MATH3001 RINGS AND FIELDS

DR. G.BERHUY

Contents

Date: Academic year 2006/2007.

1. Basic ring theory

1.1. **The definition.**

Definition. A *ring* is a set R equipped with two binary operations

$$+ : R \times R \to R, (a,b) \mapsto a+b,$$

$$\cdot : R \times R \to R, (a,b) \mapsto a \cdot b,$$

a unary operation

$$- : R \to R, a \mapsto -a,$$

and elements 0_R and 1_R such that the following properties hold:

(1) $(R, 0_R, +, -)$ is an abelian group, that is for all $a,b,c \in R$, we have :
   (a) $a + b = b + a$
   (b) $(a+b)+c = a+(b+c)$
   (c) $a + 0_R = 0_R + a = a$
   (d) $a + (-a) = 0_R$

(2) Multiplication is associative:

$$a.(b.c) = (a.b).c \text{ for all } a,b,c \in R$$

(3) 1_R is the neutral element for multiplication:

$$a.1_R = 1_R.a = a \text{ for all } a \in R$$

(4) Multiplication is distributive over addition:

$$a.(b+c) = a.b + a.c \qquad (b+c).a = b.a + c.a \text{ for all } a,b,c \in R$$

A *commutative ring* is a ring R satisfying $a.b = b.a$ for all $a, b, \in R$.

Note that we have written $a+b$ for the first operation applied to the pair (a,b) and $a.b$ for the second operation applied to (a,b). Often we will omit the . from the notation altogether, and write ab instead of $a.b$ for the result of applying . to the pair (a,b).

The notation is the same as that used for ordinary arithmetic. This is no coincidence, but a reflection of the fact that "everyday" objects such as the integers, $\mathbb{Z}$, the rational numbers, $\mathbb{Q}$, and the real numbers, $\mathbb{R}$, with their usual sum and product operations are examples of rings. Note however that the natural numbers $\mathbb{N}$ do not form a ring, because the unary operation $-$ is not defined on them.

1.2. **First examples.**

As mentioned above, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are commutative rings.

Another important example is the ring $\mathbb{Z}/n\mathbb{Z}$ for $n \geq 1$. Elements of $\mathbb{Z}/n\mathbb{Z}$ are equivalence classes of integers, under the relation given by $a \sim b$ if and only if n divides $a-b$ (You may like to think of these equivalence classes as the cosets of the additive subgroup of $\mathbb{Z}$ generated by n.) This is denoted by $a \equiv b[n]$.

In other words, the elements of $\mathbb{Z}/n\mathbb{Z}$ are the **sets** $n\mathbb{Z}, 1+n\mathbb{Z}, \cdots, (n-1)+n\mathbb{Z}$.

Writing $\overline{a}$ for the equivalence class of a, the ring operations are defined by

$$\overline{a}+\overline{b}=\overline{a+b}, -\overline{a}=\overline{-a}, \overline{a}.\overline{b}=\overline{a.b}.$$

The zero element of this ring is the class $\overline{0}$ containing 0, and the unity in this ring is the class $\overline{1}$ containing 1. The classes of $0, 1, \ldots, n-1$ are all distinct, and any a is equivalent to one of these elements. It follows that $\mathbb{Z}/n\mathbb{Z}$ has exactly n elements.

Of course, we should check that these operations does not depend on the choice of the integers a and b representing $\overline{a}$ and $\overline{b}$. Let us do it for addition:

if $\overline{a'}=\overline{a}$ and $\overline{b'}=\overline{b}$, we have to check that $\overline{a+b}=\overline{a'+b'}$. But by definition of the equivalence relation, we have $a'=a+nm$ for some $m \in \mathbb{Z}$, and $b'=b+nk$ for some $k \in \mathbb{Z}$. We then have $a'+b'=a+b+n(m+k)$, which means that $\overline{a'+b'}=\overline{a+b}$, since $(a'+b')-(a+b)$ is divisible by n.

Another notation for $\mathbb{Z}/n\mathbb{Z}$ is $\mathbf{Z}_n$.

We are now going to define two families of rings, which are of real importance in ring theory. Let $d \in \mathbb{Z}, d \neq 0$ be a square-free integer (that is not divisible by any m^2, $m \in \mathbb{Z}$).

If $d>0$, $\sqrt{d}$ is meant to be the positive square root of d. If $d<0$, $\sqrt{d}$ is meant to be the purely imaginary complex number $i\sqrt{-d}$.

Proposition 1.1. *Let $d \in \mathbb{Z}, d \neq 0$ be a square-free integer. Let $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Q}[\sqrt{d}]$ be the subsets of $\mathbb{C}$ defined by:*

$$\mathbb{Z}[\sqrt{d}]=\{z \in \mathbb{C}\,|z=a+b\sqrt{d} \textit{ for some } a,b \in \mathbb{Z}\},$$

$$\mathbb{Q}[\sqrt{d}]=\{z \in \mathbb{C}\,|z=a+b\sqrt{d} \textit{ for some } a,b \in \mathbb{Q}\}.$$

Then $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Q}[\sqrt{d}]$ are commutative rings for the $0,1,+,.,-$ of $\mathbb{C}$. Moreover, if $d \equiv 1[4]$, the set

$$\mathbb{Z}[\frac{1+\sqrt{d}}{2}]=\{z \in \mathbb{C}\,|z=a+b\frac{1+\sqrt{d}}{2} \textit{ for some } a,b \in \mathbb{Z}\}$$

is also a commutative ring for the $0,1,+,.,-$ of $\mathbb{C}$.

Proof. Left as an exercise. □

Warning: The last part is not true if $d \not\equiv 1[4]$.

For example, the sets $\mathbb{Z}[i], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i\sqrt{2}], \mathbb{Q}[i], \mathbb{Q}[\sqrt{2}], \mathbb{Q}[i\sqrt{2}], \mathbb{Q}[j]$ are commutative rings.

Bonus exercise: Assume that $d \in \mathbb{Z}$ is a square-free integer, such that $d \equiv 1 \mod 4$. Let

$$\mathbb{Z}[\frac{-1+\sqrt{d}}{2}] = \{z \in \mathbb{C} \mid z = a + b \cdot \frac{-1+\sqrt{d}}{2} \text{ for some } a, b \in \mathbb{Z}\}.$$

Show that $\mathbb{Z}[\frac{-1+\sqrt{d}}{2}] = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$.

In particular, $\mathbb{Z}[\frac{-1+\sqrt{d}}{2}]$ is also a ring for the $0, 1, +, -.$ of $\mathbb{C}$.

For example, $\mathbb{Z}[j]$ is a ring, where j is the complex number

$$j = e^{2i\pi/3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}.$$

1.3. **New rings from old.** There are lots of ways to construct new rings from old:

a) Polynomial rings: If R is a ring, then $R[X]$, the ring of polynomials in the indeterminate X, has as elements the finite sequences

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n,$$

for $a_i \in R$, or equivalently sums of the form

$$\sum_{i \geq 0} a_i X^i,$$

where all but finitely many of the a_i are zero. The operations are defined in the usual way:

$$\left(\sum_i a_i X^i\right) + \left(\sum_i b_i X^i\right) = \left(\sum_i (a_i + b_i) X^i\right),$$

$$\left(\sum_i a_i X^i\right) \cdot \left(\sum_i b_i X^i\right) = \left(\sum_i c_i X^i\right),$$

where

$$c_i = a_0.b_i + a_1.b_{i-1} + a_2.b_{i-2} + \cdots + a_i.b_0.$$

The zero element is the polynomial $\sum_i 0X^i$, and the unity element is $1 + \sum_{i>0} 0X^i$.

Note that the polynomial X^n is equal to X multiplied by itself n times using the above definition. This is thus consistent with usual notation.

We write $R[X, Y] = R[X][Y]$ for the ring of polynomials in two unknowns X and Y.

b) Matrix rings: Denote by $M_n(R)$ the set of all $n \times n$ matrices over the ring R. If A is the matrix with entries (a_{ij}), and similarly for B, C, D, then $A + B = C$, and $A.B = D$, where

$$c_{ij} = a_{ij} + b_{ij}, \qquad d_{ij} = \sum_k a_{ik} b_{kj}.$$

c) Products of rings: If R_1 and R_2 are rings, then

$$R_1 \times R_2 := \{(r_1, r_2), r_1 \; R_1, r_2 \in R_2\}$$

is a ring, where the operations are defined componentwise by

$$(r_1, r_2) + (r_1', r_2') = (r_1 + r_1', r_2 + r_2')$$

$$-(r_1, r_2) = (-r_1, -r_2)$$

$$(r_1, r_2).(r_1', r_2') = (r_1.r_1', r_2.r_2')$$

The zero element is $(0_{R_1}, 0_{R_2})$ and the identity element is $(1_{R_1}, 1_{R_2})$.

d) Subrings:

Definition. Let R be a ring. A subset S of R is called a *subring* if S contains 0_R and 1_R, and S is closed under the operations $+, -, .$, that is for all $a, b \in S$, we have $a + b, a.b, -a \in S$.

Proposition 1.2. *Let R be a ring, and let S be a subring of R. Then S is a ring with the same operations and 0 and 1 as R.*

Proof. Left as an exercise. □

The interest of this proposition is that it can be used to prove that a set S is a ring, by proving that it is a subring of a well-known ring, allowing us to check only 3 easy properties instead of a bunch of axioms.

Examples.

(1) $\mathbb{Z}[\sqrt{d}], \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ and $\mathbb{Q}[\sqrt{d}]$ are subrings of $\mathbb{C}$.
(2) $\mathbb{Z}$ is a subring of $\mathbb{Q}$.
(3) $M_n(\mathbb{Z})$ is a subring of $M_n(\mathbb{R})$.
(4) More generally, if S is a subring of R, then $M_n(S)$ is a subring of $M_n(R)$.
(5) R is a subring of $R[X]$, the elements of R being viewed as constant polynomials.

Proposition 1.3. *Let R be a ring. The intersection of a family of subrings of R is a subring of R.*

Proof. Left as an exercise. □

e) Rings of functions: If W is any set, then the set R^W of all functions $f : W \to R$ becomes a ring, where the operations are defined by

$$(f+g)(w) = f(w) + g(w), \qquad (f.g)(w) = f(w).g(w).$$

The zero of this ring is the constant function that sends every element of W to 0_R, and the identity is the constant function that sends every element of W to 1_R.

Other similar examples include the ring of continuous functions from $[0,1]$ to $\mathbb{R}$ (usually denoted by $C([0,1];\mathbb{R})$).

Other constructions, such as factor rings and fields of fractions, will be considered later.

We use familiar notation because many of the usual properties of addition and multiplication hold in all rings. But we shouldn't assume them, because some will not always hold. They should be deduced from the axioms.

Proposition 1.4. *Let R be a ring, and let a, b be elements of R. Then*

i) $a.0 = 0 = 0.a$ *ii)* $a.(-b) = -(a.b) = (-a).b$

iii) $(-1).a = -a$ *iv)* $-(-a) = a$

v) $(-a).(-b) = a.b$ *vi)* $(-1).(-1) = 1$

vii) the identity element 1_R is unique viii) the zero element 0_R is unique.

Example of proof: For viii), since $(R, +, -, 0)$ is an abelian group, we can quote the result that the identity element of a group is unique. Similarly, iv) involves only the additive group structure. For i), since $0 = 0 + 0$,

$$a.0 = a.(0+0) = a.0 + a.0.$$

Now add $-(a.0)$ to each side, and obtain $0 = a.0 + (-(a.0)) = a.0 + a.0 + (-(a.0)) = a.0$. The other equation is similar.

1.4. **Special types of rings.** Let R be a ring in which $1 = 0$. Then R has exactly one element. Indeed, if $a \in R$, then $a = 1.a = 0.a = 0$.

Definition. A ring in which $1 = 0$ is called *the trivial ring.*

Definition. Let R be a ring. We say that R *has zero divisors* if there exist $x, y \in R, x, \neq 0, y \neq 0$ such that $x.y = 0$. In this case, x is called *a left zero divisor* and y is called *a right zero divisor.*

Definition. A ring R is called an *integral domain* if it is non-trivial, commutative and for all $x, y \in R$, we have

$$xy = 0 \Rightarrow x = 0 \text{ or } y = 0$$

Remarks 1.5. (1) A ring R is an integral domain if it is non-trivial, commutative and has no zero-divisors.

(2) Any subring S of an integral domain R is an integral domain. Indeed, a subring of R is a ring itself, and since $1_R \neq 0_R$, S is not trivial either since it has same 0 and 1. Also S is commutative since R is. Finally, assume that $xy = 0$ in S. Viewing this equality in R, we deduce that $x = 0$ or $y = 0$.

Examples.

1. The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are integral domains.
2. In the ring $\mathbb{Z}/4\mathbb{Z}$, consider the element $\overline{2} = 2 + 4\mathbb{Z}$. This element is non-zero, since 4 does not divide 2, but $\overline{2}.\overline{2} = \overline{2.2} = \overline{4} = \overline{0}$, so $\overline{2}$ is a zero divisor in $\mathbb{Z}/4\mathbb{Z}$.
3. For $n \geq 2$, $M_n(\mathbb{C})$ is not an integral domain, since we can find two non zero matrices A, B such that $AB = 0$ (**Can you find an example ?**). Of course, this is cheating a bit, since $M_n(\mathbb{C})$ is not commutative so it cannot be an integral domain anyway.
4. If R is not trivial, the ring $R \times R$ is not an integral domain, since we have
$$(1,0) \cdot (0,1) = (0,0).$$
5. The rings $\mathbb{Z}[\sqrt{d}]$, and $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ are integral domains (exercise).

Definition. Let R be a ring. Let $f \in R[X]$, $f \neq 0$. Since $f \neq 0$, we can write $f = a_n X^n + a_{n-1}X^{n-1} + \cdots + X + 1$, with $a_i \in R$, $a_n \neq 0$. The integer $n \geq 0$ is called the *degree* of f, and is denoted by $\deg(f)$.

The coefficient a_n is called the *leading coefficient* of P.

For example, $P(X) = 1 \in \mathbb{C}[X]$ has degree 0, and $P(X) = -2X^3 + 5X + 2 \in \mathbb{Z}[X]$ has degree 3.

Proposition 1.6. *Let R be a ring. Then the following properties hold:*

1. *$R[X]$ is non trivial if and only if R is non trivial*
2. *$R[X]$ is commutative if and only if R is commutative*
3. *$R[X]$ has no zero divisors if and only if R has no zero divisors*
4. *$R(X]$is an integral domain if and only if R is an integral domain.*

Moreover, if (3) *or* (4) *holds, then we have*

$$\deg(PQ) = \deg(P) + \deg(Q) \textit{ for all } P, Q \in R[X], P \neq 0, Q \neq 0.$$

Proof. Point (1) comes from the fact that R and $R[X]$ have same 0 and 1 .

Let us prove (2). Assume that $R[X]$ is commutative, and let $r, s \in R$. Viewing r and s as constant polynomials, and using the fact that $R[X]$ is commutative, then we have $rs = sr$. Conversely, if R is commutative, it is easy to check that $R[X]$ is commutative (simply use the definition of multiplication. **Do it!!!**).

Let us prove (3). Assume first that $R[X]$ has no zero divisors, and take $r, s \in R$ such that $rs = 0$. Viewing r and s as constant polynomials, and using the fact that $R[X]$ has no zero divisors by assumption, we conclude that $r = 0$ or $s = 0$ in $R[X]$, and therefore in R. Assume now that R has no zero divisors. We need to show that if P, Q are non-zero polynomials, then PQ is non-zero. Suppose that $\deg(P) = m$ and $\deg(Q) = n$, so $P = a_mX^m + \cdots + a_1X + a_0$ where $a_m \neq 0$, and $Q = b_nX^n + \cdots + b_1X + b_0$, with $b_n \neq 0$. The coefficient of X^{m+n} in PQ is $a_mb_n \neq 0$ since R has no zero divisors, so $PQ \neq 0$ and $\deg(PQ) = m + n = \deg(P) + \deg(Q)$.

