infinite and consists of the set of distinct elements $\{g^n : n \in \mathbf{Z}\}$.

The corresponding groups are denoted by $C(n)$ resp. $C(\infty)$ and are called the **cyclic groups** of order n resp. of infinite order. Note that in a certain sense (which will be made precise below), all cyclic groups of the same order are identical and so we are justified in introducing the same name for each of them.

The simplest non-trivial example (which is of some importance despite its simplicity) is $C(2)$. This can be realised as the subgroup $\{-1, 1\}$ of the multiplicative group of non-zero real numbers.

If H is a subgroup of G, then we define the right **cosets** of H to be the sets of the form $Hg = \{hg : h \in H\}$. Then the following simple calculation shows that two such cosets Hg and Hg_1 are identical if and only if $g_1 g^{-1} \in H$. (If this is not the case, then they are disjoint). For suppose that $g_2 \in Hg \cap Hg_1$. Then g_2 has representation hg and $h_1 g_1$ with h and h_1 from H. We deduce that $g_1 f^{-1} = h_1^{-1} h$ and so is in H. On the other hand, it is clear that if $g_1 g^{-1} \in H$, then $Hg = Hg_1$.

Since G is the union of the cosets of H and this is a disjoint union, the cardinality $|H|$ of H divides that of G (provided the latter is finite). For each coset of H has the same cardinality as H since the mapping $h \mapsto hg$ is a bijection from H onto Hg. If we apply this to the cyclic subgroup generated by an element g of G we see that the order of g is a divisor of $|G|$ when the latter is finite. This is known as **Lagrange's theorem.**

The collection of *left* cosets of H is defined analogously i.e. as the sets of the form gH $(g \in G)$. Particularly important is the case of subgroups for which the left and right cosets coincide. This means that for each $g \in H$, $gH = Hg$ or $g^{-1}Hg = H$. Such subgroups are called **normal**. (Of course, if G is abelian, *all* subgroups are normal). Then the set of cosets has a natural group structure which is defined as follows: we define the product of two cosets Hg and Hg_1 to be $Hgg - 1$. Note that this is independent of the choice of g and g_1. For if we replace g by gh and g_1 by $g_1 h_1$ (here we are using the fact that a right coset is also a left coset), then gg_1 is replaced by $h(gg_1)h_1$ which defines the same coset.

It is then rather simple, if tedious, to carry out the computations required to show that the group axioms are satisfied for this multiplication. The corresponding group is called the **quotient** of G by H–denoted by G/H. The mapping $\pi : G \to G/H$ which maps g onto Hg is called the **canonical projection** from g onto the quotient.

If G and G_1 are groups, then a mapping $\phi : G \to G_1$ is called a **homomorphism** if it preserves the group operations i.e. if $\phi(gh) = \phi(g)\phi(h)$ for $g, h \in G$ (this implies that $\phi(e) = e$, a fact that the reader may like to check). Note here that we are committing the cardinal sin of denoting the units of G and G_1 by the same letter. This will hardly lead to any confusion and a more careful notation would be hopelessly turgid). Then $\phi(g^{-1} = \phi(g)^{-1}$ since the product of the left-hand side with $\phi(g)$ is clearly e.

If ϕ is, in addition, a bijection, then its inverse ϕ^{-1} is also a homomorphism. ϕ is then called a **(group) isomorphism** and G and $G-1$ are said to be **isomorphic**. This means that as abstract mathematical objects, they are indistinguishable. The reader will have noted that any two cyclic groups of the same order are isomorphic which justifies the use of a common name for them.

In this context, we can now state and prove three basic results on isomorphism, called the *three isomorphism theorems:*

Proposition 39 *I. If $\phi : G \to H$ is a surjective homomorphism, then $G/Ker\phi \simeq H$ where $Ker\phi = \{g \in G : \phi(g) = e\}$. (The symbol $\simeq$ denotes the fact that the two groups are isomorphic);*
II. If G_1 and G_2 are subgroups of a group G with G_2 normal, then $G_1 \cap G_2$ is normal in G_1 and

$$G_1/G_1 \cap G_2 \simeq G_1G_2/G_2$$

;
III. If $G_2 subset G_2$ where both are normal subgroups of G, then G_1/G_2 is normal in G/G_2 and

$$(G/G_2)/(G_1/G_2) \simeq G/G_1.$$

Proof. *I. Implicit in the statement is the fact that $K = Ker\phi$ is a normal subgroup of G. This can be checked very easily. We then define mappings*

$$\phi_1 : G/Kto \quad by \quad Kg \mapsto g$$

resp.

$$\phi_2 : H \to G/K \quad by \quad h \mapsto Kg \quad where\ g\ is\ such\ that\ \phi(g) = h.$$

Then one can verify that phi$_1$ and ϕ_2 are well-defined, mutually inverse homomorphisms.
II. This can be deduced from I as follows: consider the mapping $\phi : g \mapsto gG_2$ from G_1 onto G/G_2. Its image is the set of cosets of the form g_1G_2 ($g_1 \in G$) and this is clearly G_1G_2/G_2 (note that $G_1G_2 = \{g_1g_2 : g_1 \in G_1, g_2 \in G_2\}$ is a subgroup. This uses the normality of G_2. Hence, by I, $G_1/Ker\phi$ is isomorphic to G_1G_2/G_2. Bur $Ker\phi$ is easily seen to be just $G_1 \cap G_2$.
III. We leave the prooff III to the reader.

■

We now identify various special subgroups and subsets of groups which are of basic importance: If $g \in G$, then any element of the form $h^{-1}gh$ is called a **conjugate** *of g (more precisely,k the conjugate of g by h). $C_G(g) = \{h \in G : h^{-1}gh = g\}$ is called the* **centraliser** *of g (i.e. it is the set of elements of the group which commute with g).*

We remark that the mapping $g \mapsto h^{-1}gh$ is a homomorphism from G into itself (homomorphism from G into itself are called **automorphisms**. *Those of the above form are called* **inner automorphisms**. *A subgroup is normal if and only if it is stable under any such automorphism.*

The intersection $Z(G) = \bigcap_{g \in G} C_G(g)$ is called the **centraliser** *of G i.e. it consists of those elements which commute with each element of G. it is easy to check that $C_G(g)$ is a subgroup of G and that $Z(G)$ is a normal subgroup.*

Notice that two conjugates $h^{-1}gh$ and $h_1^{-1}gh_1$ of g coincide if and only if $hh_1^{-1} \in C_G(g)$. This means that the **conjugacy class** *of g (i.e. the set of members of the group which are conjugate to g) is in one-one correspondence with the set of right cosets of $C_G(g)$. Since the cardinality of the latter is a divisor of $|G|$ (namely $|G|/|C_G(g)|$), it follows that the number of elements conjugate to g is always a divisor of $|G|$ (G finite).*

We shall now discuss briefly some methods of constructing new groups from old ones. One of the simplest and most useful is that of taking Cartesian products. We begin with the case of two groups. If G_1 and G_2 are groups, then we can provide the Cartesian product $G_1 \times G_2$ with a multiplication in the natural way–we define

$$(g_1, g_2)(h_1, h_2) = (g_1h_1, g_2h_2).$$

The reader will have no trouble in verifying that $G_1 \times G_2$ is itself a group. Also the mappings

$$g \mapsto (g, e) \quad \textit{resp.} \quad h \mapsto (e, h)$$

display G_1 and G_2 as normal subgroups of the product (i.e. they are isomorphisms from G_1 resp. G_2 onto normal subgroups of the product).