Point (4) is simply a consequence of (1), (2) and (3). □

Remark 1.7. The last part of the proposition is not true anymore if R fails to be an integral domain. **Can you give an example?**

Definition. An element $a \in R$ is a *unit* if there exists $b \in R$ with $a \cdot b = 1 = b \cdot a$. The set of units is denoted by R^*.

Notice that the element b above is unique: indeed, if $b \cdot a = 1 = a \cdot b$ and $c \cdot a = 1 = a \cdot c$, then $b = 1 \cdot b = (c \cdot a) \cdot b = c \cdot (a \cdot b) = c \cdot 1 = c$. It is denoted by a^{-1}, and called the *inverse* of a.

Proposition 1.8. *Let R be a ring. Then R^* is a group for the multiplication.*

Proof. The multiplication is associative, ans has a neutral element, which is 1. Now if $a \in R^*$, the inverse $a^{-1} \in R$ is also a unit with inverse a, since we have

$$a^{-1} \cdot a = a \cdot a^{-1} = 1.$$

Thus R^* is stable by inverse. If $a, b \in R^*$, then $(a \cdot b) \cdot b^{-1} \cdot a^{-1} = a(b \cdot b^{-1}) \cdot a^{-1} = a \cdot 1 \cdot a^{-1} = a \cdot a^{-1}$. Similarly, $b^{-1} \cdot a^{-1}(a \cdot b) = 1$, so $a \cdot b \in R^*$ and we have $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$. □

Remark 1.9. We proved along the way that if $a, b \in R^*$, then

$$(a^{-1})^{-1} = a, (ab)^{-1} = b^{-1}a^{-1}.$$

Be careful!!!The set R is **not** a group for multiplication.

Examples.

(1) 0 is never a unit, except in the trivial ring.
(2) 1 and -1 are always units in any ring R.
(3) 2 is a unit in $\mathbb{Q}$, but is not a unit in $\mathbb{Z}$. Indeed, since $\mathbb{Z} \subset \mathbb{Q}$, if 2 has an inverse in $\mathbb{Z}$, it has an inverse in $\mathbb{Q}$, which is necessarily $\frac{1}{2}$. But $\frac{1}{2} \notin \mathbb{Z}$. In fact, we have $\mathbb{Z}^* = \{\pm 1\}$ (**Check it!!!**).

The following proposition will be useful in the sequel.

Lemma 1.10. *Let R be an integral domain. Then $R[X]^* = R^*$.*

Proof. Let $P \in R[X]^*$. Then there exists $Q \in R[X]$ such that $PQ = QP = 1$. It implies in particular that $P \neq 0$.

Applying the degree on both sides of the equation, we get $\deg(PQ) = \deg(1)$, so $\deg(P) + \deg(Q) = 0$, since R is an integral domain. Since $\deg(P)$ and $\deg(Q)$ are non negative integers, we deduce that $\deg(P) = \deg(Q) = 0$. Therefore P and Q are non zero constant polynomials, say $P = a$ and $Q = b$. But then we have $ab = ba = 1$, so $a \in R^*$. Therefore $P = a$ is invertible as well, with inverse $Q = a^{-1}$.

Conversely, if $a \in R^*$, it is invertible in $R[X]$, with inverse a^{-1}. Thus $R[X] = R^*$ if R is an integral domain.

□

Once again, this is NOT true if R fails to be an integral domain. Can you find an counterexample? (*Hint: Try* $R = \mathbb{Z}/4\mathbb{Z}$).

Definition. A ring R is called a *division ring* if it is non-trivial, and $R^* = R\backslash 0$, that is every element except 0 is a unit in R.

A ring R is called a *field* if it is non-trivial, commutative, and every element except 0 is a unit in R.

Thus a field is the same thing as a commutative division ring.

Examples.

(1) The ring $\mathbb{Z}$ is not a field (the only units in $\mathbb{Z}$ are 1 and -1).
(2) The rings $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields.
(3) The ring $\mathbb{Z}/3\mathbb{Z}$ is a field, because $\overline{2}\cdot\overline{2} = \overline{4} = \overline{1}$, so both $\overline{1}$ and $\overline{2}$ are units.
(4) The rings $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ are not fields (exercise)
(5) The ring $\mathbb{Q}[\sqrt{d}]$ is a field (exercise)
(6) If K is a field, then the set of rational fractions
$$K(X) = \{\frac{P(X)}{Q(X)}, P, Q \in K[X], Q \neq 0\}$$
is a field. Recall that the operations on $K(X)$ are defined by:
$$\frac{P_1}{Q_1} + \frac{P_2}{Q_2} = \frac{P_1Q_2 + P_2Q_1}{Q_1Q_2}, \frac{P_1}{Q_1}.\frac{P_2}{Q_2} = \frac{P_1P_2}{Q_1Q_2}, -\frac{P}{Q} = \frac{-P}{Q},$$
and $0_{K(X)} = \frac{0}{1}, 1_{K(X)} = \frac{1}{1}$

Definition. Let F be a field. We say that a subset $K \subset F$ is a *subfield* of F, if K is a subring of F and for all $x \in K - \{0\}, x^{-1} \in K$.

Examples.

(1) $\mathbb{Q}, \mathbb{Q}[\sqrt{d}], \mathbb{R}$ are subfields of $\mathbb{C}$
(2) $\mathbb{Z}$ is not a subfield of $\mathbb{Q}$
(3) If K is a field, $K[X]$ is not a subfield of $K(X)$ (**do you see why?**).

Proposition 1.11. *Let F be a field. A subfield K of F is a field for the same operations. Morever, the intersection of a family of subfields of F is a subfield of F.*

Proof. As for the case of subrings. □

Proposition 1.12. *A zero divisor cannot be a unit. In particular, a division ring contains no zero divisors, and every field is an integral domain.*

Proof. Assume that $a \in R, a \neq 0$ is a left zero divisor, so there exists $b \neq 0$ such that $a \cdot b = 0$. If a is a unit then

$$b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0.$$

This is a cotnradiction. The case of right zero divisors is left to the reader. The remaining part of the proposition follows easily from the first one. □

Warning: An integral domain is not necessarily a field. For example, $\mathbb{Z}$ is a integral domain which is not a field. However, the following proposition shows that $\mathbb{Z}/z\mathbb{Z}$ is either a field whenever it is an integral domain, but this phenomenon is really particular to rings with finitely many elements.

Proposition 1.13. *The ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is a prime number. If $n \geq 2$ is not a prime number, then $\mathbb{Z}/n\mathbb{Z}$ has zero divisors.*

Proof. Let us show that $\mathbb{Z}/p\mathbb{Z}$ is a field when p is a prime number. We already now that it is a non-trivial commutative ring, so it remains to show that any non-zero element has an inverse. Let $\overline{a} \in \mathbb{Z}/p\mathbb{Z}, \overline{a} \neq \overline{0}$. We then have $p \not| a$, so a and p a relatively prime since p is a prime number. By Bézout theorem, there exists $u, v \in \mathbb{Z}$ such that $ua + vp = 1$. Hence we have

$$\overline{1} = \overline{ua + vp} = \overline{u}.\overline{a} + \overline{v}.\overline{p} = \overline{u}.\overline{a}.$$

Thus $\overline{u}$ is an inverse for $\overline{a}$.

Now assume that n is not prime. If $n = 1$, all the the integers are congruent modulo n, so $\mathbb{Z}/n\mathbb{Z}$ is the trivial. If $n \geq 2$, the assumption implies that $n = n_1 n_2$, with $1 < n_1, n_2 < n$. In this case, $\overline{n_i} \neq \overline{0}$ since n_i is not a multiple of n, and we have

$$\overline{n}_1 \cdot \overline{n}_2 = \overline{n_1 n_2} = \overline{n} = \overline{0}.$$

□

Notation: If p is a prime number, the field $\mathbb{Z}/p\mathbb{Z}$ will be denoted by $\mathbb{F}_p$.

1.5. An example of a division ring: Hamilton quaternions. Applications to coding theory.

1.5.1. *Hamilton quaternions.* Remember that if $z = a + ib \in \mathbb{C}$, then $\overline{z}$ is by definition $\overline{z} = a - ib$, and that we have $z\overline{z} = a^2 + b^2$.

We are now going to define a non-commutative division ring, the ring $\mathbb{H}$ of Hamilton quaternions.

We define the set $\mathbb{H}$ by

$$\mathbb{H} = \{M \in M_2(\mathbb{C}) | M = \begin{pmatrix} z_1 & -z_2 \\ \overline{z}_2 & \overline{z}_1 \end{pmatrix}, z_1, z_2 \in \mathbb{C}\}$$

Proposition 1.14. *The set $\mathbb{H}$ is a non-commutative subring of $M_2(\mathbb{C})$. Moreover, it is a division ring.*

Proof. The 0 and 1 of $M_2(\mathbb{C})$ are the zero matrix and the identity matrix, which clearly both belong to $\mathbb{H}$.

Now let $M, M' \in \mathbb{H}$, so

$$M = \begin{pmatrix} z_1 & -z_2 \\ \overline{z}_2 & \overline{z}_1 \end{pmatrix}, M' = \begin{pmatrix} z'_1 & -z'_2 \\ \overline{z'}_2 & \overline{z'}_1 \end{pmatrix},$$

for some $z_1, z'_1, z_2, z'_2 \in \mathbb{C}$.

We have

$$-M = \begin{pmatrix} -z_1 & z_2 \\ -\overline{z}_2 & -\overline{z}_1 \end{pmatrix} = \begin{pmatrix} -z_1 & -(-z_2) \\ \overline{-z_2} & \overline{-z_1} \end{pmatrix},$$

so $-M \in \mathbb{H}$.

We have

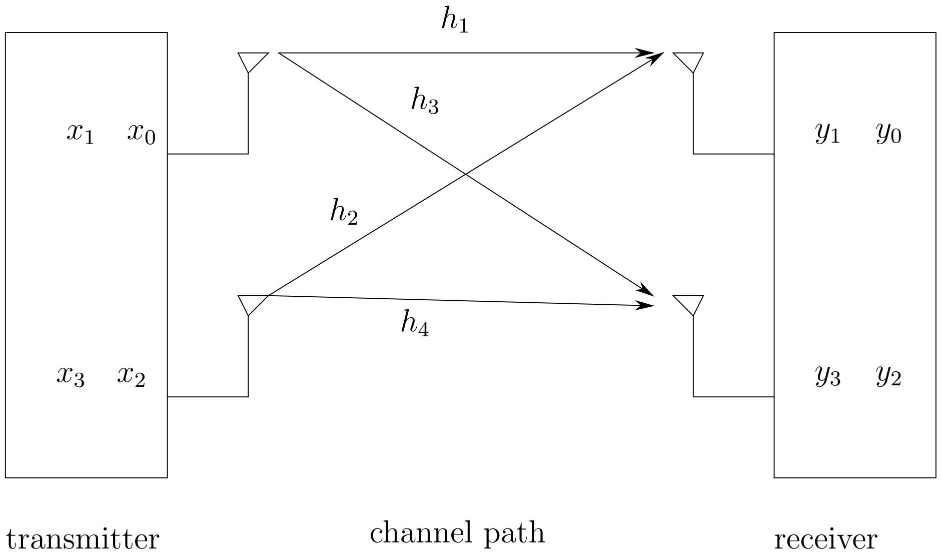
$$M + M' = \begin{pmatrix} z_1 + z'_1 & -z_2 - z'_2 \\ \overline{z}_2 + \overline{z'}_2 & \overline{z}_1 + \overline{z'}_1 \end{pmatrix} = \begin{pmatrix} z_1 + z'_1 & -(z_2 + z'_2) \\ \overline{z_2 + z'_2} & \overline{z_1 + z'_1} \end{pmatrix},$$

so $M + M' \in \mathbb{H}$.

Finally, we have

$$\begin{aligned} MM' &= \begin{pmatrix} z_1z'_1 - z_2\overline{z'}_2 & -z_1z'_2 - z_2\overline{z'}_1 \\ \overline{z}_2z'_1 + \overline{z}_1\overline{z'}_2 & -\overline{z}_2\overline{z'}_2 + \overline{z}_1\overline{z'}_1 \end{pmatrix} \\ &= \begin{pmatrix} z_1z'_1 - z_2\overline{z'}_2 & -(z_1z'_2 + z_2\overline{z'}_1) \\ \overline{z_1z'_2 + z_2\overline{z'}_1} & \overline{z_1z'_1 - z_2\overline{z'}_2} \end{pmatrix} \end{aligned}$$

so $MM' \in \mathbb{H}$. Hence $\mathbb{H}$ is a subring of $M_2(\mathbb{C})$,clearly non-trivial.



We now prove that $\mathbb{H}$ is not commutative.

We have $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{H}$, and one can check easily that

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix},$$

but that

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

so $\mathbb{H}$ is not commutative.

Finally, we prove that every non zero element of $\mathbb{H}$ is invertible. First notice that

$$\det\begin{pmatrix} z_1 & -z_2 \\ \overline{z}_2 & \overline{z}_1 \end{pmatrix} = z_1\overline{z}_1 + z_2\overline{z}_2 = |z_1|^2 + |z_2|^2,$$

so it is non-zero, unless $z_1 = z_2 = 0$.

So any non-zero element of $\mathbb{H}$ is an invertible matrix, so we just have to check that the inverse of this matrix remains in $\mathbb{H}$. But we have

$$M^{-1} = \frac{1}{|z_1|^2 + |z_2|^2}\begin{pmatrix} \overline{z}_1 & z_2 \\ -\overline{z}_2 & z_1 \end{pmatrix} = \frac{1}{|z_1|^2 + |z_2|^2}\begin{pmatrix} \overline{z}_1 & -(-z_2) \\ \overline{-z_2} & \overline{\overline{z}}_1 \end{pmatrix},$$

so $M^{-1} \in \mathbb{H}$, and since $MM^{-1} = M^{-1}M = I_2$, it shows that every non-zero element of $\mathbb{H}$ has an inverse in $\mathbb{H}$. □

1.5.2. *Hamilton quaternions and coding theory: the Alamouti code.* In this section, we give an application of ring theory to wireless communication.

Suppose that we want to transmit information symbols without using any wire. Typically, it is the case when you are using wireless internet connections or cellular phones. During transmission via the channel (the air for example, in the case of cellular phones), two phenomenons may occur: fading, that is a loss of intensity of the transmitted information, due to the fact that it may go through obstacles, such as trees or buildings (this is way your voice may not appear as loud as it is actually is to your interlocutor), and noise, due for example to interferences with other waves (this is why your interlocutor may not hear you properly sometimes). This is why the information transmitted to the receiver is not the original one.

The problem is to encode your information and transmit it in such a way that the probablity error is minimal, that is only very few errors occur during transmission. Of course, one way to proceed to send the same information several times, but it costs computer memory, it increases the amount of energy necessary for transmission, and no so much information is transmitted, so it is not worth it.

Suppose that we have two transmitting antennas and two receiving antennas. The information symbols we want to transmit are complex numbers. Each transmitting antenna sends an information symbol which will be received by each of the two receiving antennas, the information symbol going through two different paths. We will assume that the channel does not have time to change during two successive uses.

During the first use, the first antenna transmits x_0 and the second one transmits x_2. Each of these two symbols go through the two possible paths and are received by the receiving antennas.

The symbol x_0 arrives to the first receiving antenna as h_1x_0 and to the second one as h_3, where h_1, h_3 are coefficients representing fading. The symbol x_2 arrives to the first receiving antenna as h_2x_0 and to the second one as h_4, where h_2, h_4 are once again coefficients representing fading.