An internal characterisation of when a given group G splits up into a product of two smaller ones is as follows:

Proposition 40 *If G is a group with normal subgroups H and K so that $H \cap K = \{e\}$ and $HK = G$, then G is isomorphic to the product $H \times K$.*

PROOF. *The condition $HK = G$ means that each element of G has a representation hk with $h \in K$, $k \in K$. On the other hand, this representation is unique since if $hk = h_1 k_1$, we have $h_1^{-1} h = k_1 k^{-1}$ and the left hand side belongs to H while the right hand side is in K. hence both are in $H \cap K$ ie. are equal to e.*

We now show that if $h \in H$, $k \in K$, then $hk = kh$. For consider $h^{-1}k^{-1}hk$. If we write this as $h^{-1}(k_{-1}hk)$ we see that it is in H. Similarly, it is in K and so is equal to e which means that $hk = kh$. It is now a routine matter to check that the mapping $(h, k) \mapsto hk$ is an isomorphism from $H \times K$ onto G.

■

In this connection we have the following isomorphism theorem:

Proposition 41 *Let G be the product $G_1 \times G_2$ of two groups and let H_1 resp. H_2 be normal subgroups of G_1 resp. G_2. Then $H_1 \times H_2$ is normal in $G_1 \times G - 2$ and there is a natural isomorphism*

$$G?(H_1 \times H_2) \simeq (G_1/H_1) \times (G_2/H_2).$$

PROOF. *We define a homomorphism π from $G_1 \times G_2$ onto the right hand side by putting*

$$\phi(g_1, g_2) = (H_1 g_1, H_2 g_2).$$

Its kernel is just $H_1 \times H_2$ and so the result follows from the first isomorphism theorem.

■

The above construction of products can be generalised to the case of infinite products as follows: let $(G_\alpha)_{\alpha \in A}$ be a family of groups. We define two new groups G and H. As a set G is the Cartesian product $\prod G_\alpha$ and H is the subset of G consisting of those sequences (g_α) for which $g_\alpha = e$ except for finitely many of the α. We define a product on G in the natural way. If $g = (g_\alpha)$, $h = (h_\alpha)$, then $gh = (g_\alpha h_\alpha)$. It is easy to check that G is a group under this operation and that H is a subgroup. G is called the **Cartesian product** *of the G_α's–written $\prod G_\alpha$–and H is called the* **direct sum**–*written $\oplus G_\alpha$.*

We remark that we have natural mappings

$$G_\beta \to \prod_{\alpha\in A} G_\alpha \to \prod_{\alpha\in A} G_\alpha \to G_\beta$$

all of which are homomorphisms. We can thus regard each G_β as a subgroup and a quotient group of the direct sum and the Cartesian product. In addition G_β is a normal subgroup as is $\oplus_{\alpha\in A} G_\alpha$ of $\prod_{\alpha\in A} G_\alpha$.

Representations *Although (or rather because) groups are defined abstractly, one is interested in obtaining concrete representations of them in terms of the more concrete objects which arise in geometry. This is subsumed in the notation of a representation as defined below. There are two main types of such representations:*
I. As transformation groups. Here we have a set S and a homomorphism Φ from G into the group Per(S) of all bijections of S. it is convenient to write Φ_g for the image of g under the representation. It is traditional to call the representation **faithful** *if it is injective (this is a not uncommon example in mathematics of a special case of a general concept retaining an older name which, strictly speaking, is superfluous). This means that Φ is an isomorphism from G onto a subgroup of Per(S) i.e. is a transformation group on S in the sense of ??? It is also tradition to consider Φ as a mapping from $G \times S$ into S by putting $\Phi(g, s) = \Phi_g(s)$ in which case the following equations hold:*

$$\Phi(g, \Phi(h, s)) = \Phi(gh, s)$$

and

$$\Phi(e, s) = s \qquad (g, h \in G, s \in S).$$

The reason for this is that the classical examples were drawn from physics (more precisely, the theory of dynamical systems). here the group G is **R** *(as a time variable), the set S is the phase space of some physical system and $\Phi(g, s)$ describes the state (at time g) of the system which starts in a state s at time 0 (the neutral element of the group). Then the above equations have the natural interpretations.*
II. Linear representations. here the set S is replaced by a vector space and the representation is a homomorphism from G into $GL(V)$, the group of nvertible linear mappings on V. hence to each $g \in G$, we associate an element $T_g \in L(V)$ in such a way that

$$T_e + Id \quad T_{gh} = T_g \circ T_h.$$

If V is finite dimensional (as it will be in our examples), then we can choose a basis for V and so obtain a so-called **matrix-representation** *for G i.e. a mapping $g \mapsto A_g$ from G into M_n (where $n = dim V$) so that*

$$A_e = I_n \quad and \quad A_{gh} = A_g A_h.$$

If we have a representation of a group G as a transformation group on a set S it is often convenient to shift the emphasis and call S (more pedantically, the triple (G, Φ, S)) a **G-space** *and say that G* **acts** *on S. Then two G-spaces (G, Φ, S) and $G, \Phi', S')$ are said to be* **isomorphic** *if there is a bijection $f : S \to S'$ so that $\Phi'_g = f^{-1} \circ \Phi_g \circ f (g \in G)$. This is just a fancy way of saying that f does not merely establish the fact that S and S' are isomorphic (as sets) but it also preserves their structures as G-spaces.*

More generally, a **G-morphism** *from S into S' is a mapping $f : S \to S'$ so that $f \circ \Phi'_g = \Phi_g \circ f$ for each $g \in G$.*

We list some examples of G-spaces:

I. The regular representation. Each group acts on itself by right multiplication. More precisely, if we write r_g for the mapping $h \mapsto hg$ on G, then $g \mapsto r_g$ is a homomorphism from G into $Per(G)$. It is called the **(right) regular representation***. Of course, this representation is faithful and this shows that every group "is" a transformation in a certain sense.*

II. Each group acts on itself by conjugation i.e. if we define for each $g \in G$, the mapping Φ_g by the equation

$$\Phi_g(h) = ghg^{-1}$$

then $g \mapsto \Phi_g$ is a representation of G in $Per G$ (even in $Aut(G)$, the group of automorphisms of G). Note that this representation need not be faithful. In fact, for an Abelian group, it is trivial (i.e. Φ_g is the identity for each g).