Therefore, the first receiving antenna receives a signal y_0 which is the sum of 3 different signals: h_1x_0, h_3x_1 and some noise ν_1, so

$$y_0 = h_1x_0 + h_3x_3 + \nu_1$$

Similarly, the second receiving antenna receives a signal y_2 of the form

$$y_2 = h_2x_0 + h_4x_2 + \nu_2$$

During the second use, the first transmitting antenna sends x_1, and the second one sends x_3. Since the channel does not have time to change between the two uses, the fading coefficients will remain the same, and the first and second receiving antennas will receveive signals y_1 and y_3 of the form

$$y_1 = h_1x_1 + h_3x_3 + \nu_3$$

and

$$y_3 = h_2x_1 + h_4x_3 + \nu_4.$$

Therefore, setting

$$H = \begin{pmatrix} h_1 & h_3 \\ h_2 & h_4 \end{pmatrix}, N = \begin{pmatrix} \nu_1 & \nu_3 \\ \nu_2 & \nu_4 \end{pmatrix}$$

and

$$X = \begin{pmatrix} x_0 & x_1 \\ x_2 & x_3 \end{pmatrix}, Y = \begin{pmatrix} y_0 & y_1 \\ y_2 & y_3 \end{pmatrix},$$

we get the following matrix equation

$$Y = HX + N.$$

The matrices H and N are random matrices following a Gaussian law.

We send in fact a matrix $X \in M_2(\mathbb{C})$, and we receive a matrix $Y \in M_2(\mathbb{C})$. The receiver is supposed to know the set $\mathcal{C}$ of all matrices X we send, called the *codebook* . An element $X \in \mathcal{C}$ is called a *codeword*. He is supposed to know also the channel, that is the matrix H. The main problem is that $Y \notin \mathcal{C}$ in general. How to decode? That is, how to recover a codeword $\hat{X} \in \mathcal{C}$ from Y, in such a way that the probability $\mathbb{P}(X \to \hat{X})$ of sending X and decoding $\hat{X} \neq X$ is as small as possible?

The recipe is as follows: for any $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, set

$$||M||_2 = \sqrt{|a|^2 + |b|^2 + |c|^2 + |d|^2}$$

The codeword $\hat{X}$ will be a codeword such that $||Y - HX'||_2$ is minimal among all the codewords $X' \in \mathcal{C}$ (if there is more than one codeword with this property, one is chosen at random). The receiver can always compute $\hat{X}$ since he knows $\mathcal{C}$ and H.

With this way of decoding, we have

$$\mathbb{P}(X \to \hat{X}) \leq \frac{C}{\min\limits_{X \neq X' \in \mathcal{C}} |\det(X - X')|^4},$$

where C is a constant depending on the channel and X, X' describe $\mathcal{C}$. So the main question is now: how to design the codebook $\mathcal{C}$? The criterion is: reliability ! To have an interesting upper bound, and ensure that $\mathbb{P}(X \to \hat{X})$ is small, we need to maximize $\min\limits_{X \neq X' \in \mathcal{C}} |\det(X - X')|^4$, the first step being that we need to ensure that $\mathcal{C}$ is chosen in such a way that $\det(X - X') \neq 0$ for all $X \neq X'$.

The main difficulty to achieve this is the non-linearity of the determinant. The idea is then to take for $\mathcal{C}$ a finite subset of a subring R of $M_2(\mathbb{C})$ which is also a division ring. In this way, we will have $X - X' \in R\backslash\{0\}$ since R is a ring and the fact that R is a division ring will ensure that every $X - X'$ is invertible in R, and therefore in $M_2(\mathbb{C})$, which means that we will have $\det(X - X') \neq 0$ for all $X \neq X' \in C$.

For example, the Alamouti code sends two information symbols $z_1, z_2 \in \{\pm 1, \pm i\}$ as follows: with the previous notation, we set $x_0 = z_1, x_1 = -z_2, x_3 = \overline{z}_2, x_3 = \overline{z}_1$, so the two antennas transmit four information symbols made from the original data z_1, z_2. In this case we take

$$\mathcal{C} = \{M \in M_2(\mathbb{C}) | M = \begin{pmatrix} z_1 & -z_2 \\ \overline{z}_2 & \overline{z}_1 \end{pmatrix}, z_1, z_2 \in \{\pm 1, \pm i\}\}$$

Now to evaluate $\min\limits_{X \neq X' \in \mathcal{C}} |\det(X - X')|^4$, instead of testing $\frac{16 \times 15}{2} = 120$ possibilities, observe that $\mathcal{C} \subset \mathbb{H} \cap M_2(\mathbb{Z}[i])$. Since $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, $M_2(\mathbb{Z}[i])$ is a subring of $M_2(\mathbb{C})$, and therefore so is $S := \mathbb{H} \cap M_2(\mathbb{Z}[i])$. Hence $X - X' \in S$ for all $X \neq X' \in \mathcal{C}$. Therefore

$$\min_{X \neq X' \in \mathcal{C}} |\det(X - X')|^4 \geq \min_{0 \neq M \in S} |\det(M)|^4$$

But if $M = \begin{pmatrix} z_1 & -z_2 \\ \overline{z}_2 & \overline{z}_1 \end{pmatrix} \in S$, then $\det(M) = |z_1|^2 + |z_2|^2$. If $M \neq 0$, z_1 or z_2 is not 0, say z_1, and since $z_1 \in \mathbb{Z}[i]$, we have $|z_1|^2 \in \mathbb{N}$. Since $z_1 \neq 0$, we get $|z_1|^2 \geq 1$, so $\min |\det(M)| \geq 1$. This lower bound is easily obtained for $z_1 = 1, z_2 = 0$ for example. Hence $\min\limits_{0 \neq M \in S} |\det(M)|^4 = 1$. Putting things together, we get

$$\min_{X \neq X' \in \mathcal{C}} |\det(X - X')|^4 \geq 1$$

It appears that the Alamouti code has really good performances, thanks to this last property (The lower bound 1 may not appear to be very big, but the other existing codes have very small lower bounds).

2. Ring homomorphisms. Definition and basic examples.

We start by recalling some basic definitions of set theory.

Definition. Let E, E' be two sets, and let $f : E \to E'$ be a map.

We define $\mathrm{Im}(f)$ to be the subset of E' defined by

$$\mathrm{Im}(f) = \{x' \in E' | x' = f(x) \text{ for some } x \in E\}$$

In other words, $\mathrm{Im}(f) = \{f(x), x \in E\}$, that is the set of all possible values attained by the map f.

We say that $f : E \to E'$ is *surjective* if $\mathrm{Im}(f) = E'$. Another way to say it is that for all $x' \in E'$, the equation $f(x) = x'$ has **at least one** solution $x \in E$.

We say that f is *injective* if

$$\text{for all } x_1, x_2 \in E, f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$

In other words, f is injective if for all $x' \in E$, the equation $f(x) = x'$ has **at most one** solution $x \in E$.

We say that $f : E \to E'$ is *bijective* if it is injective and surjective. In other words, f is bijective if for all $x' \in E$, the equation $f(x) = x'$ has **exactly one** solution $x \in E$.

One can show that f is bijective if and only if there exists a function $g : E' \to E$ satisfying

$$g(f(x)) = x \text{ for all } x \in E, \text{ and } f(g(x')) = x' \text{ for all } x' \in E'$$

The map g is unique in this case, and denoted by f^{-1}. If f is bijective, then f^{-1} is bijective as well and $(f^{-1})^{-1} = f$.

Definition. A *ring homomorphism* is a function $\phi \colon R \to R'$, where R and R' are rings, such that for all a and b in R,

i) $\phi(1) = 1$ ii) $\phi(a + b) = \phi(a) + \phi(b)$ iii) $\phi(a.b) = \phi(a).\phi(b)$.

Note that the operations + and . on the left of the equations are the operations in R and the operations on the right are the operations in R'. Similarly, i) should really be written $\phi(1_R) = 1_{R'}$.

Exercises:

1) If $\phi : R \to R'$ is a ring homomorphism, then $\mathrm{Im}(\phi)$ is a subring of R'.

2) If $\phi_1 : R_1 \to R_2, \phi_2 : R_2 \to R_3$, then $\phi_2 \circ \phi_1 : R_1 \to R_3$ is a ring homomorphim as well.

Examples.

(1) Each of the inclusions $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ is a ring homomorphism. An inclusion is injective, but not surjective in general.
(2) More generally, if S is any subring of R, then the inclusion of S in R is an injective ring homomorphism.
(3) If R_1 and R_2 are rings, the projection maps $\pi_1\colon R_1 \times R_2 \to R_1$ and $\pi_2\colon R_1 \times R_2 \to R_2$, given by $\pi_1((r_1, r_2)) = r_1$, $\pi_2((r_1, r_2)) = r_2$, are ring homomorphisms. Both are surjective, but not injective (**check it!**).
(4) The map $m \in \mathbb{Z} \mapsto \overline{m} \in \mathbb{Z}/n\mathbb{Z}$ is a ring homomorphism. It is surjective, but not injective (**check it!**).

Proposition 2.1. *If $\phi\colon R \to R'$ is a ring homomorphism then*

(1) $\phi(0_R) = 0_{R'}$
(2) $\phi(-a) = -\phi(a)$ *for all $a \in R$.*
(3) *If $a \in R^*$, then $\phi(a) \in {R'}^*$ and we have*

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

In particular, $\phi : R \to R'$ restricts to a group homomorphism $\phi : R^* \to {R'}^*$.

Proof. For (1), note that $\phi(0_R) + \phi(0_R) = \phi(0_R + 0_R) = \phi(0_R)$, now add $-\phi(0)$ to each side of this equation. For (2), we have

$$\phi(-a) + \phi(a) = \phi((-a) + a) = \phi(0_R),$$

the first equality coming from the fact that ϕ is a ring homomorphism. But by (1), $\phi(0_R) = 0_R$, so $\phi(-a) + \phi(a) = 0_R$. Now add $-\phi(a)$ to each side. For (3), if a_inR^*, we have

$$\phi(aa^{-1}) = \phi(a^{-1}a) = \phi(1_R) = 1_{R'}.$$

Since ϕ is a ring isomorphism, we get

$$\phi(a)\phi(a^{-1}) = \phi(a^{-1})\phi(a) = 1_{R'}.$$

By definition of a unit, the equalities above show that $\phi(a^{-1})$ is a unit with inverse $\phi(a)^{-1}$. The last statement is clear. □

Definition. A ring homomorphism $\phi\colon R \to R'$ is called an *isomorphism* if it is bijective. In this case, the inverse of ϕ, denoted by ϕ^{-1} is necessarily a ring homomorphism (and then a ring isomorphism!).

Indeed

$$\phi^{-1}(1_{R'}) = \phi^{-1} \circ \phi(1_R) = 1_R,$$

and if x and y are elements of R', let $\phi^{-1}(x) = a$, $\phi^{-1}(y) = b$. Then

$$\phi(a) = x, \quad \phi(b) = y, \quad \phi(a+b) = x + y, \quad \text{and} \quad \phi(ab) = xy.$$

Therefore

$$\phi^{-1}(x+y) = \phi^{-1}(\phi(a+b)) = a+b = \phi^{-1}(x) + \phi^{-1}(y)$$

and

$$\phi^{-1}(xy) = \phi^{-1}(\phi(ab)) = a.b = \phi^{-1}(x)\phi^{-1}(y).$$

Definition. Let R be a ring, let $r \in R$, and let $m \in \mathbb{Z}$. We define $m.r$ as follows:

$$m.r = \begin{cases} 0_R & \text{if} \quad m = 0 \\ r + \cdots + r & \text{if} \quad m \geq 1 \\ (-r) + \cdots + (-r) & \text{if} \quad m \leq -1 \end{cases}$$

where r is summed m times in the second case, and $-r$ is summed $-m$ times in the third case.

Proposition 2.2. *For any ring R, there is a unique ring homomorphism $\Theta_R : \mathbb{Z} \to R$. It is defined by*

$$\Theta_R(m) = m.1_R \textit{ for all } m \in \mathbb{Z}\,.$$

Proof. If $\phi : \mathbb{Z} \to R$ is a ring homomorphism, then we have by definition $\phi(1) = 1_R$. In particular, for each $m > 0$, we get

$$\phi(m) = \phi(1_R + \ldots + 1_R) = \phi(1_R) + \ldots + \phi(1_R) = m.1_R$$

Also, by the previous proposition, we have $\phi(0) = 0_R$. Hence we get

$$\phi(m) = m.1_R \text{ for all } m \geq 0.$$

Now if $m < 0$, then $\phi(m) = \phi(-(-m)) = -\phi(-m)$. Since $-m > 0$, the previous case gives

$$\phi(m) = -(-m).1_R.$$

Notice now that we have

$$\begin{array}{rcl} (-m).1_R + m \cdot (-1_R) & = & (1_R + \ldots + 1_R) + ((-1_R) + \ldots + (-1_R)) \\ & = & (1_R + (-1_R)) + \ldots + (1_R + (-1_R)) \\ & = & 0_R + \ldots + 0_R \\ & = & 0_R. \end{array}$$

This implies easily that we have

$$-((-m).1_R) = (-m).(-1_R).$$

Therefore, we get

$$\phi(m) = -m.(-1_R) = m.1_R,$$

the last equality coming from the definition of $m.1_R$ (since $m < 0$).

Hence we proved that $\phi(m) = m.1_R = \Theta_R(m)$ for all $m \in \mathbb{Z}$. Now it remains to show that this function Θ_R is indeed a ring homorphism. We know that $\Theta_R(1) = 1_R$ already. We need to check the equalities

$$(n+m).1_R = n.1_R + m.1_R, nm.1_R = (n.1_R)(m.1_R), \text{ for all } n, m \in \mathbb{Z}$$

This is not difficult but extremely tedious, so we are skipping this part of the proof and leave it to the courageous reader. □

Let R be a commutative ring. There is an injective homomorphism from R into the polynomial ring $R[X]$ sending $r \in R$ to the constant polynomial whose only term is in X^0 with coefficient r. We shall identify r with this constant polynomial, i.e. we view R as a subring of $R[X]$.

Proposition 2.3. *Let R and R be commutative rings, let $f : R \to R'$ be a homomorphism and let $y \in R'$. Then there is a unique homorphism $\psi_{f,y} : R[X] \to R'$ such that*

$$\psi_{f,y}(a) = f(a) \text{ for all } a \in \text{ and } \psi_{f,y}(X) = y.$$

It is defined by

$$\psi_{f,y}\left(\sum_i a_i X^i\right) = \sum_i f(a_i) y^i.$$

Moreover, every homomorphism $R[X] \to R'$ is obtained in this way.

Proof. Define $\psi_{f,y}$ by

$$\psi_{f,y}\left(\sum_i a_i X^i\right) = \sum_i f(a_i) y^i.$$

Clearly we have $\psi_{f,y}(a) = f(a)$ for all $a \in$ and $\psi_{f,y}(X) = y$ (**Check it!** This uses the fact that $f(0_R) = 0_{R'}$ and $f(1_R) = 1_{R'}$). Let us check that this is a ring homomorphism:

We have $\psi(1_R) = f(1_R) = 1_{R'}$. Moreover,

$$\psi_{f,y}(\sum_i a_i X^i + \sum_i b_i X^i) = \psi_{f,y}(\sum_i (a_i + b_i) X^i) = \sum_i f(a_i + b_i) y^i$$

Since f is a ring homomorphism, we get

$$\psi_{f,y}(\sum_i a_i X^i + \sum_i b_i X^i) = \sum_i (f(a_i) + f(b_i)) y^i = \sum_i f(a_i) y^i + \sum_i f(b_i) y^i,$$

hence

$$\psi_{f,y}(\sum_i a_i X^i + \sum_i b_i X^i) = \psi_{f,y}(\sum_i a_i X^i) + \psi_{f,y}(\sum_i b_i X^i)$$

Finally, we have

$$\psi_{f,y}((\sum_i a_i X^i).(\sum_i b_i X^i)) = \psi_{f,y}(\sum_i (\sum_{j+k=i} a_j.b_k) X^i) = \sum_i f(\sum_{j+k=i} a_j.b_k) y^i$$

Since f is a ring homomorphism, we obtain

$$\psi_{f,y}((\sum_i a_iX^i).(\sum_i b_iX^i)) = \sum_i \sum_{j+k=i} f(a_j).f(b_k)y^i = (\sum_i f(a_i)y^i)(\sum_i f(b_i)y^i),$$

hence

$$\psi_{f,y}((\sum_i a_iX^i).(\sum_i b_iX^i)) = \psi_{f,y}(\sum_i a_iX^i).\psi_{f,y}(\sum_i b_iX^i)$$

It is unique, because for any such other ring homomorphism ϕ satisfying the conditions $\phi(a) = f(a)$ for all $a \in R$, and $\phi(X) = y$, we have (since we have a ring homorphism)

$$\phi(\sum_i a_iX^i) = \sum_i \phi(a_i)(\phi(X))^i = \sum_i f(a_i)y^i = \psi_{f,y}\left(\sum_i a_iX^i\right),$$

and therefore $\phi = \psi_{f,y}$.