III. Each of the matrix groups

$$Aff(n) \quad Isom(n) \quad GL(n) \quad SL(n) \quad O(n) \quad SO(n)$$

(see ???) act on $\mathbf{R}^n$.

IV. We can generalise I as follows: let G be a group, H a subgroup (not necessarily normal). Then G acts on the set of right cosets of H by right multiplication i.e. if we define, for $g \in G$,

$$\Phi_g : Hh \mapsto Hhg$$

then Φ is a representation of G on the set of right cosets of H. G-spaces of this type are called **symmetric spaces** *and are particularly in applications.*

V. In general, a group action on a set defines a related action on spaces of functions on this set. A typical example of this is the following: first note that the permutation group S_n acts on $\mathbf{R}^n$ as follows: if $\pi \in S_n$ and $x = (\xi_1, \dots, \xi_n)$, then we define

$$x_\pi = (\xi_{\pi(1)}, \dots, \xi_{\pi(n)}).$$

This induces an action on the set of functions f on $\mathbf{R}^n$ as follows: if $f : \mathbf{R}^n \to \mathbf{R}$, then we define a new function f^π by putting $f^\pi(x) = f(x_{\pi^{-1}})$ i.e.

$$f^\pi((\xi_1, \dots, \xi_n)) = f((\xi_{\pi^{-1}(1)}, \dots, \xi_{\pi^{-1}(n)})).$$

We now introduce some general terminology for G-spaces. The group G is said to act **transitively** *on S if for each s and t in S there is a $g \in G$ so that $\Phi_g(s) = t$. For example, the action of Aff(n) on $\mathbf{R}^n$ is transitive, whereas that of $)(n)$ is not.*

if $s \in S$, then the set

$$Orb(s) = \{\Phi_g(s) : g \in G\}$$

of images of s under the group action is called the **orbit** *of s under the action. The classical example is the $(n-1)$-dimensional sphere S^{n-1} which is the orbit of the point $(1, 0, \dots, 0)$ in $\mathbf{R}^n$ under the action of $)(n)$. More generally, the orbit of any point in $\mathbf{R}^n$ is the sphere with radius equal to the length of the corresponding vector. If G acts on itself by conjugacy, then the orbits are just the conjugacy classes. note that two orbits in S are either disjoint or identical. Hence S is the disjoint union of the orbits. Of ;course, there is precisely one orbit if and only if the action of G is transitive.*

The **stabiliser** *of a point s in S (written Stab(s)) is the set of group elements which leave s fixed i.e.*

$$Stab(s) = \{g \in G : \Phi_g(s) = s\}.$$

For example, under the action of G on itself by conjugacy, the stabiliser of g is just the set $C_G(g)$.

It is easy to check that Stab(s) is a subgroup. Also the right cosets of this subgroup are in correspondence with the points of the orbit of s. More precisely, if we consider the surjection $g \mapsto \Phi_g(s)$ from G onto Orb(s), then the elements of the coset Stab$(s)h$ are mapped onto $\Phi_h(s)$. Conversely, two elements h and h_1 of G are mapped onto the same point in the orbit if and only if they belong to the same coset of Stab(s). From this it follows that if G is a finite group, then its cardinality is the product of the cardinalities of Stab(s) and Orb(s) (and, in particular, the latter two quantities are divisors of $|G|$).

3.2 Abelian groups

In this section, we shall discuss some topics which are special to abelian groups. It is traditional to emphasise their commutativity by writing the operation additively. This means that we write $g + h$ *for what up until now has been* gh*. As a consequence we write* 0 *for the unit,* $-g$ *for the inverse of* g *and* ng *for* G^n *(*$n \in \mathbf{Z}$*).*

It follows from the commutativity of addition that we have the equation

$$n(g+h) = ng + nh \quad (n \in \mathbf{Z}, g, h \in G).$$

(Of course it is not true that $(gh)^n = g^n h^n$ *in an arbitrary group).*

The classical examples of abelian groups are $\mathbf{Z}$*,* $\mathbf{Q}$*,* $\mathbf{R}$ *and* $\mathbf{C}$ *(all with addition) resp.* $\mathbf{Q} \setminus \{0\}$*,* $\mathbf{R} \setminus \{0\}$ *and* $\mathbf{C} \setminus \{0\}$*, with multiplication. The most important finite abelian groups are the cyclic ones i.e. the groups* $C(n)$ *(*$n \in \mathbf{N}$*). In fact, the main result of this section will show that all finite abelian groups can be obtained from groups of this type in a very simple manner. We begin with the remark that general cyclic groups can be split up into products of such groups whose order is a power of a prime:*

Proposition 42 *Let* n *be a positive integer with prime factorisation*

$$n = p_1^{\alpha_1} \dots p_k^{\alpha_k}.$$

Then $C(n)$ *is isomorphic to the product* $C(p_1^{\alpha_1} \times \dots \times C(p_k^{\alpha_k})$*.*

This is a special case of the following result.

Proposition 43 *Let* $m = m_1 \dots m_k$ *be a product of mutually prime numbers* $m_1, \dots, m_k$*. Then*

$$C(m) \simeq C(m_1) \times \dots \times C(m_k).$$

Note that groups of the form $C(p^\alpha)$ *do not split up into smaller ones. For example, if* p *is prime, then* $C(p^2)$ *is not isomorphic to* $C(p) \times C(p)$*. This can be seen by examining the corresponding multiplication tables. The simplest case is* $p = 2$ *where* $C(2) \times C(2)$ *is not isomorphic to* $C(4)$ *(the former is the Klein four group).*

Our main result in this section will be a complete description of all finite abelian groups–indeed all finitely generated ones i.e. those of the form $\langle g_1, \dots, g_n \rangle$*. This means that each element in the group has a representation* $k_1 g_1 + \dots + k_n g_n$ *where the* k_i *are integers. Such a family is called a* **basis** *if* *this representation is unique i.e. if whenever $k_1 g_1 + \cdots + k_n g_n = 0$, then the k_i vanish. Of course, this is reminiscent of the corresponding concepts for vector spaces. However, things are rather more delicate here since we cannot divide by the coefficients. Nevertheless we can prove that every finitely generated abelian group has a basis. The proof used the following result on integral matrices:*

Lemma 5 *Let $k_1, \dots, k_n$ be integers with greatest common divisor 1 (i.e. the integers are jointly prime). Then there is an $n \times n$ matrix over the integers with determinant 1 whose first row is $(k_1, \dots, k_n)$.*

Proof. *The proof is by induction on n. The case $n = 1$ is trivial as usual but it is instructive to consider the case $n = 2$ also. Then we are assuming that k_1 and k_2 are relatively prime and so there are integers s and t with $sk_2 + tk_2 = 1$. Then*

$$\begin{bmatrix} k_1 & k_2 \\ t & -s \end{bmatrix}$$

is a matrix with the desired properties.

For the case n, we let d denote the greatest common denominator of $(k_1, \dots, k_{n-1})$. Then if we apply the induction hypothesis to $\frac{k_1}{d}, \dots, \frac{k_{n-1}}{d}$ we get an $(n-1) \times (n-1)$ matrix of the form

???

with determinant d. Now there are integers s and t so that $sk_n + td = 1$. Then

$$\begin{bmatrix} a \end{bmatrix}$$

is the required matrix (the choice of sign in the last row depends on the order of the matrix).

■

We shall this result in order to verify the following observation. let $\{g_1, \dots, g_n\}$ be a set of generators for the abelian group G and let $k_1, \dots, k_n$ be a jointly prime family of integers. Then there is a set of generators whose first element is $k_1 g_1 + \cdots + k_n g_n$. For suppose that A is the matrix whose existence is proved above. Then A_{-1} is also a matrix over the whole numbers. if we define elements $\tilde{g}_1, \dots, \tilde{g}_n$ by the equation

$$A \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} = \begin{bmatrix} \tilde{g}_1 \\ \vdots \\ \tilde{g}_n \end{bmatrix}$$

so that

$$A^{-1}\begin{bmatrix}\tilde{g}_1\\ \vdots\\ \tilde{g}_n\end{bmatrix}=\begin{bmatrix}g_1\\ \vdots\\ g_n\end{bmatrix}$$

These equations mean that the $\tilde{g}_i$'s are suitable combinations of the g's (which we knew already) and conversely. This of course implies that the $\tilde{g}$'s also generate G.

We can now state and prove our main result:

Proposition 44 *Every finitely generated abelian group has a basis.*

PROOF. *Amongst all sets of generators of G there is (at least) one with the smallest number of generators. Let this number be n and denote this basis by $g_1, \dots, g_n$. The proof is by induction on n. if $n = 1$, then G is cyclic and the result is clear.*

Now among all families of such sets of generators there is one for which the order of the last element is smaller than the orders of all other members of sets of generators with n elements. We suppose that our generators are chosen so that g_n has this property. Now let

$$G_1 = \langle g_1, \dots, g_{n-1}\rangle \quad G_2 = \langle g_n\rangle.$$

Then, by the induction hypothesis, G_1 has a basis (as does G_2 of course). hence it suffices to show that $G_1 \cap G_2 = \{o\}$. For then G is the Cartesian product of G_1 and G_2 and it is clear that the product of two groups with basis has a basis.

Suppose then that this is ot the case. This means that we have a non-trivial relation of the form

$$k_1 g_1 + \dots k_{n-1}g_{n-1} - k_n g_n = 0.$$

Then if d is the greatest common divisor of $(k_1, \dots, k_n)$, d is a divisor of k_n and the element

$$\frac{k_1 g_1}{d} + \dots + \frac{k_{n-1}g_{n-1}}{d} = \frac{k_n g_n}{d}$$

is, by the above, a member of a set of n generators. But this element has order $\leq d < k_n$ which contradicts the minimal property of g_n.

■

Note that if m_r is the order of the basis element g_r (of course, m_r can be infinite), then this result means that G is isomorphic to the product

$$C(m_1) \times \cdots \times C(m_n).$$

Using the result proved above we can reduce each of these cyclic groups to products of groups of the form $C(p^\alpha)$. Hence we have the following description of the structure of finitely generated abelian groups. Let G be such a group. Then there is a finite sequence $p_1, \dots, p_r$ of primes and an $s \in \mathbf{N}$ so that

$$G \simeq G_1 \times \cdots \times G_r \times \mathbf{Z}^s$$

where each G_i is an abelian group of order $p_i^{\alpha_i}$ for some $\alpha_i \in \mathbf{N}$. Further G_i has a representation of the form

$$G_1^i \times \cdots \times G_{r_i}^i$$

where $G_k^i = C(p_i^{n_k^i})$ and $n_1^i \leq n_2^i \leq \cdots \leq n_k^i$ and $\sum_k n_k^i = n_i$.

We note that this representation of the group G is unique (up to the order of the factors), a fact which we shall not prove here.

3.3 Rings

A **ring** is a set R with two operations, called addition and multiplication, and written as

$$(x,y) \mapsto x+y \quad \text{resp.} \quad (x,y) \mapsto xy$$

so that the following holds:

- R is an abelian group (with unit 0) under addition;
- multiplication is associative i.e. $(xy)z = x(yz)$ $(x,y,z \in R)$;
- multiplication is distributive over addition i.e.

$$(x(y+z) = xy + xz \quad (x+y)z = xz + yz \quad (x,y,z \in R).$$

We have already met several examples of rings. The basic one is the set $\mathbf{Z}$ of whole numbers. Further examples are $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{C}$. More elaborate examples are:
I. $M_n(\mathbf{R})$ (resp. $M_n(\mathbf{C})$), the set of $n \times n$ (resp. complex) matrices.
II. The quaternions: this is the set of 2×2 complex matrices of the form

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

with not both a and b non-zero.
III. The polynomials (over **R** *or* **C***).*

These are all provided with the usual operations of addition and multiplication.

unit *for a ring is an element e so that $ex = xe = x$ for each $x \in R$. All of the above examples have units. The ring is* **commutative** *if $xy = yx$ for $x, y \in R$. The typical example of a non-commutative ring is M_n (for $n \geq 2$ of course). We note here that if R is commutative, then we have the following identities*

$$x^2 - y_2 = (x - y)(x + y) \tag{180}$$

$$(x + y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r \quad (n \in \mathbf{N}). \tag{181}$$

The proofs are exactly as for the classical case of real numbers. In non-commutative rings, one must be careful in the use of such identities.