To prove that any homomorphism $\varphi : R[X] \to S$ is obtained in this way, set $f = \varphi_{|R}$ (the restriction of φ to R) and $y = \varphi(X)$. Since φ is a ring homomorphism, $f : R \to R'$ is a ring homomorphism (**do you see why?**).

We have just seen that

$$\phi(\sum_i a_iX^i) = \sum_i \phi(a_i)(\phi(X))^i.$$

Now by **definition** of f and y, this rewrites

$$\phi(\sum_i a_iX^i) = \sum_i f(a_i)(y)^i = \psi_{f,y}(\sum_i a_iX^i).$$

Hence $\varphi = \psi_{f,y}$. This concludes the proof. □

Examples.

(1) We can apply this proposition to $R = R', f = \mathrm{Id}_R$ and $y = r \in R$. We then get a morphism called *the evaluation at r*. In this case, the image of a polynomial $P \in R[X]$ is commonly denoted by $P(r)$. If $P = \sum_i a_iX^i \in R[X]$ and $r \in R$, we have $P(r) = \sum_i a_ir^i$, so this is just taking the value of P at $X = r$. In other words, for a given $r \in R$, the map

$$R[X] \to R, P := \sum_i a_iX^i \mapsto P(r) := \sum_i a_ir^i$$

is a ring homomorphism.

(2) If $f : R \to R'$ is a ring homomorphism, we can also view f as a morphism $f : R \to R'[X]$, since $R' \subset R'[X]$. Setting $y = X \in R'[X]$ and applying the previous proposition, we obtain a ring homomorphism $\psi : R[X] \to R'[X]$ such that

$$\psi(\sum_i a_i X^i) = \sum_i f(a_i) X^i.$$

In particular, for any integer $n \geq 2$, we have a ring homomorphism $\psi : \mathbb{Z}[X] \to \mathbb{Z}/n\mathbb{Z}[X]$ satisfying

$$\psi(\sum_i a_i X^i) = \sum_i \overline{a}_i X^i.$$

It is called *the reduction modulo n*; the image of P is denoted in general by $\overline{P}$.

For example, if $P = 3X^3 + 2X^2 - X + 5 \in \mathbb{Z}[X]$, we have

$$\overline{P} = \overline{2}X^2 - X + \overline{2} = -X^2 - X - 1 \in \mathbb{F}_3[X]$$

and

$$\overline{P} = X^3 - X + \overline{1} = X^3 + X + \overline{1} \in \mathbb{F}_2[X].$$

Recall that $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for a prime number p.

Let us come back to the general properties of ring homomorphisms.

Definition. Let R, R' be arbitrary rings. If $\phi \colon R \to R'$ is a ring homomorphism, define $\ker(\phi)$, *the kernel of* ϕ, to be

$$\ker(\phi) = \{x \in R \mid \phi(x) = 0_{R'}\}.$$

Proposition 2.4. *A homomorphism* $\phi \colon R \to R'$ *is injective if and only if* $\ker(\phi) = \{0\}$.

Proof. Suppose ϕ is injective. Clearly $\{0\} \subseteq \ker(\phi)$, and if $a \in \ker(\phi)$, then $\phi(a) = 0 = \phi(0)$, so by injectivity of ϕ, $a = 0$. Conversely, suppose $\ker(\phi) = \{0\}$. If $\phi(a) = \phi(b)$, then $\phi(a-b) = \phi(a) - \phi(b) = 0$, so $(a - b) \in \ker(\phi) = \{0\}$, and hence $a = b$. □

Examples.

(1) The kernel of $\pi_1 : (r_1, r_2) \in R_1 \times R_2 \mapsto r_1 \in R_1$ is $\ker(\pi_1) = \{0\} \times R_2$.
(2) The kernel of $m \in \mathbb{Z} \mapsto \overline{m} \in \mathbb{Z}/n\mathbb{Z}$ is $n\mathbb{Z}$. Do you see why ?

3. Ideals and factor rings.

3.1. **Definitions and first properties.**

Definition. A subset I of a ring R is called an *ideal* if

(i) I is an additive subgroup of R, that is $0 \in I$ and for all $a, b \in I$, $a + b \in I$ and $-a \in I$,

(ii) for any $a \in I$ and any $r \in R$, $a.r \in I$ and $r.a \in I$.

Proposition 3.1. *Let* $\phi\colon R \to S$ *be a ring homomorphism. Then* $\ker(\phi)$ *is an ideal of* R*.*

Proof. Let a and b be elements of $\ker(\phi)$. Then $\phi(a \pm b) = \phi(a) \pm \phi(b) = 0$, so $a \pm b \in \ker(\phi)$, giving i). For ii), $\phi(a.r) = \phi(a).\phi(r) = 0.\phi(r) = 0$, and similarly the other way around. □

Examples.

- The set $\{0\}$ is an ideal in any ring R. (Either check the properties, or note that it is the kernel of the identity map.) The set R is an ideal in R (check definition, or as kernel of the map from R to the trivial ring).

- Let I be an ideal of $\mathbb{Z}$. If I is not the zero ideal, it contains some element greater than zero. Let n be the smallest positive element of I. We claim that $I = n\mathbb{Z} = \{nm, m \in \mathbb{Z}\}$.

Indeed, since $n \in I$ and I is an ideal of $\mathbb{Z}$, we have $n\mathbb{Z} \subset I$. Now if $\ell \in I$, we can write $\ell = mn + \ell'$ with $0 \leq \ell' < n$ (Euclidean division of integers). Hence $\ell' = \ell - mn \in I$, since I is an ideal. Then $\ell' = 0$, otherwise we would have a positive element of I smaller than n. Hence $\ell = mn$; and so $I \subset n\mathbb{Z}$. Conversely, it is a good exercise to show that $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

Hence we get

Proposition 3.2. *The ideals of* $\mathbb{Z}$ *are exactly the sets* $n\mathbb{Z} = \{mn\colon m \in \mathbb{Z}\}$*, where* $n \geq 0$*.*

Proposition 3.3. *The intersection of a family of ideals of a ring* R *is an ideal of* R*.*

Proof. Let Λ be an indexing set for the family, and write I_λ for the ideals, where $\lambda \in \Lambda$. Let I be the intersection of the I_λ. Then x is in I if and only if x is in each I_λ. If $a, b \in I_\lambda$ for all λ, then $a + b \in I_\lambda$ for all λ, and hence if $a, b \in I$, so is $a + b$. Similarly, if $r \in R$ and $a \in I_\lambda$ for all λ, then $r.a, a.r \in I_\lambda$ for all λ, and so if $r \in R$ and $a \in I$, then $r.a$ and $a.r$ are in I. □

Definition. If A is a subset of a ring R, then (A), the *ideal generated by* A, is the smallest ideal of R containing A. Equivalently, (A) is the intersection of all ideals containing A.

If $A = \{a_1, \cdots, a_s\}$ is a finite subset of R, we will write $(a_1, \cdots, a_s)$ instead of $(\{a_1, \cdots, a_s\})$. An ideal generated by one element is called a *principal ideal.*

Proposition 3.4. *Let R be a ring and let A be a subset of R. Then*

$$(A) = \{\sum_{i=1}^{n} r_i a_i s_i, \qquad \textit{for some } n \geq 1, a_i \in A, r_i, s_i \in R.\}$$

In particular, $(a) = \{ras, r, s \in R\}$.

To see this, check that elements of this form do form an ideal that contains A, and that any ideal containing A contains these elements.

Remarks: If R is a commutative ring, this simplifies to give

$$(A) = \{\sum_{i=1}^{n} a_i.r_i, n \geq 1,\ r_i \in R,\ a_i \in A\}.$$

Similarly, in a commutative ring R the principal ideal (a) is

$$(a) = \{r.a, r \in R\} = \{a.r, r \in R\}$$

This last equation could be taken as the definition of (a) in a commutative ring.

Notation: If $a \in R$, we denote by aR the set

$$\{a.r, r \in R\}$$

Therefore if R is **commutative**, then

$$(a_1, \cdots, a_s) = a_1R + \cdots + a_sR = \{a_1.r_1 + \cdots + a_sr_s, r_i \in R\}$$

In particular, if R is commutative, a ideal I is principal if and only if $I = aR$ for some $a \in R$.

Warning: If R is not commutative, this is not true anymore.

Examples.

- If $R = \mathbb{Z}$, then $(2) = 2\,\mathbb{Z}$.

- If $R = \mathbb{Z}$, then $(15, 3) = 15\,\mathbb{Z} + 3\,\mathbb{Z} = \{15m + 3n, m, n \in \mathbb{Z}\}$. This ideal is principal, since one can easily show that $15\,\mathbb{Z} + 3\,\mathbb{Z} = 3\,\mathbb{Z}$.

- If $R = \mathbb{Z}[X]$, then
$(2, X) = 2\,\mathbb{Z}[X] + X\,\mathbb{Z}[X] = \{2P(X) + XQ(X), P, Q \in \mathbb{Z}[X]\}$. One can show that this ideal is not principal.

- If $R = M_2(\mathbb{C})$ and $a = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, then the set aR is **NOT** an ideal of R. In fact $aR = \{\begin{pmatrix} x & y \\ x & y \end{pmatrix}, x, y \in \mathbb{C}\}$, so

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \notin aR$$

and thus axiom (ii) is not satisfied.

Proposition 3.5. *Let I be an ideal of R. Then $I = R$ if and only if I contains a unit.*

Proof. If $I = R$, then I contains 1, which is a unit. Conversely, let $u \in I$ be a unit. Then there exists $u^{-1} \in R$, and so $1 = u^{-1}.u \in R.I = I$. But then for any $r \in R$, $r = r.1 \in I$, since I is an ideal. □

Proposition 3.6. *If R is a field, any homomorphism $\phi\colon R \to R'$ is injective, unless R' is the trivial ring.*

Proof. It suffices to check that $\ker(\phi) = \{0\}$. Assume that $\ker(\phi)$ contains a non-zero element $a \in R$. Since $a \neq 0$ and R is a field, a is a unit. Since $\ker(\phi)$ is an ideal, we get $\ker(\phi) = R$. In particular, $1 = \phi(1) = 0 \in R'$, so R' is the trivial ring. □

Recall now that there exists a unique ring homomorphism $\Theta_R : \mathbb{Z} \to R$.

Definition. For any ring R, the *characteristic* of R is defined to be the unique integer $c \geq 0$ such that $\ker(\Theta_R) = (n) = c\,\mathbb{Z}$.

It is denoted by $\mathrm{char}(R)$.

In other words, the characteristic of R is 0 if the equation $n.1_R = 0$ has no solution $n \neq 0, n \in \mathbb{Z}$, and it is the smallest positive solution c of the equation $n.1_R = 0$ if this equation has a non zero solution.

Moreover in this last case, we have

$$\text{For all } n \neq 0, n \in \mathbb{Z}, n.1_R = 0_R \Rightarrow \mathrm{char(R)}|n.$$

For example, $\mathrm{char}(\mathbb{Q}) = 0, \mathrm{char}(\mathbb{Z}\,/2\,\mathbb{Z}) = 2$.

Proposition 3.7. *If R is an integral domain, the characteristic of R is either zero or a prime number.*

Proof. Suppose R is a ring whose characteristic c is neither zero nor a prime. Then either $c = 1$, or $c = l.m$ with $l, m \geq 2$. If $c = 1$, then $1 = 0$ in R, so R is the trivial ring and hence is not an integral domain. If $c = l.m$, $l, m \geq 2$, then $l.1_R \neq 0$, $m.1_R \neq 0$, but $(l.1_R).(m.1_R) = n.1_R = 0$, so R is not an integral domain either. □

Warning: The converse is not true! For example $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ has characteristic 2, but it is not an integral domain since it has zero divisors (could you find some?).

3.2. **Factor rings.** The definition of an ideal was inspired by the list of properties enjoyed by kernels of ring homomorphisms. Compare this with the definition of a normal subgroup in group theory. Just as in group theory, where the quotient of a group by any normal subgroup may be defined, it turns out that we can define a *factor ring.*

Let R be a ring, and $I \subset R$ an ideal of R. Let $r, s \in R$. Say that r is congruent to s modulo I,

$$r \equiv s \mod I \qquad \text{if } r - s \in I.$$

This is an equivalence relation, so R splits into disjoint equivalence classes under it. Notice that if $r \in R$, its equivalence class, denoted by $\overline{r}$ is nothing but the set $r + I := \{r + a, a \in I\}$.

We will write R/I for the set of equivalence classes:

$$R/I = \{\overline{r} : r \in R\}.$$

We define an addition and a multiplication on R/I by

$$\overline{r} + \overline{s} = \overline{r+s}, \qquad \overline{r}.\overline{s} = \overline{r.s}.$$

Proposition 3.8. *The operations described above are well-defined (i.e., do not depend on the choice of a representative of the equivalence class $\overline{r}$). The set R/I is a ring with these operations, with zero element $\overline{0}$, unity $\overline{1}$, and negatives given by $-\overline{r} = \overline{-r}$, and is called the factor ring.*

Proof. Check that + and . are well defined:

$$\overline{r} = \overline{r'} \quad \text{if and only if} \quad r = r' + \lambda \quad \text{for some} \quad \lambda \in I.$$

Similarly, $\overline{s} = \overline{s}'$ if and only if $s = s' + \mu$ for some $\mu \in I$. To show that + and . are well defined, it suffices to show that $\overline{r+s} = \overline{r'+s'}$, and $\overline{r.s} = \overline{r'.s'}$, but

$$(r+s) - (r'+s') = (r-r') + (s-s') = \lambda + \mu \in I \qquad \text{and}$$

$$r.s - r'.s' = (r'+\lambda).(s'+\mu) - r'.s' = r'.\mu + \lambda.s' + \lambda.\mu \in I.$$

Given that the operations are well-defined, it's easy to check that R/I is a ring, because the axioms follow immediately from those in R, for example to check associativity of ., we want to show that $(\overline{a}.\overline{b}).\overline{c} = \overline{a}.(\overline{b}.\overline{c})$. The left hand side is the equivalence class containing $(a.b).c$ and the right hand side is the equivalence class containing $a.(b.c)$, but these are equal, so the classes containing them are too. □

Remark: It is easy to check that R/I is the trivial ring if and only if $I = R$.

Definition. The *canonical projection* is the map

$$\pi : R \to R/I, r \mapsto \overline{r}$$

This is a surjective ring homomorphism with $\ker(\pi) = I$ (**Check it!**).

The rest of this section is devoted to the study of ring homomorphisms $R/I \to R'$. Let's start with an illustrative problem:

Problem: Describe all the ring homomorphisms $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$.

Let's try to analyze the problem. If φ is such a ring homomorphism, then by definition $\varphi(\overline{1}) = 1$, and $\varphi(\overline{0}) = 0$. Since $\mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}$, we could say that we already described all the possible homomorphisms. In fact, there is just this one!

Yes, but there is an hidden difficulty. Elements of $\mathbb{Z}/2\mathbb{Z}$ are equivalence classes , and therefore can be represented by several different elements. For example, $\overline{0} = \overline{2} = \overline{4} = \cdots$.

So we should get $\varphi(\overline{0}) = \varphi(\overline{2})$. But $\varphi(\overline{2}) = \varphi(\overline{1}+\overline{1}) = \varphi(\overline{1})+\varphi(\overline{1})$, and therefore we get $0 = 2$ in $\mathbb{Z}$, which is a contradiction !

Thus there is **NO** homomorphism $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$.

The crucial thing to understand is that constructing the quotient R/I is in some sense adding new relations between the elements of R. Roughly speaking, you add the relations $a = 0$ whenever $a \in I$. For example, you construct $\mathbb{Z}/2\mathbb{Z}$ from $\mathbb{Z}$ by imposing the extra relation "2=0" in some sense.