A subset R_1 of a ring is a **subring** *if it is closed under addition and multiplication. it is then a ring in its own right. A subset I is an* **ideal** *if it is closed under addition and we also have*

$$ax \in I \quad and \quad xb \in I \quad for \quad x \in I, a, b \in R.$$

The motivation for the latter definition lies in the fact that we can take quotients of rings modulo ideal. More precisely, if I is an ideal and R/I denotes the set of cosets of R (in the sense of the additive structure of R), then we can define a multiplication on the latter by putting

$$(x + I)(y + I) = xy + I.$$

With this multiplication, R/I is a ring. The classical example of such a quotient is $\mathbf{Z}_m$. Here the ideal which is factored out is that set of all integers which are divisible by m. This is an example[le of a so-called **principal ideal***. These are defined as follows: let R be a commutative ring and a a non-zero element. Then $R = \{ab : b \in R\}$ is easily seen to be an ideal and it is called the* **principal ideal** *generated by a. it is an easy consequence of our discussion in ?? that the only ideals in* **Z** *are those of this type. The same holds for the ring of polynomials over* **R** *or* **C** *as we shall see shortly.*

A commutative ring is called an **integral domain** *if the product xy of two non-zero elements is never zero. Some of the simple result on divisibility can be interpreted as stating that $\mathbf{Z}_m$ is an integral domain if and only if m is prime.*

If R is a ring with unit, then not every element need have an inverse (the latter is defined in the obvious way). Those which have are called **units**. *In M_n this coincides with the usual notion of invertibility for matrices.*

A ring with unit in which every non-zero element is invertible is called a **skew-field**. *If it is, in addition, commutative, it is called a* **field**. *Examples of fields are* $\mathbf{Q}$, $\mathbf{R}$ *and* $\mathbf{C}$. *If p is a prime, then $\mathbf{Z}_p$ is a finite field. The quaternions provide an example of a skew-field which is no a field.*

If R and R_1 are rings, a **(ring) homomorphism** *from R into R_1 is a mapping $\phi : R \to R_1$ so that*

$$\phi(x+y) = \phi(x) + \phi(y) \quad \phi(xy) = \phi(x)\phi(y) \quad (x, y \in R).$$

if R and $R-1$ have units, we shall tacitly assume that $\phi(e) = eL$. We remark that this does not follows from the multiplicativity.

R and R_1 are **isomorphic** *if there is a bijection ϕ from R onto R_1 so that both ϕ and ϕ^{-1} are homomorphisms.*

If $\phi : R \to R_1$ is a homomorphism, then its kernel

$$Ker\phi = \{x \in R : \phi(x) = 0\}$$

is an ideal in R. In this context, we have the following analogue of the first isomorphism theorem for groups (the proof is virtually identical):

Proposition 45 *Let $\phi RtoR_1$ be a surjective homomorphism. Then $R/Ker\phi = R_1$.*

In the context of rings, constructions which are almost identical to those for groups allow us to define

- *the Cartesian product $\prod_{\alpha\in A} R_\alpha$ of a family of rings;*
- *the direct sum $\oplus_{\alpha\in A} R_\alpha$ of a family of rings;*
- *the free ring $R(s)$ over a set S.*

The reader is invited to carry out these constructions in investigate their elementary properties.

Examples of rings: *One rich source of examples of rings is the following construction. G is a group and R is a ring (to be concrete, it is useful to concentrate on the case where the ring R is $\mathbf{Z}$. In any case, this is the most important example). As a set, the new ring, which we denote by $R(G)$, is the direct sum $\oplus_{g\in G}R_g$ where each R_g is a copy of R. It is useful and suggestive to write $\sum_{g\in G} r_g g$ for the generalised sequence $\jmath(r_g)_{g\in G}$ (of course, only a finite number of elements in this sum are non-zero). Then we define addition and multiplication in the natural way by putting*

$$(\sum r_g g) + (\sum s_g g) = \sum (r_g + s_g) g$$

and

$$(\sum_g r_g g) \cdot (\sum_h s_h h) = \sum_{g,h} r_g \cdot s_h (gh).$$

The reader can check that this is a ring. It is commutative if and only if both R and G are.

The particular case where $r = \mathbf{Z}$ produces the so-called **group ring $\mathbf{Z}(G)$** *of G. its elements can be regarded as formal integral combinations of the elements of G i.e. as formal sums of the form $\sum_{g\in G} n_g g$. This group has the following universal property.*

3.4 Matrices over rings

It is rather trivial (but will prove to be useful) to generalise some of the basic topics of linear algebra by replacing the fields $\mathbf{R}$ and $\mathbf{C}$ by a general ring R. We consider briefly matrices and polynomials over general rings. if R is a ring, then an $m \times n$ matrix over R is an array

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

whereby each $a_{ij} \in R$ (more pedantically, a matrix is a mapping from the set

$$\{(i,j) : i \in \{1,\dots,m\}, j \in \{1,\dots,n\}\}$$

into R).

We can define the sum of two $m \times n$ matrices over R resp. the product of an $m \times n$ with an $n \times p$ matrix just as in the real case. Then the family $M_n(R)$ of $n \times n$ matrices over R is itself a ring.

If R is commutative, we can define the **determinant** *of the $n \times n$ matrix by the formula*

$$\det A = \sum_{\pi \in S_n} \epsilon_\pi a_{1\pi(1)} \dots a_{n\pi(n)}$$

just as in the real case. The following facts hold in the general situation:

$$\det A = \sum_k (-1)^{i+k} a_{ik} \det A_{ik} \tag{182}$$

$$= \sum_i (-1)^{i+k} a_{ik} \det A_{ik} \tag{183}$$

where A_{ik} is the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the k-th column of A. As a consequence we have the formula

$$A \cdot (adj\, A) = (adj\, A) \cdot A = \det A \cdot I_n$$

where $adj\, A$ is the matrix $[(-1)^{i+j} \det A_{ji}]$. Hence A is invertible if and only if $\det A$ is a unit of R and in this case we have the formula

$$A^{-1} = (\det A)^{-1} \cdot adjA.$$

3.5 Rings of polynomials

In a similar manner, we can generalise the notion of a polynomial by introducing the space $R[t]$ of polynomials over a ring R. A typical element has the form

$$p(t) = a_0 + a_1 t + \dots + a_n t^n \quad (a_0, \dots, a_n \in R).$$

if the leading coefficient a_n is one (we are assuming that R has a unit), then the polynomial is **monic**. *The set $R[t]$ itself forms a ring, where multiplication and addition are defined as usual. In particular, the product of polynomials p and q whereby*

$$p(t) = a_0 + a_1 t + \dots + a_n t^n$$

and

$$q(t) = b_0 + b_1 t + \dots + b_m t^m$$

is the polynomial

$$c_0 + c_1 t + \dots + c_{n+m} t^{n+m}$$

where $c_k = \sum_{r+s=k} a_r b_s$.

At this point, it is perhaps appropriate to remark that if R is not an integral domain, then the latter polynomial need not have degree $m+n$ (since C_{n+m} can vanish, even if a_n and b_m do not).