Any homomorphism $\varphi : R/I \to R'$ has to respect these new relations to be well-defined, which was not the case in the previous example.

A natural way to construct such a φ would be to start from a ring homomorphism $\psi : R \to R'$, and to set $\varphi : R/I \to R', \overline{r} \mapsto \psi(r)$. It is really easy to see that φ satisfies all the necessary properties to be a ring homomorphism (because ψ satisfies them), **BUT** such a φ is not necessarily a well-defined map !

Indeed, let $\overline{r} \in R/I$. Let $s \in R$ such that $\overline{r} = \overline{s}$, that is s is equivalent to r. Then $s - r \in I$ by definition of the equivalence relation , so $s = r + a, a \in I$.

Now $\varphi(\overline{r}) = \varphi(\overline{s}) = \varphi(\overline{r+a})$, and so $\varphi(\overline{r}) = \varphi(\overline{r}) + \varphi(\overline{a})$. Hence we get $\varphi(\overline{a}) = 0$, that is $\psi(a) = 0$ for all $a \in I$.

Therefore, if it happens that this last condition is not satisfied, then we could get two different images by φ for the same element of R/I ! For example, if $\psi : m \in \mathbb{Z} \mapsto m \in \mathbb{Z}$, and $\varphi : \overline{m} \in \mathbb{Z}/2\mathbb{Z} \mapsto m \in \mathbb{Z}$, then we have $0 = \varphi(\overline{0}) = \varphi(\overline{2}) = 2$!!!

However, if $\psi(a) = 0$ for all $a \in I$, then we get a well-defined map.

Therefore, we have almost proved the following result:

Proposition 3.9. *Let R, R' be two rings, and let I be an ideal of R. Let $\psi : R \to R'$ be a ring homomorphism such that $\ker(\psi) \supset I$, and let $\pi : R \to R/I$ the canonical projection.*

Then there exists a unique well-defined ring homomorphism

$$\overline{\psi} : R/I \to R'$$

such that $\overline{\psi} \circ \pi = \psi$, that is such that $\overline{\psi}(\overline{r}) = \psi(r)$ for all $r \in R$.

Proof. If such $\overline{\psi}$ exists, then it is unique because it has to be defined by the formula $\overline{\psi}(\overline{r}) = \psi(r)$ for all $r \in R$. Now it is enough to check that the map defined by this formula is a well-defined homomorphism. If $\overline{r} = \overline{s} \in R/I$, then $s = r + a$ for some $a \in I$, and then $\overline{\psi}(\overline{s}) = \overline{\psi}(\overline{r + a}) = \overline{\psi}(\overline{r}) + \overline{\psi}(\overline{a})$.

Hence $\overline{\psi}(\overline{s}) = \psi(r) + \psi(a)$. Since $a \in I \subset \ker(\psi)$, we get $\overline{\psi}(\overline{s}) = \psi(r)$, and therefore $\overline{\psi}(\overline{s}) = \overline{\psi}(\overline{r})$, so $\overline{\psi}$ is well-defined. The fact that it is a ring homomorphism is left to the reader. □

Remark: In particular, any ring homomorphism $\psi : R \to R'$ induces a ring homomorphism $\overline{\psi} : R/\ker(\psi) \to R'$.

Lemma 3.10. *Let $\psi : R \to R'$ be a ring homomorphism. Then the ring homomorphism $\overline{\psi} : R/\ker(\psi) \to R'$ is injective.*

Proof. It is enough to prove that its kernel is trivial; but $\overline{\psi}(\overline{r}) = \overline{0} \iff \psi(r) = 0 \iff r \in \ker(\psi) \iff \overline{r} = \overline{0}$. □

Example. Let R be a ring, and let $c = \mathrm{char}(R)$. If $c = 0$, then $\Theta_R : \mathbb{Z} \to R$ is injective, and then $\mathbb{Z}$ can be identified to a subring of R. If $c > 0$, then $\overline{\Theta}_R : \mathbb{Z}/c\mathbb{Z} \to R$ is injective by the previous lemma (because $\ker(\Theta_R) = c\mathbb{Z}$) and therefore $\mathbb{Z}/c\mathbb{Z}$ can be identified to a subring of R.

Corollary 3.11 (First isomorphism theorem)**.** *Let $\psi : R \to R'$ be a ring homomorphism. Then we have a ring isomorphism*

$$R/\ker(\psi) \simeq \mathrm{Im}(\psi)$$

given by $\overline{r} \in R/\ker(\psi) \mapsto \psi(r) \in R'$.

Proof. By the previous lemma, the map $\overline{\psi} : R/\ker(\psi) \to R'$ is an injective ring homomorphism. Therefore it induces an isomomorphism between $R/\ker(\psi)$ and its image. By definition, its image consists of all the elements of R' of the form $\psi(r)$ for some $r \in R$, that is exactly $\mathrm{Im}(\psi)$, so we are done. □

Examples.

(1) If R is any ring, the first isomorphism theorem applied to $\psi = Id_R$ gives that $R/(0) \simeq R$ (do you see why?).

(2) Let $\psi : P \in \mathbb{Q}[X] \mapsto P(i) \in \mathbb{C}$. We saw in the exercise sheets that we have $\ker(\psi) = (X^2+1)$ (the ideal of $\mathbb{Q}[X]$ generated by X^2+1), and $\mathrm{Im}(\psi) = \mathbb{Q}[i]$. Therefore, we get an isomorphism
$$\mathbb{Q}[X]/(X^2+1) \simeq \mathbb{Q}[i].$$

(3) Let $\psi : P \in \mathbb{Z}[X] \mapsto \overline{P} \in \mathbb{F}_2[X]$. We saw in the exercise sheets that we have $\ker(\psi) = (2)$ (the ideal of $\mathbb{Z}[X]$ generated by 2), and $\mathrm{Im}(\psi) = \mathbb{F}_2[X]$. Therefore, we get an isomorphism
$$\mathbb{Z}[X]/(2) \simeq \mathbb{F}_2[X].$$

Example. Let K be a field, and let's try to identify the factor ring $K[X]/(X)$.

The best way to do this is to find a ring homorphism $\psi : K[X] \to R'$, where R' is a suitable ring, such that $\ker(\psi) = (X)$, and use the first isomorphism theorem.

In particular, we need to have $\psi(X) = 0$. Let us consider
$$\psi : P \in K[X] \mapsto P(0) \in K.$$

We have $\psi(P) = 0 \iff P(0) = 0 \iff P$ is a multiple of X (the last equivalence is easy). Hence $\ker(\psi) = (X)$. Now in order to use the first isomorphism theorem, we need to identify the image of ψ. But here ψ is surjective. Indeed, for any $a \in K$, we have $\psi(a) = a$, where a is viewed as a constant polynomial on the left-hand side of the equation.

Hence, applying the first isomorphism theorem to ψ gives
$$K[X]/(X) \simeq K.$$

From now on, all the rings will be commutative!!!!

3.3. Maximal and prime ideals.

Definition. Let R be a ring.

An ideal $\mathfrak{p}$ is called *a prime ideal* if $\mathfrak{p} \neq R$ and for every $a, b \in R$, we have
$$a.b \in \mathfrak{p} \Rightarrow a \in \mathfrak{p} \text{ or } b \in \mathfrak{p}$$

An ideal $\mathfrak{m}$ is called *a maximal ideal* if $\mathfrak{m} \neq R$ and whenever an ideal I satisfies $\mathfrak{m} \subseteq I \subseteq R$, either $I = \mathfrak{m}$ or $I = R$.

Proposition 3.12. 1) *An ideal* $\mathfrak{p}$ *of* R *is prime if and only if* $R/\mathfrak{p}$ *is an integral domain.*

2) *An ideal* $\mathfrak{m}$ *of* R *is maximal if and only if* $R/\mathfrak{m}$ *is a field.*

Proof. 1) Assume that $R/\mathfrak{p}$ is an integral domain. In particular, this is not the trivial ring, and therefore $\mathfrak{p} \neq R$. Now let $a, b \in R$ such that $a.b \in \mathfrak{p}$. Then we have $\overline{a}.\overline{b} = \overline{a.b} = \overline{0}$. Since $R/\mathfrak{p}$ is an integral domain, we get that $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$, which exactly means that $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. Therefore, $\mathfrak{p}$ is a prime ideal.

Conversely, if $\mathfrak{p}$ is a prime ideal, then $\mathfrak{p} \neq R$, so $R/\mathfrak{p}$ is not the trivial ring. Now, if $\overline{a}.\overline{b} = \overline{0}$ in $R/\mathfrak{p}$, then $\overline{a.b} = \overline{0}$, that is $a.b \in \mathfrak{p}$. Since $\mathfrak{p}$ is a prime ideal, we get $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$, that is $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$. Thus $R/\mathfrak{p}$ is an integral domain.

2) Assume that $R/\mathfrak{m}$ is a field. Once again, the factor ring is not the trivial ring in this case, so $\mathfrak{m} \neq R$. Let I be an ideal of R satisfying $\mathfrak{m} \subset I \subset R$, and assume that $I \neq \mathfrak{m}$. We have to show that $I = R$. Let $a \in I$, $a \notin \mathfrak{m}$. Then $\overline{a} \neq \overline{0}$, so $\overline{a}$ is invertible by assumption. Therefore, there exists $\overline{b} \in R/\mathfrak{m}$ such that $\overline{a}.\overline{b} = \overline{1}$. Thus $\overline{1 - a.b} = \overline{0}$, so $1 - a.b \in \mathfrak{m} \subset I$. Since $a \in I$ and I is an ideal, $a.b \in I$, so $1 = (1 - a.b) + a.b \in I$. Hence I contains a unit, so $I = R$, and $\mathfrak{m}$ is a maximal ideal.

Conversely, if $\mathfrak{m}$ is a maximal ideal, then $\mathfrak{m} \neq R$, so $R/\mathfrak{m}$ is not the trivial ring. Now let $\overline{a} \in R/\mathfrak{m}, \overline{a} \neq \overline{0}$. Thus $a \notin \mathfrak{m}$. Let $I = R.a + \mathfrak{m}$; it is an ideal of R (in fact, this is the ideal generated by a and the elements of $\mathfrak{m}$). Then $\mathfrak{m} \subset I$ and $I \neq \mathfrak{m}$, since $a \in I, a \notin \mathfrak{m}$. Since $\mathfrak{m}$ is a maximal ideal, we then get $I = R$. Therefore, I contains 1, so there exists $r \in R$ and $m \in \mathfrak{m}$ such that $1 = r.a + m$. Thus $\overline{1} = \overline{r}.\overline{a}$ and $\overline{a}$ is therefore a unit.

□

Corollary 3.13. *Every maximal ideal is prime.*

Proof. Every field is an integral domain. □

Remark: Using Zorn's lemma, which is equivalent to the Axiom of Choice, it can be shown that any proper ideal of a ring R is contained in a maximal ideal.

Examples.

(1) If K is a field and $\alpha \in K$, the ideal $(X - \alpha)$ is maximal in $K[X]$, since the corresponding quotient ring is isomorphic to K, as we will see in the exercises sheets.
(2) If p is a prime number, (p) is maximal in $\mathbb{Z}$, because the factor ring $\mathbb{Z}/p\mathbb{Z}$ is a field.

(3) $(1+i)$ is maximal in $\mathbb{Z}[i]$: one can show that the quotient is isomorphic to $\mathbb{Z}/2\,\mathbb{Z}$ (see exercise sheets).

(4) (0) is a prime ideal of $\mathbb{Z}$, or of any integral domain R. Note that in $\mathbb{Z}$, $(0) \subset (p)$, so (0) is prime but not maximal.

(5) The ideal (X) in $\mathbb{Z}[X]$, and the ideal (p) in $\mathbb{Z}[X]$, since the respective factor rings are isomorphic to $\mathbb{Z}$ and $\mathbb{F}_p[X]$, which are integral domains. Neither of these is maximal, and both are contained in (p, X) which is a maximal ideal (see exercise sheets).

(6) The ideals (0), (X), (X, Y), (X, Y, Z) are all prime ideals in the ring $\mathbb{Q}[X, Y, Z]$, because the respective factor rings are isomorphic to $\mathbb{Q}[X, Y, Z]$, $\mathbb{Q}[Y, Z]$, $\mathbb{Q}[Z]$ and $\mathbb{Q}$, which are all integral domains. Only (X, Y, Z) is maximal.

4. The field of fractions of an integral domain.

Let R be an integral domain. Let S be the set $R \times (R - \{0\})$. Define an equivalence relation $\sim$ on S by $(a, b) \sim (c, d)$ if and only if $ad = bc$.

Let's check that this is really an equivalence relation. First, we have $(a, b) \sim (a, b)$, since $ab = ba$ (recall that R is commutative). Now if $(a, b) \sim (c, d)$, we have $ad = bc$, and so $cb = da$, which means that $(c, d) \sim (a, b)$. Finally, if $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$ then $ad = bc$ and $cf = de$, so $adf = bcf = bde$. Thus $d(af - be) = 0$, but $d \neq 0$ so $af = be$ since R is an integral domain.

Denote the equivalence class of (a, b) by $\dfrac{a}{b}$.

Therefore, $\dfrac{a}{b} = \dfrac{c}{d} \iff ad = bc$.

We define operations on $S/\sim$ by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$

$$-\frac{a}{b} = \frac{-a}{b},$$

and

$$\frac{a}{b}.\frac{c}{d} = \frac{ac}{bd}.$$

We also set

$$0_{S/\sim} := \frac{0}{1}, 1_{s/\sim} = \frac{1}{1}$$

Notice that the definitions make sense since $b \neq 0$ and $d \neq 0 \Rightarrow bd \neq 0$ (R is integral domain). We now check that these operations are well-defined.

For example, if $\dfrac{a}{b} = \dfrac{a'}{b'}$ and $\dfrac{c}{d} = \dfrac{c'}{d'}$, we have $ab' = ba', cd' = dc'$, so

$$(ad + bc)(b'd') = adb'd' + bcb'd' = ba'dd' + bb'c'd = (a'd' + b'c')(bd),$$

which means that

$$\frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'}, \text{ that is } \frac{a}{b} + \frac{c}{d} = \frac{a'}{b'} + \frac{c'}{d'}.$$

The cases of . and $-$ are similar but easier, and left to the reader.

One can check that $S/\sim$ is a commutative ring, which is not the trivial ring, with the above operations. We won't do it here. Now suppose that $\dfrac{a}{b} \in S/\sim, \dfrac{a}{b} \neq \dfrac{0}{1}$.

This means that $a \neq 0$ (Indeed $\dfrac{a}{b} = \dfrac{0}{1} \iff a.1 = 0.b \iff a = 0$). Then we can consider $\dfrac{b}{a}$, and we have $\dfrac{a}{b}.\dfrac{b}{a} = \dfrac{1}{1}$, so every non-zero element of $S/\sim$ is a unit.

Hence $S/\sim$ is a field. Now we write K_R for $S/\sim$.

Definition. The field of fractions of an integral domain R is the field K_R defined above.

Examples.

(1) $K_{\mathbb{Z}} = \mathbb{Q}$
(2) If K is a field and $R = K[X]$, then $K_R = K(X)$, the field of rational fractions in one variable.
(3) If $R = \mathbb{Z}[\sqrt{d}]$ or $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$then $K_R = \mathbb{Q}[\sqrt{d}]$ (it is a good exercise to check it).

Lemma 4.1. *The map $i \colon R \to K_R$ given by $i(r) = r/1$ is an injective ring homomorphism.*

Proof. If $i(r) = i(s)$, then $(r, 1) \sim (s, 1)$, so $r.1 = 1.s$, and thus $r = s$. Hence it is injective. This is also a ring homomorphism. Indeed:

$$i(r+s) = \frac{r+s}{1} = \frac{r}{1} + \frac{s}{1} = i(r) + i(s),$$
$$i(rs) = \frac{rs}{1} = \frac{r}{1}.\frac{s}{1} = i(r).i(s),$$
$$i(1_R) = \frac{1_R}{1_R} = 1_{K_R}.$$

□

Corollary 4.2. *If F is a field, then i induces an isomorphism $F \simeq K_F$.*

Proof. We already know that i is an injective ring homomorphism, so we only need to prove that it is surjective. Let $\frac{a}{b} \in K_F$. Since $b \neq 0$ and F is a field, b is invertible in F. Thus we have

$$\frac{a}{b} = \frac{ab^{-1}}{bb^{-1}} = \frac{ab^{-1}}{1} = i(ab^{-1}),$$

hence i is surjective. □

Proposition 4.3. *Let F be a field, and let $\theta \colon R \to F$ be an injective ring homomorphism, where R is an integral domain, and F is a field. Then there is a unique homomorphism $\tilde{\theta} \colon K_R \to F$ such that $\tilde{\theta}(\dfrac{a}{1}) = \theta(a)$, and this homomorphism is injective.*

It is defined by $\tilde{\theta}(\dfrac{a}{b}) = \theta(a)(\theta(b))^{-1}$.

Proof. Define $\tilde{\theta}(\frac{a}{b}) = \theta(a)(\theta(b))^{-1}$. We first check this is well-defined: if $\frac{a}{b} = \frac{c}{d}$, then $ad = bc$, so $\theta(ad) = \theta(bc)$. Therefore, $\theta(a)\theta(d) = \theta(b)\theta(c)$, and $\theta(a)(\theta(b)^{-1}) = \theta(c)(\theta(d))^{-1}$.

Now we check that $\tilde{\theta}$ is a ring homomorphism:

$$\tilde{\theta}\left(\frac{a}{b}.\frac{c}{d}\right) = \theta(ac)(\theta(bd))^{-1} = \theta(a)(\theta(b))^{-1}\theta(c)(\theta(d))^{-1} = \tilde{\theta}(\frac{a}{b}).\tilde{\theta}(\frac{c}{d}).$$

The other properties to check are left to the reader as an easy exercise.

Finally, we check uniqueness: If ϕ is another homomorphism $K_R \to F$ extending θ, then $\phi(\frac{r}{1}) = \phi(i(r)) = \theta(r)$. Now for all $r \in R$ and $s \in R - \{0\}$, $\frac{r}{s} = \frac{r}{1}.\frac{1}{s}$, and

$$\phi(\frac{r}{s}) = \phi(\frac{r}{1}).\phi(\frac{1}{s}) = \phi(\frac{r}{1})(\phi(\frac{s}{1}))^{-1} = \theta(r)(\theta(s))^{-1} = \tilde{\theta}(r/s).$$

Finally, by a previous result, $\tilde{\theta}$ is injective since F is not the trivial ring. □

5. Factorisation

5.1. **Prime and irreducible elements.**

Definition. For $a, b \in R$, we say that a *divides* b, denoted by $a|b$, if there exists $\lambda \in R$ such that $b = a\lambda$, or equivalently if $(b) \subseteq (a)$.

Remark. The crucial point in this definition is that $\lambda \in R$.

Examples.

(1) 2 does not divide 3 in $\mathbb{Z}$. Otherwise we would have $3 = 2\lambda$ for $\lambda \in \mathbb{Z}$. But this implies that $\lambda = \frac{3}{2}$, which is not in $\mathbb{Z}$.
(2) $2|3$ in $\mathbb{Q}$ since $3 = 2.\frac{3}{2}$
(3) 2 does not divide $1 + i$ in $\mathbb{Z}[i]$. Otherwise, we would have $1 + i = 2.z, z \in \mathbb{Z}[i]$. Writing $z = a + bi, a, b, \in \mathbb{Z}$, we obtain $1 + i = 2a + 2bi$, so in particular $1 = 2a$ in $\mathbb{Z}$, which is impossible.
(4) $1 + i|2$ in $\mathbb{Z}[i]$ since $2 = (1 + i)(1 - i)$, and $1 - i \in \mathbb{Z}[i]$.

Definition. For $a, b \in R$, we say that a and b are *associate* if $a|b$ and $b|a$, or equivalently, if $(a) = (b)$.

Being associate is an equivalence relation, and the equivalence classes are partially ordered by divisibility.

Examples.

(1) 2 and $1 + i$ are not associate in $\mathbb{Z}[i]$ since 2 does not divide $1 + i$ in $\mathbb{Z}[i]$.
(2) 10 and -10 are associate in $\mathbb{Z}$.
(3) $1 - i$ and $1 + i$ are associate in $\mathbb{Z}[i]$ since we have $1 + i = i(1 - i)$ and $1 - i = -i(1 + i)$.
(4) X and $2X$ are associate in $\mathbb{Q}[X]$ but not in $\mathbb{Z}[X]$ (**do you see why ?**).

Lemma 5.1. *If R is an integral domain, then a and b are associate if and only if $a = bu$ for some unit u in R.*

Proof. If $a = bu, u \in R^*$, then $au^{-1} = b$. So $b|a$ and $a|b$ (in this direction we didn't need to assume that R is an integral domain). Conversely, if $a = b\lambda$ and $b = a\mu$, then $a = a\mu\lambda$, so $a.(1 - \lambda\mu) = 0$. If $a = 0$, then $b = 0$ as well, in which case $a = 1.b$). If $a \neq 0$, then $1 - \lambda\mu = 0$ since R is an integral domain. Hence $\lambda\mu = 1$, and therefore λ and μ are units. □

Definition. An element $\pi \in R$ is irreducible if $\pi \neq 0$, $\pi \notin R^*$, and whenever $\pi = ab$, $a, b \in R$, then either a is a unit or b is a unit.

Example. Any prime number is irreducible in $\mathbb{Z}$.

Remark 5.2. If $\pi \in R$ is irreducible, then for any $u \in R^*$, $u\pi$ is irreducible.

Indeed, we have $u\pi \neq 0$. Otherwise $u\pi = 0$ implies $\pi = 0$, multiplying by u^{-1}, which is not the case since π is irreducible. Now assume that $u\pi = ab, a, b, \in R$. Then $\pi = (u^{-1}a)b$, and since π is irreducible, we get that either $u^{-1}a$ or b is a unit. But if $u^{-1}a$ is a unit, so is $uu^{-1}a = a$.

Remark 5.3. Irreducible elements do not necessarily exist. For example, a field does not contain any irreducible element, since every non zero element is a unit. It also exists some examples of rings which are not fields, and which do not have irreducible elements. For example, the set

$$\mathcal{O} = \{z \in \mathbb{C} \mid \text{ There exists } P \in \mathbb{Z}[X], P \neq 0, P(z) = 0\}$$

can be proven to be a subring of $\mathbb{C}$ (**Difficult!**).

It is a good exercise to check that it is not a field (**Hint:** $1/2 \notin \mathcal{O}$) and that it does not have any irreducible element (**Hint:** If $z \in \mathbb{C}$, let $y \in \mathbb{C}$ such that $z = y^2$. Show that if $z \in \mathcal{O}$, then $y \in \mathcal{O}$).

Definition. An element $\pi \in R$ is prime if $\pi \neq 0$, $\pi \notin R^*$, and whenever $\pi | ab$, $a, b \in R$, then either $\pi | a$ or $\pi | b$.

Example. Once again, any prime number in $\mathbb{Z}$ is prime in this new sense.

Remark 5.4. If $\pi \in R$ is prime, then for all $u \in R^*$, $u\pi$ is prime (**Check it!**).

Lemma 5.5. *Let $\pi \in R$. Then π is a prime element if and only if (π) is a non zero prime ideal.*

Proof. If π is prime, it is non zero, and therefore (π) is a non zero ideal. Moreover, $(\pi) \neq R$. Otherwise, we would have $1 \in (\pi)$, and therefore, we would have an element $r \in R$ such that $1 = \pi r$, meaning that $\pi \in R^*$. This is not the case since π is prime. Hence $(\pi) \neq R$. Now let $a, b \in R$ such that $ab \in (\pi)$. Then $\pi | ab$, so either $\pi | a$ or $\pi | sb$, so either $a \in (\pi)$ or $b \in (\pi)$. Thus (π) is a non zero prime ideal. Conversely, assume that (π) is a non zero prime ideal. Then $\pi \neq 0$. Moreover, π is not a unit since otherwise we would have $(\pi) = R$, which is not the case since (π) is a prime ideal. Now let $a, b \in R$ such that and $\pi | ab$. Then $ab \in (\pi)$, so either $a \in (\pi)$ or $b \in (\pi)$, so either $\pi | a$ or $\pi | b$. Hence π is a prime element of R. □

Lemma 5.6. *Let R be an integral domain. Then every prime element of R is irreducible.*

Proof. Let π be a prime element, and suppose that $\pi = ab$. Need to show that either a or b is a unit. Since $\pi = ab$, then $\pi \mid ab$, now by primality, either $\pi \mid a$ or $\pi \mid b$. Without loss of generality assume $\pi \mid a$. Then $a = \pi\lambda$ for some $\lambda \in R$, so $\pi = ab = \pi\lambda b$, so $\pi.(1 - b\lambda) = 0$. Since R is an integral domain and $\pi \neq 0$, we get $b\lambda = 1$ and hence b is a unit. □

In general, the notions of prime and irreducible are distinct, as we show in the following example:

Important example:

Let $R = \mathbb{Z}[i\sqrt{6}]$. Recall that $R = \{z = a + ib\sqrt{6} \in \mathbb{C}, a, b, \in \mathbb{Z}\}$. We claim that the element 2 is irreducible but not prime. To see that, let's introduce a function $\delta : R \to \mathbb{Z}$, defined by $\delta(a+ib\sqrt{6}) = |a+ib\sqrt{6}|^2 = a^2 + 6b^2$.

By the properties of the modulus, we have $\delta(z.z') = \delta(z).\delta(z')$ for all $z, z' \in R$. Notice that $\delta(z) \in \mathbb{N}$ for all $z \in R$. Indeed, if we write $z = a + ib\sqrt{6}$ for some $a, b \in \mathbb{Z}$, then $\delta(z) = a^2 + 6b^2 \in \mathbb{N}$.

We start we a simple remark: if $z \in R^*$, then $zz' = z'z = 1$ for some $z' \in R$. We then get in particular $\delta(z).\delta'(z') = \delta(1) = 1$, and since $\delta(z), \delta(z') \in \mathbb{N}$, we get $\delta(z) = 1$.

Hence if $z \in R^*$, then $\delta(z) = 1$. In particular 2 is not a unit.

Now suppose that $2 = zz'$, for some $z, z' \in R$. We need to prove that z or z' is a unit of R. Applying δ on both sides of the previous equality gives

$$\delta(2) = 4 = \delta(z)\delta(z')$$

Since $\delta(z), \delta(z') \in \mathbb{N}$, we get that $\delta(z) = 1, 2$ or 4. Write $z = a + ib\sqrt{6}$.

Assume first that $\delta(z) = 2$. We have to solve $a^2 + 6b^2 = 2$ for $a, b, \in \mathbb{Z}$. If $|b| \geq 1$, then $4 = a^2 + 6b^2 \geq 6b^2 \geq 6$, which is a contradiction. Then $b = 0$, and therefore $a^2 = 2$. But 2 is not a square in $\mathbb{Z}$.

So the equation $\delta(z) = 2$ has no solution in R, and hence $\delta(z) = 1$ or 4.

Assume first that $\delta(z) = 1$. Therefore $a^2 + 6b^2 = 1$, and so $b = 0$ (argue as before), $a = \pm 1$, and so $z = \pm 1$ is a unit.

If $\delta(z) = 4$, then $\delta(z') = 1$, and by the previous point, z' is a unit of R.

This prove that 2 is irreducible in R.

On the other hand, 2 is not prime in $\mathbb{Z}[i\sqrt{6}]$, because $2 | i\sqrt{6}.i\sqrt{6}$, but $2 \not| i\sqrt{6}$, as it can be shown easily (**check it !**).

A similar argument can be used to show that 2 is irreducible but not prime in $\mathbb{Z}[i\sqrt{3}]$, since 2 divides the product $(1 - i\sqrt{3})(1 + i\sqrt{3})$ but does not divide either factor.

5.2. **ED, UFD and PID.**

Definition. A ring R is called a *unique factorisation domain*, or *UFD*, if R is an integral domain and

(1) Every $a \in R$ that is neither zero nor a unit can be written as a finite product $a = a_1 \dots a_n$ of irreducible elements of R;

(2) If two such products are equal: $a_1 \dots a_n = b_1 \dots b_m$, then $m = n$, and there is a permutation $\sigma \in S_n$ such that a_i and $b_{\sigma(i)}$ are associate.

Examples include $\mathbb{Z}$, any field, $K[X]$, $K[X_1, \dots, X_n]$. You will have seen this already for $\mathbb{Z}$. The other examples will be verified later in the chapter and the next one.

Proposition 5.7. *Let R be a UFD. Then every irreducible element of R is prime.*

Proof. Let π be a irreducible element. Then $\pi \neq 0, \pi \notin R^*$. it remains to show that for any element $a, b \in R$ such that $\pi | ab$, then $\pi | a$ or $\pi | b$.

Write $ab = \pi c$ for some $c \in R$.

If $a = 0$ or $b = 0$, there is nothing to prove. If a is a unit, then we have $\pi a^{-1} c = b$, so $\pi | b$. Similarly if b is a unit, then $\pi | a$. So assume now that a and b are both non zero, and are not units. Then we can write $a = \pi_1 \cdots \pi_r, b = \pi'_1 \cdots \pi'_s$, where the π_i's are the π'_j's are irreducible. We then have

$$\pi_1 \cdots \pi_r \pi'_1 \cdots \pi'_s = \pi c.$$

If c is a unit then $c\pi$ is irreducible, and uniqueness of the decomposition into product of irreducible elements implies that $c\pi$ is associate to some π_i or π'_j. It easily implies that $\pi | a$ or $\pi | b$.

If c is not a unit, then one can decompose it into products of irreducible elements as well, namely

$$c = \pi''_1 \cdots \pi''_t, \text{ for some irreducible elements } \pi''_1, \dots, \pi''_t \in R.$$

We then have

$$\pi_1 \cdots \pi_r \pi'_1 \cdots \pi'_s = \pi \pi''_1 \cdots \pi''_t,$$

and uniqueness of the decomposition into product of irreducible elements implies this time that π is associate to some π_i or π'_j, and we conclude as before. □

Recall that we showed that in any integral domain, prime elements are always irreducible. Thus in a UFD, prime and irreducible are equivalent. This can be useful to prove that a give ring is **NOT** a UFD.

Example. Recall that we showed that in $\mathbb{Z}[i\sqrt{6}]$, 2 is irreducible but not prime. It follows that $\mathbb{Z}[i\sqrt{6}]$ cannot be a UFD. Note that $6 = 2.3 = -i\sqrt{6}i\sqrt{6}$ has two distinct expressions as a product of irreducibles (one can show that 3 is irreducible).

We now would like to precise the result of the previous proposition. We start with a lemma.

Lemma 5.8. *Let R be an integral domain. Assume that every irreducible element of R is prime. Then for all integers $r, s \geq 1$ and all irreducible elements $\pi_1, \ldots, \pi_r, \pi'_1, \ldots, \pi'_s \in R$ such that*

$$\pi_1 \cdots \pi_r = \pi'_1 \cdots \pi'_s,$$

we have $r = s$ and there exists a permutation $\sigma \in S_r$ such that π_i and $\pi'_{\sigma(i)}$ are associate.

Proof. We proceed by induction on r. More precisely, the induction hypothesis is the following one:

(H_r) If we have an equality

$$\pi_1 \cdots \pi_r = \pi'_1 \cdots \pi'_s,$$

with $s \geq r$ and the $\pi_1, \ldots, \pi_r, \pi'_1, \ldots \pi'_s$ are irreducible, then $r = s$ and there is a permutation $\sigma \in S_r$ such that π_i is associate to $\pi'_{\sigma(i)}$.