Note that the concepts of polynomials with coefficients in a matrix ring and of matrices over a polynomial ring are essentially the same. For example, in the case of real polynomials, the object

$$\begin{bmatrix} 2+3t+t^2 & 1-t & 1+t \\ t^2-1 & 2t & 1 \\ 1 & t^2-1 & 3 \end{bmatrix}$$

can be regarded as a matrix whose elements are polynomials or as the polynomial

$$\begin{bmatrix} 2 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -1 & 3 \end{bmatrix} + \begin{bmatrix} 3 & -1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} t + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} t^2$$

whose coefficients are matrices.

We now consider the division algorithm for polynomials over rings with a unit. Since we are not assuming that the ring is commutative (in our applications it will be a matrix ring), more care is required than in the classical case. For example, if p is a polynomial

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n$$

in $R[t]$ and $x \in R$, there are two possibilities for substituting x for t in p, namely

$$a_0 + a_1 x + \cdots + a_n x^n$$

(on the right) or

$$a_0 + x a_1 + \cdots + x^n a_n$$

(on the left).

It is customary to indicate in the notation which type of substitution is being used. however, we shall only substitute on the right and so we shall simply use the symbol $p(x)$ which will denote the first of the above two formulae. If p is the polynomial there and $a_n \neq 0$, then n is the **degree** *of p (written deg(p)). Then we have the simple inequalities*

$$deg(p+q) \leq \max(degp, degq)$$

and

$$degpq \leq degp + degq.$$

As noted above, we need not have inequality in the second inequality, in contrast to the classical case. However, we do have if, for example, R is a field–more generally, if one of the leading coefficients is invertible). It follows immediately from this that if R is a division ring then a polynomial is invertible in $R[t]$ if and only if it is constant (i.e. an element of R) which is invertible in R.

In $R[t]$, the division algorithm takes on the following form: let p and q be polynomials whereby the leading coefficient of q is invertible. Then there are unique polynomials s and r so that

$$p = qs + r \quad \textit{where} \quad degr < degq.$$

(Note that we also have a corresponding formula $p = s_1 q + r_1$ for division on the right. However, there is no reason why the remainders r and r_1 should be the same if R is not commutative).

We would like to deduce the usual criteria for an element x of the ring to be a root of the polynomial i.e. for $p(x)$ to vanish. In doing so, we shall have to be careful about calculating expression such as $pq(x)$ (i.e. the substitution of x in the product pq. This will not, in general, be the same as $p(x)q(x)$. We merely have the formula

$$pq(x) = a_0 + a_1 q(x) + \cdots + a_n q(x)^n$$

where p is the polynomial

$$p(t) = a_0 + \cdots + \_nt^n.$$

In particular, if $q(x) = 0$, then $pq(x) = 0$. (There is no reason why $pq(x)$ should be zero if $p(x) = 0$. This apparent assymetry arises from the fact that we are substituting on the right.

If we now apply the division algorithm to a polynomial p (with degree at least one) and the divisor $t - x$ (where x is an element of R), then we can write

$$p(t) = q(t)(t - x) + r$$

where $r \in R$ (we are dividing by $(t - x)4$ on the **right***). Substituting x in both sides and using the above remarks, we see that $r = p(x)$. hence we see that the linear polynomial $t - x$ is a right factor of p if and only if $p(x) = 0$. (Of course, it is a left factor if and only if we get zero when we substitute x in p on the left).*

As an application, we give a proof of the Cayley-Hamilton theorem which does not use any of the canonical forms and is valid for any commutative ring:

PROOF. *proof A is an $n \times n$ matrix over a commutative ring R with unit. As in the real case, the characteristic polynomial is defined by the formula*

$$\chi_A(t) = \det(tI - A).$$

of course, this is a polynomial of degree n over R. As we know, we have the equation

$$\begin{aligned} adj(tI - A)(tI - A) &= \det(tI - A) \cdot I & (184) \\ &= \chi_A(t) \cdot I. & (185) \end{aligned}$$

This is an equation in $M_n(R[t])$ i.e. between ;matrices whose elements are polynomials over R (or between polynomials whose coefficients are matrices over R). In particular, this means that the linear polynomial $(tI - A)$ is a right factor of the polynomial $\chi_A(t) \cdot I$ and so, by the above, $\chi_A(A) = 0$.

■

We shall require some results on the factorisation of polynomials. The formulations and proofs are so very similar to those that we have studied for the integers that, rather than duplicate the proofs in this new setting, we prefer to adapt a more abstract approach which will allow us to subsume both sets of result in a common framework. Our policy will be not to repeat the proofs of results where this involves the mere translation of the proofs for the more concrete situation into this abstract language.

We begin with a definition which embodies the essentials of the division algorithm:

Definition: *A* **euclidean domain** *is a commutative integral domain R with unit, together with a function $d : R \setminus \{0\} \to \mathbf{N}$ so that*

- *$d(a) = 0$ if and only if $a = 0$;*
- *$d(ab) = d(a) + d(b)(a, b \neq 0)$;*
- *if a and b are non-zero elements of R, then there exist q and r in R so that $a = qb + r$ and $r = 0$ or $d(r) < d(b)$.*

of course, important examples are $\mathbf{Z}$ itself (where $d(n) = |n|$) and the ring of polynomials over $\mathbf{R}$ (or, more generally, a field) where $d(a)$is the degree of a.

Proposition 46 *If R is a euclidean domain, then every ideal in R is a principal ideal.*

Proof. *Let I be an ideal in R which we suppose to be non-trivial (i.e. $I \neq \{0\}$). Now we choose a non-zero element x of I with the smallest degree of any non-zero element of the ideal. We claim that $I = xR$ i.e. it is the principal ideal generated by x. To prove this, we suppose that $y \in I$ and show that y is a multiple of x. By the division algorithm, $y = qx + r$ where $r = 0$ or $d(r) \leq x$. But in the latter case we have $r = y - qx$ is an element of I and this contradicts the minimal property of x.*

∎

This result allows us to define the g.c.d. of a collection of elements of a euclidean domain. For if we denote by I the set of elements of the form

$$r_1 a_1 + \cdots + r_n a_n \quad (r_1, \dots, r_n \in R)$$

for a given sequence $(a_1, \dots, a_n)$ in R, then it is clear that this is an ideal–in fact, the smallest one which contains the a's. Now by the above result, this is a principal ideal, generated by an element x say. This x is called a greatest common divisor of $a_1, \dots, a_n$ and written g.c.d.$(a_1, \dots, a_n)$.