Assume first that $r = 1$. We then have

$$\pi_1 = \pi'_1 \cdots \pi'_s.$$

Since $\pi_1 | \pi'_1 \cdots \pi'_s$, then π_1 divides one of the elements π'_j by assumption, say π'_1 (that we may always assume after renumbering). We then have

$$\pi'_1 = u\pi_1, u_1 \in R.$$

Since π'_1 is irreducible, we have $\pi_1 \in R^*$ or $u \in R^*$. Since π_1 is irreducible, it is not a unit and so $u \in R^*$. Hence, π_1 et π'_1 are associate. If $s = 1$, we are done. If $s \geq 2$, after simplification by π_1 (which is possible since R is an integral domain), on obtient

$$1 = u\pi'_2 \ldots \pi'_s.$$

This implies that π'_2 is a unit, which contradicts the fact that π'_2 is irreducible. Hence $s = 1$ and (H_1) is proved.

Assume now that (H_r) is true for some $r \geq 1$, and let us show that (H_{r+1}) is true as well. Let us consider an equality of the type

$$\pi_1 \cdots \pi_{r+1} = \pi'_1 \cdots \pi'_s,$$

where $s \geq r + 1 \geq 2$. Reasoning as previosuly, we see that π_1 is associate to some π'_j, say π'_1. We then have $\pi'_1 = u\pi_1, u \in R^*$, and after simplification by π_1, we get

$$\pi_2 \cdots \pi_{r+1} = (u\pi'_2)\pi'_3 \cdots \pi'_s.$$

Since $u\pi'_2$ is irreducible, we get by induction $r = s$ and each $\pi_i, i \geq 2$ is associate to a unique $\pi'_j, j \geq 3$ or to $u\pi'_2$, that is associate to a unique

$\pi'_j, j \geq 2$. Sinceπ_1 is associate to π'_1, this concludes the proof of (H_{r+1}), and the proof by induction. □

We then have the folowing characterization of UFDs.

Theorem 5.9. *Let R be an integral domain. Then R is a UFD if and only if the two following conditions are satisfied:*

(1) *Every $a \in R$ that is neither zero nor a unit can be written as a finite product $a = a_1 \dots a_n$ of irreducible elements of R.*

(2) *Every irreducible element of R is prime.*

Proof. A UFD satisfies condition (1) by definition, and condition (2) by Proposition 5.7. Conversely, an integral domain satisfying (1) and (2) is a UFD, since the uniqueness of the decomposition of a is ensured by the previous lemma. □

Definition. Let R be a ring. *A complete system of irreducible elements of R* is a set $\mathcal{P}$ satisfying the following properties:

1) Every element of $\mathcal{P}$ is irreducible

2) Two different elements of $\mathcal{P}$ are non-associate

3) Every irreducible element of R is associate to an element of $\mathcal{P}$

Remark 5.10. Such $\mathcal{P}$ always exists. Indeed, consider all the equivalence classes of the set of irreducible elements under the relation 'being associate', and construct your set $\mathcal{P}$ by choosing exactly one element in each equivalence class. Then by definition, $\mathcal{P}$ satisfies 1), 2) and 3).

You have in general infinitely many different choices for $\mathcal{P}$.

Examples. The following sets are complete systems of irreducible elements for $\mathbb{Z}$:

(1) $\mathcal{P} = \{p, p \text{ prime number }\}$
(2) $\mathcal{P} = \{-p, p \text{ prime number }\}$
(3) $\mathcal{P} = \{-2, 3, 5, 7, 11, 13, \cdots\}$

Proposition 5.11. *Let R be a UFD, and let $\mathcal{P}$ be a complete system of irreducible elements. Then every $a \in R, a \neq 0$ can be written in a unique way as*

$$a = u \prod_{\pi \in \mathcal{P}} \pi^{n_\pi}$$

where $u \in R^*$ and the n_π's are non-negative integers which are almost all zero (except a finite number).

Proof. If a is a unit, we just take $u = a$ and $n_\pi = 0$ for all $\pi \in \mathcal{P}$. Now assume that a is not a unit. Since R is a UFD, then $a = \pi'_1 \cdots \pi'_s$ where the π'_j's are irreducible. Now each π'_j is associate to an element of $\mathcal{P}$.

Since R is an integral domain, this is equivalent to say that $\pi'_j = u_j\pi_j$ where $u_j \in R^*$ and $\pi_j \in \mathcal{P}$. Now set u to be the product of the u'_js and collect all the π'_js which are equal. We then get a in the desired form.

To prove uniqueness, let

$$a = u\pi_1^{n_1}\cdots\pi_k^{n_k} = v\pi_1^{'m_1}\cdots\pi_r^{'m_r},$$

where the π_i's (resp. the π'_j's) are pairwise distinct elements of $\mathcal{P}$, and $u, v \in R^*$.

One can always assume that $r = k$ and $\pi_i = \pi'_i$ for all i, renumbering and writing some terms of the form π_i^0 or $\pi_j^{'0}$ if necessary.

Now if $n_i \neq m_i$ for some i, then after simplification by a suitable power of π_i , we would obtain an equality of the form

$$(\text{unit}).\pi_i^s.(\text{product of } \pi_j\text{'s}, j \neq i) = (\text{unit}).(\text{product of } \pi_j\text{'s}, j \neq i)$$

with $s \geq 1$.

So π divides the right hand side. Since R is a UFD and π_i is irreducible, then π_i is prime and therefore divides one of the π'_js (it cannot divide the unit, otherwise π_i would be a unit itself, which is not the case since π_i is irreducible). Hence $\pi_j = c\pi_i$ for some $j \neq i$ and $c \in R$. Since π_j is irreducible, it necessarily implies that c is a unit (do you see why?), and so π_j and π_i are associate, which is impossible by choice of $\mathcal{P}$. Hence $n_i = m_i$ for all i, and we get finally $u = v$ after simplification, hence the uniqueness is established. □

Examples.

The decomposition obviously depends on the choice of $\mathcal{P}$. Take $R = \mathbb{Z}$ and consider the three previous choices of $\mathcal{P}$:

(1) If $\mathcal{P} = \{p, p \text{ prime number }\}$, then 6 decomposes as $6 = 2.3$
(2) If $\mathcal{P} = \{-p, p \text{ prime number }\}$, then 6 decomposes as $6 = (-2).(-3)$
(3) $\mathcal{P} = \{-2, 3, 5, 7, 11, 13, \cdots\}$, then 6 decomposes as $6 = -1.(-2).(3)$

Definition. Let R be a ring. If $a, b \in R$, then $d \in R$ is called a *highest common factor* or *h.c.f.* of a and b if

(1) $d|a$ and $d|b$;

(2) Whenever $c|a$ and $c|b$, then $c|d$.

Remark 5.12. A *h.c.f.* does not necessarily exist. For example, one can show that 9 and $3(2 + i\sqrt{5})$ do not have any h.c.f. in $\mathbb{Z}[i\sqrt{5}]$ (this is not totally immediate).

If moreover a h.c.f. does exist, it is not unique. Indeed, one may check that if d is a h.c.f. of a and b, then ud is also a h.c.f. of ab and b for all $u \in R^*$.

Proposition 5.13. *Let R be a UFD, and $a, b \in R$. If $(a, b) \neq (0, 0)$, then a, b have a h.c.f. It is unique up to multiplication by a unit of R.*

Proof. If d and d' are two *h.c.f.* of a and b, then $d|d'$ and $d'|d$ by definition, so d and d' are associate, and therefore differ by a unit, since R is an integral domain. Hence if a *h.c.f.* of a and b exists, then it is unique up to multiplication by a unit of R.

Now let us prove the existence. If $a = 0$, then clearly a *h.c.f.* of a and b is b. Similarly, if $b = 0$, then then clearly a *h.c.f.* of a and b is a. (**Check it**!) Now assume that a and b are both non zero, and let $\mathcal{P}$ be a complete system of irreducible elements of R. Write

$$a = u \prod_{\pi \in \mathcal{P}} \pi^{n_\pi}, b = v \prod_{\pi \in \mathcal{P}} \pi^{m_\pi}$$

and let $d = \prod_{\pi \in \mathcal{P}} \pi^{k_\pi}$, where k_π is the minimum of n_π and m_π. It is easy to check that d is a h.c.f. of a and b (**Do it !**). □

Recall that an ideal in a ring R is said to be *principal* if it can be generated by a single element.

Definition. A ring R is a *principal integral domain* or *PID* if R is an integral domain and every ideal of R is principal.

Example. The ring $\mathbb{Z}$ is a PID, as is any field.

Definition. A ring R is a *Euclidean domain* or *ED* if R is an integral domain and it has a norm function $\delta : R - \{0\} \to \mathbb{N}$ (remember that $0 \in \mathbb{N}$) such that for all $a, b \in R - \{0\}$, there exist elements $q, r \in R$ with $a = qb + r$ and either $r = 0$ or $\delta(r) < \delta(b)$.

In this case, we say that R is Euclidean with respect to δ. The elements q and r are somatimes called respectively a *quotient* and a *remainder.*

Remark 5.14. We do not require q and r to be unique. In fact, they are not unique in general.

Examples.

(1) $\mathbb{Z}$ is a Euclidean domain with norm given by $\delta(n) = |n|$.
(2) Any field is a Euclidean domain with norm given by $\delta(r) = 1$ for all non-zero r.
(3) For any field K, the polynomial ring $K[X]$ is a Euclidean domain with norm $\delta(f) = \deg(f)$, as we will see later.

Theorem 5.15. *Every Euclidean domain is a PID.*

Proof. Let R be a Euclidean domain, I an ideal of R. If $I = (0)$, then I is principal. If $I \neq 0$, pick $b \in I, b \neq 0$ such that $\delta(b) \leq \delta(a)$ for all $a \in I - \{0\}$ (We can do this, because any set of elements of $\mathbb{N}$ has a least element). We claim that $(b) = I$: if $a \in I$, then there exist q and r such that $a = qb + r$, where either $r = 0$ or $\delta(r) < \delta(b)$. But $a, b \in I$, so $-qb \in I$ and therefore $r = a - qb \in I$, since I is an ideal. Since $\delta(r) < \delta(b)$, we cannot have $r \neq 0$, since this would contradict the choice of b. Hence we must have $r = 0$ and $a = qb$. Hence $I \subset (b)$. Conversely, since $b \in I$ and I is an ideal $(b) = bR \subset I$. Hence we get $I = (b)$, so I is principal. □

This is proof is just a generalization of the argument we used earlier to classify ideals of $\mathbb{Z}$.

Theorem 5.16. *Every PID is a UFD.*

Proof. Step 1: We show that every non zero element which is not a unit decomposes as a product of irreducible elements. We will do this by assuming that this property does not hold and obtaining a contradiction. Let R' be the set of non zero elements of R which are not units, and cannot be factored as a product of irreducible elements. We assume that R' is not empty. Let $\mathcal{F} = \{(a), a \in R'\}$. By assumption this set is not empty. Consider a non empty chain of elements of $\mathcal{F}$, that is a family of elements $(a_i)_{i \geq 0} \in \mathcal{F}$ which is totally ordered for the inclusion: $(a_1) \subseteq (a_2) \subseteq \cdots (a_i) \subseteq \cdots$.

Consider $I = \bigcup_{i \geq 1} (a_i)$. This is an ideal since the ideals (a_i) form of chain (this would be not true otherwise!!!!). By assumption, $I = (x)$ for some $x \in R$. In particular $x \in (a_j)$ for some j and therefore $(x) \subset (a_j) \subset I = (x)$ and $(x) = (a_j)$. Assume now that there exists $i > j$ for which $(a_j) \subsetneq (a_i)$. Then we would have $I = (x) = (a_j) \subsetneq (a_i) \subset I$, so we would get $I \subsetneq I$, which is a contradiction. Therefore, for all $i > j$, we have $(a_j) = (a_i)$, which means that the chain is finite.

So there exists a chain of elements of $\mathcal{F}$, say $(a_1) \subsetneq (a_2) \subsetneq \cdots \subsetneq (a_n)$ which cannot be extended. Otherwise, we could construct by induction a chain of infinite length, which is a contradiction.

Since $a_n \in R'$, a_n is not a unit and is not irreducible. Therefore $a_n = b.c$ for some non-zero $b, c \in R$ which are not units. Assume that $(a_n) = (b)$. Then $a_n = u.b, u \in R^*$, therefore $b.c = b.u$ and then $b.(c - u) = 0$, so $c = u$ (R is an integral domain). Therefore $c \in R^*$, a contradiction. We then have $(a_n) \subsetneq (b)$. In this case, $b \notin R'$ (otherwise the chain $(a_1) \subsetneq \cdots \subsetneq (a_n)$ of elements of $\mathcal{F}$ can be extended into the chain $(a_1) \subsetneq \cdots \subsetneq (a_n) \subsetneq (b)$). It means that b can be factored. A similar reasoning shows that c can be factored as well, so is $bc = a_n$, which is a contradiction.

Step 2: Assume that $p_1 \cdots p_n = q_1 \cdots q_m$, where the p_i's and the q_j's are irreducible elements. We can assume $n \geq m$.

Consider the ideal (p_1, q_j). By assumption, $(p_1, q_j) = (b_j)$ for some $b_j \in R$. Therefore, $p_1 = b_j c_j$ for some $c_j \in R$. Since p_1 is irreducible, we get $b_j \in R^*$ or $c_j \in R^*$.

Assume first that $b_j \in R^*$ for all j. So $(p_1, q_j) = R$ for all j. In particular, we have $1 = u_j p_1 + v_j q_j$ for all j. Multiplying all these relations, we get that $1 = up_1 + vq_1 \cdots q_m$. So $\overline{vq_1 \cdots q_m} = \overline{1}$ in $R/(p_1)$. In particular, $\overline{q_1 \cdots q_m} \neq \overline{0}$ in $R/(p_1)$. But this is a contradiction, since $q_1 \cdots q_m = p_1 \cdots p_n \in (p_1)$.

Therefore, there exists some j for which $b_j \notin R^*$, and so $c_j \in R^*$. In this case $(p_1) = (b_j) = (p_1, q_j)$, so $q_j \in (p_1)$ and therefore $q_j = cp_1$. Since p_1 and q_j are irreducible, c is a unit and p_1 and q_j are associate. Now we simplify the relation by p_1 and repeat the same process again and again. We then show that $p_1, \cdots, p_m$ are associate with some q_j's. If $n > m$, we will end up with a relation of the form "product of irreducibles= unit", which would imply that the remaining irreducible elements are invertible, a contradiction. Therefore $n = m$ and we are done. □

One can be summarise things as follows: ED ⇒ PID ⇒ UFD.

Warning: One can prove that $\mathbb{Z}[\frac{-1+i\sqrt{19}}{2}]$ is a PID but not an ED for any norm map(difficult!!!). Moreover, we will see later that $\mathbb{Z}[X]$ is a UFD; however this is not a PID:

Exercise: Prove by a way of contradiction that $(2, X)$ is not a principal ideal of $\mathbb{Z}[X]$.

Given the name, it won't be a surprise that a Euclidean domain turns out to be a ring in which there is a version of Euclid's algorithm.

Proposition 5.17. *(Euclid's algorithm). Let R be a Euclidean domain. There is an algorithm for finding a h.c.f. d of any two elements $a, b \in R$, without factorizing a and b. Moreover, there are $u, v \in R$ such that $d = au + bv$.*

Proof. We can assume that a and b or both non zero, otherwise d is easily computed. The algorithm is then as follows:

Suppose that $\delta(a) \geq \delta(b)$. Set $r_{-1} = a, r_0 = b$.

Whenever it is possible (that is whenever the new remainder is $\neq 0$), we proceed to the following divisions:

$$r_{-1} = q_0.r_0 + r_1, q_0 \in R, \delta(r_1) < \delta(r_0)$$
$$r_0 = q_1.r_1 + r_2, q_1 \in R, \delta(r_2) < \delta(r_1)$$
$$\vdots$$

$r_{i-1} = q_i.r_i + r_{i+1}, q_i \in R, \delta(r_{i+1}) < \delta(r_i)$
$\vdots$
$r_{n-2} = q_{n-1}.r_{n-1} + r_n, q_n \in R, \delta(r_n) < \delta(r_{n-1})$
$r_{n-1} = q_n.r_n + 0$ (that is $r_{n+1} = 0$).