At this point, we remark that this g.c.d. is, in contrast to the situation in $\mathbf{Z}$*, not uniquely determined by the a's..*

If we apply the above result to the ring of polynomials (over $\mathbf{R}$ *or* $\mathbf{C}$*, say), then we see that if p and q are non-zero polynomials, then there exists a greatest common divisor d for p and q. In this case d is unique (up to a scalar factor) and so there is precisely one monic polynomial with this property. d can be written in the form $d = pr + qs$ for suitable polynomials r and s.*

The fact that every ideal of the ring of polynomials is a principal ideal can be used to give a slick proof of the existence of minimal polynomials. Suppose that f is a linear operator on the (complex) vector space V. Then there is a natural mapping $p \mapsto p(f)$ of substitution which is a ring homomorphism from the ring of polynomials into the ring $L(V)$. its kernel is thus an ideal and so has the form $\{pm : p \in \mathbf{C}[t]\}$ for some monic polynomial m. m is then the minimal polynomial of f.

We now introduce a concept for euclidean domains which is the natural generalisation of that of prime numbers.

Definition: *An element x of a euclidean domain R is said to be* **irreducible** *if it cannot be factorised in the form $x = x_1 x_2$ except for the trivial case where one f the elements, say x_1 is invertible and $x_2 = x_1^{-1} x$. Of course* *the irreducible elements of* $\mathbf{Z}$ *are just the primes. more interesting is the case of polynomials. The irreducible polynomials over* $\mathbf{C}$ *are the linear ones, while in* $\mathbf{R}[t]$ *they are the linear ones and the quadratic ones of the form* $at_t+2bt+c$ *which have no real roots (i.e. with negative discriminant).*

By aping the proof of the fundamental theorem of arithmetic, one can prove the following:

Proposition 47 *Let* x *be an element of a euclidean domain. Then* x *has a representation*

$$z = p_1^{\alpha_1} \dots p_n^{\alpha_n}$$

as a product of powers of irreducible elements. This factorisation is unique in the sense that if

$$x = q_1^{\alpha_1} \dots q_k^{\alpha_k}$$

then $k = n$ *and there is a permutation* π *of* $\{1, \dots, n\}$ *and a finite sequence* $u_1, \dots, u_n$ *of units so that* $q_i = u_i p_{\pi(i)}$ *for each* i*.*

The important application of this result if to rings of polynomials where it states that every polynomial over a field can be expressed as a product of irreducible ones.

We shall now show how to use these rather abstract results to obtain sharper ones on canonical forms for matrices. We shall consider matrices over a field K *(the reader who prefers to remain in the more concrete situation of the real or complex field will losing nothing essential if he reads* $\mathbf{R}$ *of* $\mathbf{C}$ *for* K*).*

We begin by introducing an equivalence relationship for matrices of polynomials which is the exact analogue of the one considered in ??? for real matrices.

Definition: *We say that two matrices* A *and* B *(both* $m \times n$*) over* K *are* **equivalent** *if we can transform* A *into* B *by means of a finite sequence of operations of the following types:*

- *exchanging rows or columns;*
- *multiplying a row (or column) by a non-zero scalar (i.e. an element of* K*);*
- *adding* p *times row* i *(resp. column* i*) to row* j *(resp. column* j*) (*p *a polynomial).*

As in the scalar case, these operations are implemented by matrices of the following types:

We shall see shortly that this relationship can be restated in the more natural form that there are invertible matrices P and Q over $K[t]$ (where P is $m \times m$ and Q is $n \times n$) so that $PAQ = B$. (we remark that a matrix of polynomials is invertible if and only if its determinant is a non-zero scalar).

Using suitable refinements of the methods of ??? we can prove the following result:

Proposition 48 *Every $m \times n$ matrix A over $K[t]$ is equivalent to one of the form*

$$\left[\begin{array}{c} p_1 \end{array} \right]$$

where each p_i is a monic polynomial and $p_i | p_{i-1}$ for each i. Moreover the polynomials p_i are uniquely determined by A.

Proof. *We begin by considering all matrices which are equivalent to A. Amongst the elements of all such matrices, there is one with lowest degree. There is no loss of generality if we suppose that this is an element of A and by employing suitable row and column exchanges we can arrange for it to be a_{11}. Then we claim that a_{11} is a divisor of the remaining elements of the first row and column of A. For if this is not the case–say if a_{11} does not divide a_{21}, then by the division algorithm, $a_{21} = qa_{11} + r$ where $d(r) < d(a_{11}$. Then we can subtract q times the first row from the second and so obtain a matrix which is equivalent to A and has the polynomial f in the $(2, 1)$-th place. This contradicts the minimal property of a_{11}.*

We now proceed exactly as in the scalar case to obtain a matrix of the form

$$\left[\begin{array}{c} a_{11} \end{array} \right]$$

which is equivalent to A. At this point, we note that a_{11} divides each element of B.

The proof is now completed by an obvious induction argument.

Uniqueness:

■

Since the (p_i) are uniquely determined by A we are justified in calling them the **invariant factors** *of A. Thus tow matrices A and B are equivalent if and only if they have the same invariant factors.*

If we consider the special case of an $n \times n$ matrix and not that the P and Q obtained in the proof are products of elementary matrices (i.e. those which implement the elementary row and column operations), we can show *that every invertible matrix in $F[t]$ is a product of such matrices. For suppose that A is invertible and that*

$$PAQ = \left[\ p_1\ \right].$$

Then of course, the diagonal matrix is also invertible which means that each p_i is invertible. Now the only invertible polynomials are the constant ones and since they are monic we must have that $p_i = 1$. Thus the right hand side is the unit matrix and so $A = P^{-1}Q^{-1}$ is a product of elementary matrices.

It follows that two polynomial matrices A and B are equivalent if and only if there exist invertible matrices P and Q of suitable dimensions over $K[t]$ so that $PAQ = B$.

The rational canonical form *:*

We now use the above theory to deduce a result on canonical forms which applies to matrices over a general field. If

$$p(t) = a_0 + a_1 t + \cdots + t^n$$

is a monic polynomial over a field K, the matrix

$$C(p) = \left[\ 0\ \right]$$

is called the **companion matrix** *of p. For example the standard circulant matrix circ$(0, 1, 0, \dots, o)$ is the companion matrix of $t^n - 1$. The characteristic polynomial of $C(p)$ is precisely p as the reader can easily verify. The invariant factors are $1, 1, \dots, p$. This follows from the fact that for $i < n$ we have an i-minor of the matrix*

$$tI - C(p) = \left[\ t\ \right]$$

with determinant ±1 (namely the minor

$$\left[\ -1\ \right].$$

Hence $d_i = 1$ for $i < n$. Also $d_n - \det(tI - C(p)) = p(t)$.

From the previous criteria for similarly, we can deduce the following result:

Proposition 49 *Let A be an $n \times n$ matrix over a field K with invariant factors $p_1, \dots, p_r$. Then A is similar to the matrix*

$$\left[\ C(p_1)\ \right].$$

For the proof it suffices to show that A has the same invariant factors as the above and this is clear.

We can refine there considerations still further. Let A be a matrix with invariant factors $p-1, \dots, p_r$. Then each p_i has a unique factorisation as a product of powers of irreducible polynomials, say

$$p_i = q_{i1}^{\alpha_1} \dots q_{is_i}^{\alpha_{s_i}}.$$

The list $(q_{11}^{\alpha_1}, \dots, q_{rs_r}^{\alpha_{s_i}})$ of all these factors is called the set of **elementary divisors** *of A. Note that the irreducible factors of p_1 occur in those of p_2 and so on. The list is therefore made with repetitions.*

If we know the rank of a matrix, we can reconstruct the list of invariant factor from the elementary divisors as follows:

Proposition 50 *Let A be an $n \times n$ scalar matrix with elementary divisors $q_1, \dots, q_n$. Then A is similar to the matrix*

$$\left[\ C(q_1) \ \right].$$

It follows from this that two matrices with the same rank are similar if and only if they have the same list of elementary divisors.

We can use these results to obtain a criterium for the similarly of matrices over a field as follows:

Lemma 6 *Let A and B be $n \times n$ matrices over the field K. Then A and B are similar (over K) if and only if the polynomial matrices $tI - A$ and $tI - B$ are equivalent over $K[t]$.*

PROOF. *Suppose that A and B are similar, say $B = P^{-1}AP$. Then*

$$tI - B = P^{-1}(tI - A)P$$

and so the polynomial matrices are equivalent.

Suppose now that there are invertible matrices $P(t)$ and $Q(t)$ over $K[t]$ so that

$$tI - B = P(t)(tI - A)Q(t).$$

Then

$$P(t)^{-1}(tI - B) = (tI - A)Q(t).$$

This means that $(tI - B)$ is a right factor of the right-hand side and so substitution of B leads to the zero matrix i.e.

$$(B - A)Q(B) = 0 \quad \textit{or}$$

hence if $P = Q(B)$, we see that $PB = AP$. Thus it suffices to show that P is invertible. But $Q(t)$ is invertible as a matrix over $K[t]$ and so there is a matrix $R[t]$ so that

$$R(t)Q(t) = I = Q(t)R(t).$$

Substitution of B in this equation shows that $R(B)$ is an inverse for $Q(B)$.

■

Since the similarity properties of A are related to the equivalence properties of the matrix tI_A which are in turn determined by the elementary divisors of the latter, it is natural to define the **elementary divisors** *of the scalar matrix A to be those of the polynomial matrix tI_A.*

Then we can summarise the above results as follows:

Proposition 51 *Two $n \times n$ matrices A and B over the field K are similar if and only if they have the same rank and the same elementary divisors.*

3.6 Field theory

If E is a subfield of a filed F (in which case we sometimes call F an **extension** *of E), we may regard F as a vector space over E. Typical examples are* **C** *which is a two-dimensional vector space over* **R** *and* **R** *which is an infinite-dimensional vector space over* **Q**. *If F is finite dimensional over E, we say that it is a* **finite extension** *and the dimension of F as a vector space over E is called the* **degree** *of the extension and denoted by F/E.*

Proposition 52 *If F is a finite extension of E and G is a finite extension of F, then G is a finite extension of E and we have the equality*

$$G/E = (G/F) \cdot (F/E).$$

PROOF. *Let $(x_i)_{i=1}^n$ be a basis for F as a vector space over E and $(y_j)_{j=1}^m$ be a basis for G over F. The proof consists of the simple computations required to show that $(x_i y_j)_{i,j}$ is a basis for G over E.*
Existence of representations: Let $x \in G$. Then x can be written as a sum $\sum_j a_j z_j$ where the $a)_j$ are in F. These can in turn be written as

$$a_j = \sum_i b_{ij} x_i$$

where the b_{ij} are in E. Then

$$x = \sum_j a_j y_j = \sum_{i,j} b_{ij} x_i y_j$$

as required.
Linear independence: Suppose that the a_{ij} are coefficients in E so that $\sum_{i,j} a_{ij} x_i y_j = 0$. Then

$$\sum_j (\sum_i a_{ij} x_i) y_j = 0$$

and so $\sum_i a_{ij} x_j = 0$ for each j by the linear independence of the y_j. The linear independence of the x's now implies that each of the a_{ij} vanishes.

■

The usual way of obtaining extensions is by adding elements to subfields as follows: let E be a subfield of F and suppose that $x \in F \setminus E$. Then we introduce the notation $E(x)$ for the smallest subfield of F which contains x. This consists of all elements of the form $\frac{p(x)}{q(x)}$ where p and q and polynomials over E and x is not a zero of q.

A typical example is the subfield $\mathbf{Q}(\sqrt{2})$ of $\mathbf{R}$. The reader will observe that the degree of this extension is 2. On the other hand, if we add to $\mathbf{Q}$ an irrational number which is transcendental (i.e. is not a solution of a polynomial equation with integer equations–π is such a number), then we get an extension which is not finite.

For the general situation we make the following definition. If E is a subfield of F, then an element x of F is **algebraic** *(over E) if there is a polynomial p over E so that $p(x) = 0$. otherwise x is* **transcendental**.

In general, a polynomial over a field need not split into linear factors as the examples

$$t^2 - 2 \quad (\textit{over } \mathbf{Q}) \quad \textit{and} \quad t^2 + 1 \quad \textit{over } \mathbf{R})$$

show. However, we know that we can enlarge both fields to obtain one in which they do have such a factorisation. In the first case we can take $\mathbf{Q}(\sqrt{2})$, in the second case the complex field (indeed, the complex field is introduced for precisely this reason). These are special cases of a general result whose proof is an abstract form of one of the possible constructions of the complex field. We shall show that if p is a non-constant polynomial without a zero over a field E, then we can find an extension F of E in which p contains a linear factor. Hence we can continue this process a finite number of times and eventually obtain an extension in which p splits into linear factors.

We begin the construction with the remark that we can assume that p is irreducible (if not we begin with an irreducible factor of p). We consider the

quotient ring $E[t]/(p)$ which is obtained by factoring out the principal ideal generated by p in the ring of polynomials over E. We claim that this is a field with the required properties.
PROOF.

■

Using this Lemma, we can deduce the existence of a so-called **splitting field** *of a polynomial.*

Proposition 53 *Let p be a non-constant polynomial over a field E. Then there exists an extension F of E with the following properties:*

- *in F p has a factorisation $p(t) = a(t - x_1) \ldots (t - x_n)$ as a product of linear factors where $n = degp$ and $a \neq 0$;*

- *$F = E(x_1, \ldots, x_n)$.*