Here n is the least integer such that $r_n \neq 0$ but $r_{n+1} = 0$.

This integer n necessarily exists, since $\delta(r_0), \delta(r_1), \delta(r_2), \ldots$ is a strictly decreasing sequence of non-negative integers.

We claim that $d = r_n$ is a h.c.f. for a and b.

- We first prove that for all $i \geq -1$, $r_i = a.u_i + b.v_i$ for some $u_i, v_i \in R$.

We proceed by induction. Our induction hypothesis (H_i) is the following:

$$(H_i)\ r_{i-1} = a.u_{i-1} + b.v_{i-1} \text{ and } r_i = a.u_i + b.v_i$$

First, (H_0) is true. Indeed, $r_{-1} = a.1 + b.0$ and $r_0 = a.0 + b.1$.

Now assume that (H_i) is true for some $i \geq 0$, and let us prove that (H_{i+1}) is true.

We would like to prove

$$(H_{i+1})\ r_i = a.u_i + b.v_i \text{ and } r_{i+1} = a.u_{i+1} + b.v_{i+1}$$

Since we are assuming that (H_i) is true, we already know the first part. Now we have $r_{i-1} = q_i.r_i + r_{i+1}$, so $r_{i+1} = r_{i-1} - q_i.r_i$ and therefore $r_{i+1} = a.u_{i-1} + b.v_{i-1}) - q_i.(a.u_i + b.v_i)$.

Thus $r_{i+1} = a.u_{i+1} + b.v_{i+1}$, with $u_{i+1} = u_{i-1} - q_i.u_i$ and $v_{i+1} = v_{i-1} - q_i.v_i$.

- We now prove that r_n divides a and b.

We proceed by induction. This time, our induction hypothesis will be

$$(H_i)\ r_n | r_{n-i} \text{ and } r_n | r_{n-i-1}$$

First, (H_0) is true. Indeed $r_n | r_n$ and by definition of n, we have $r_{n-1} = q_n.r_n$, so $r_n | r_{n-1}$ as well.

Now assume that (H_i) is true for some $i \geq 0$, and let's prove that (H_{i+1}) is true.

We want to prove

$$(H_{i+1})\ r_n | r_{n-i-1} \text{ and } r_n | r_{n-i-2}$$

The first part is known from (H_i). Now we have $r_{n-i-2} = q_{n-i-1}.r_{n-i-1} + r_{n-i}$.

Since r_n divides r_{n-i} and r_{n-i-1} by induction hypothesis, then it divides r_{n-i-2}.

Therefore (H_{i+1}) is true.

It follows that in particular (H_n) is true, that is r_n divides $r_{-1} = a$ and $r_0 = b$.

- We are now ready to conclude. By the previous point, $r_n|a$ and $r_n|b$. Now if $c|a$ and $c|b$, then from the equation $r_n = a.u_n + b.v_n$, we get that $c|r_n$.

Therefore, $d = r_n$ is a h.c.f. of a and b. Moreover $d = u.a + v.b$ for some $u, v \in R$ (take $u = u_n$ and $v = v_n$). □

Manual disposition

There is a convenient way to perform the algorithm and compute $d = h.c.f(a, b)$ and some $u, v \in R$ satisfying $d = a.u. + b.v$ at once.

From the proof, we define sequences $(r_i)_{i\geq -1}, (q_i)_{i\geq 0}, (u_i)_{i\geq -1}, (v_i)_{i\geq -1}$ of elements of R as follows

$$\begin{gathered} r_{-1} = a, r_0 = b, r_{i-1} = q_i.r_i + r_{i+1}, \\ u_{-1} = 1, u_0 = 0, u_i = u_{i-1} - q_i.u_i, \\ v_{-1} = 0, v_0 = 1, v_i = v_{i-1} - q_i.v_i. \end{gathered}$$

We then construct an array with 4 columns corresponding to the values $-q_i, r_i, u_i, v_i$, and with rows indexed by integers $i \geq -1$ starting like this:

$$\begin{array}{cccc} -q_i & r_i & u_i & v_i \\ & a & 1 & 0 \\ & b & 0 & 1 \end{array}$$

We will workout a concrete example to illustrate the procedure at the same time: $R = \mathbb{Z}, a = 32, b = 7$.

So we have

$$\begin{array}{cccc} -q_i & r_i & u_i & v_i \\ & 32 & 1 & 0 \\ & 7 & 0 & 1 \end{array}$$

Notice that the value q_{-1} is not defined. The next step is to compute q_0. It can be not totally immediate, but it is easy for $R = \mathbb{Z}$ or $K[X]$, and there are some recipes for other rings that we will give later.

In our example, $32 = 4 \times 7 + 4$, so $q_0 = 4$.

Hence the next step gives

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1

Now to compute the values r_1, u_1, v_1, we multiply all the elements of the last line by the first one, and we add the corresponding elements lying above them.

For example, the next value of r_i is $-4 \times 7 + 32$, the next value of u_i is $-4 \times 0 + 1$, and the next value of v_i is $-4 \times 1 + 0$. So we get

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1
	4	1	-4

Now we need to compute the values for the next line. For this we consider, only the two last lines, and we compute the new quotient, that is we divide the two last values of r_i.

In this example, we have to divide 7 by 4. So $7 = 1 \times 4 + 3$, and the new quotient is 1.

Then we get

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1
-1	4	1	-4

Now we repeat the previous procedure to compute the new values for the next line. The new value for r_i is $-1 \times 4 + 7$, the new value for u_i is $-1 \times 1 + 0$, and the new value for v_i is $-1 \times (-4) + 1$.

Hence we get

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1
-1	4	1	-4
	3	-1	5

To find the new quotient, we have to divide 4 by 3: $4 = 1 \times 3 + 1$, so the new quotient is 1.

Then we have

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1
-1	4	1	-4
-1	3	-1	5

The new value for r_i is $-1 \times 3 + 4$, the new value for u_i is $-1 \times (-1) + 1$, and the new value for v_i is $-1 \times 5 + (-4)$.

Then we get

$-q_i$	r_i	u_i	v_i
	32	1	0
-4	7	0	1
-1	4	1	-4
-1	3	-1	5
	1	2	-9

To find the new quotient, we need to divide 3 by 1. But $3 = 3 \times 1 + 0$, so the new remainder is 0 and we stop here.

Hence a h.c.f of 32 and 7 is 1 (the last r_i) and we have $u = 2$ (the last u_i) and $v = -9$ (the last v_i).

Let us check: $32 \times 2 - 7 \times (-9) = 64 - 63 = 1$!!!

Remark: The method works for any ED. The problem is to find a way to compute the quotients, which can be tricky sometimes.

Let j be the complex number given by the following equivalent definitions:

$$j = e^{2\pi i/3} = \frac{-1 + i\sqrt{3}}{2}.$$

Notice that j is a root of the polynomial $X^2 + X + 1$.

We already know that $\mathbb{Z}[j] = \{a + bj, a, b \in \mathbb{Z}\}$.

Proposition 5.18. 1) *Let $R = \mathbb{Z}[i], \mathbb{Z}[j]$ or $\mathbb{Z}[i\sqrt{2}]$. Then R is an Euclidean domain for the norm $\delta(z) = |z|^2$.*

2) *Let $R = \mathbb{Z}[\sqrt{2}]$. Then R is an Euclidean domain for the norm $\delta(a + b\sqrt{2}) = |(a + b\sqrt{2})(a - b\sqrt{2})| = |a^2 - 2b^2|$.*

Proof. I will do only the second case.

Notice first that δ can be defined for elements of $\mathbb{Q}[\sqrt{2}]$ (i.e. for $a, b \in \mathbb{Q}$) in the same way, and that we have $\delta(z.z') = \delta(z).\delta(z')$ for all $z, z' \in \mathbb{Q}[\sqrt{2}]$.

Now consider $x = a + b\sqrt{2}, y = c + d\sqrt{2}$, with $\delta(y) \leq \delta(x)$. We have

$$\frac{x}{y} = \frac{(a + b\sqrt{2})(c - d\sqrt{2})}{c^2 - 2d^2} = u + v\sqrt{2}, u, v \in \mathbb{Q}\,.$$

Let m (resp. n) be the closest integer from u (resp. from v) (do you see why such m, n always exist ? Such m, n are unique, except if u or v equal 1/2. In this latter case, we have two choices). Set $q = m + n\sqrt{2}$. Now set $r = x - qy$. We have $\delta(r) = \delta(y(x/y - q)) = \delta(y)\delta(x/y - q)$. But $\delta(x/y - q) = \delta((u - m) + (v - n)\sqrt{2}) = |(u - m)^2 - 2(v - n)^2|$.

By definition of u, v, we have $|u - m| \leq 1/2, |v - n| \leq 1/2$. Therefore

$$|(u - m)^2 - 2(v - n)^2| \leq |u - m|^2 + 2.|v - n|^2 \leq 1/4 + 2 \times 1/4 = 3/4.$$

Consequently $\delta(r) \leq \frac{3}{4}\delta(y) < \delta(y)$.

□

Example. We can now illustrate the fact that a quotient and a remainder are not necessarily determined. Let $R = \mathbb{Z}[\sqrt{2}]$ with the norm defined above. Let $a = 1 + 2\sqrt{2}$, and $b = 2$. Let us divide a by b. To find a quotient, we proceed as in the proof of the previous proposition. We have

$$\frac{a}{b} = \frac{1 + 2\sqrt{2}}{2} = 1 + \frac{1}{2} + \sqrt{2}.$$

We now have to choose integers $m, n \in \mathbb{Z}$ such that

$$|\frac{1}{2} - m| \leq \frac{1}{2} \text{ and } |1 - n| \leq \frac{1}{2}.$$

We do not have a choice for n: necessarily, $n = 1$. However, we could take either $m = 0$ or $m = 1$.

If we choose $m = 0$, then we get $q = \sqrt{2}$, and then $r = a - qb = 1$.

If we choose $m = 1$, then we get $q = 1 + \sqrt{2}$, and then $r = a - qb = -1$.

Notice that in both cases, we have $\delta(r) = 1 < \delta(b) = 4$. Hence we get two possible results for the division of a by b:

$$1 + 2\sqrt{2} = \sqrt{2} \cdot 2 + 1 \text{ or } 1 + 2\sqrt{2} = (1 + \sqrt{2}) \cdot 2 - 1.$$

Remark 5.19. Since the following rings are Euclidean domains, it follows that they are principal ideal domains too:

$$\mathbb{Z}, K[X], \mathbb{Z}[i], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i\sqrt{2}], \mathbb{Z}[j].$$

5.3. **Applying Euclid's algorithm in $\mathbb{Z}[\sqrt{d}], d = -1, 2, -2$ and $\mathbb{Z}[j]$.** Euclid's algorithm can be applied directly to find the h.c.f. of two elements of these rings.

We have to find a way to compute quotients. For $\mathbb{Z}$ and $K[X]$, there is no particular problem.

Here is some recipes for $\mathbb{Z}[i], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i\sqrt{2}]$ and $\mathbb{Z}[j]$. The following recipe could seem a bit strange and coming from nowhere, but in fact it is not. Actually, in the proofs (that I didn't explain for $\mathbb{Z}[i]$ and $\mathbb{Z}[j]$), of the fact that these rings are ED, the way to construct a quotient and a remainder is exactly the following one:

Let $a, b \in R = \mathbb{Z}[i], \mathbb{Z}[\sqrt{2}], \mathbb{Z}[i\sqrt{2}]$ or $\mathbb{Z}[j]$.

Step 1: Write $\frac{a}{b}$ under the form $x + iy, x + y\sqrt{2}, x + iy\sqrt{2}$ or $x + jy$, with $x, y \in \mathbb{Q}$ (it's always possible).

Step 2: Set $q = m + in, m + n\sqrt{2}, m + in\sqrt{2}$ or $m + jn$, with $m, n \in \mathbb{Z}$, where m, n satisfy $|u - m| \leq 1/2, |v - n| \leq 1/2$. Set $r = z - q.b$.

For example, let's compute the highest common factor of $5 + j$ and 14 in $\mathbb{Z}[j]$.

Recall that $j^2 + j + 1 = 0, \bar{j} = j^2 = -1 - j$ and $\delta(a + bj) = a^2 - ab + b^2$.

We have easily $\delta(14) \geq \delta(5 + j)$.

So we set $r_{-1} = 14, r_0 = 5 + j$.

$$\frac{14}{5 + j} = \frac{14.(5 + \bar{j})}{21} = \frac{2}{3}(4 - j) = \frac{8}{3} - \frac{2}{3}j,$$

so $u = \frac{8}{3}, v = -\frac{2}{3}$. Therefore $m = 3, n = -1$.

So we set $q_0 = 3 - j$, and the remainder is

$r_1 = 14 - (3 - j).(5 + j) = 14 - (15 - 2j - j^2) = 14 - (16 - j) = -2 + j.$

But now

$$\frac{5 + j}{-2 + j} = \frac{(5 + j).(-2 + \bar{j})}{7} = \frac{-10 - 5 - 5j - 2j + 1}{7} = \frac{-14 - 7j}{7} = -2 - j.$$

and so $-2 + j$ divides exactly into $5 + j$ in the ring $\mathbb{Z}[j]$. Hence the next remainder is 0 and $-2 + j$ is a h.c.f. for $5 + j$ and 14.

Another example: to find a h.c.f. for $3 + 4i$ and 5 in $\mathbb{Z}[i]$:

$$\frac{3 + 4i}{5} = \frac{3}{5} + \frac{4}{5}i,$$

so we have $u = \frac{3}{5}, v = \frac{4}{5}$, and we can take $m = 1, n = 1$. So $q_0 = 1 + i$, and the remainder is

$$r_1 = (3 + 4i) - 5(1 + i) = -2 - i.$$

Now we have

$$\frac{5}{-2-i} = \frac{5(-2+i)}{(-2-i)(-2+i)} = \frac{5(-2+i)}{5} = -2+i,$$

so the new remainder is 0.

The last non-zero remainder was $-2-i$, so this is a h.c.f. of $3+4i$ and 5 in $\mathbb{Z}[i]$.

5.4. **Some useful tricks.** This section is purely informal, and will give you some tricks that you can use in exercises about factorisation in $\mathbb{Z}[\sqrt{d}]$ or $\mathbb{Z}[j]$.

Let $R = \mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d}, a, b, \in \mathbb{Z}\}$ or $\mathbb{Z}[j] = \{a + bj, a, b, \in \mathbb{Z}\}$.

We have a standard norm on $\mathbb{Z}[\sqrt{d}]$, which is

$$\delta(a+b\sqrt{d}) = |(a+b\sqrt{d})(a-b\sqrt{d})| = |a^2 - d.b^2| \in \mathbb{N}.$$

In the case of the ring $\mathbb{Z}[j]$, we set $\delta(a+bj) = |a+bj|^2 = a^2 - ab + b^2$.

For example, if we consider $\mathbb{Z}[i], \mathbb{Z}[j]$ or $\mathbb{Z}[i\sqrt{2}]$, we get simply $\delta(z) = |z|^2$. If $\mathbb{Z}[\sqrt{2}]$, we recover the δ we define in the last section.

We can now use δ to solve several questions.

Trick 1: To find units of $R = \mathbb{Z}[\sqrt{d}]$ or $\mathbb{Z}[j]$:

For $z \in R$, if $z \in R^*$, then $\delta(z) = 1$.

Indeed, assume $zz' = 1$ in R. Then $\delta(z.z') = \delta(z).\delta(z') = \delta(1) = 1$. Since $\delta(z), \delta(z') \in \mathbb{N}$, the only possibility is $\delta(z) = 1$.

You will have to do this in each case, since this is just a trick, and not a general theorem that you may use.

The next step is to solve the equation $\delta(z) = 1$ for $z \in R$, and then to check that all the solutions you obtain are effectively units (by finding the inverse).

In fact, one can show that if $\delta(z) = 1$, then z is a unit...but you are not supposed to know that, so you will have to prove this in each particular case (which is done by the last step).

Example: Let $R = \mathbb{Z}[i\sqrt{6}]$. We want to compute R^*. We have $\delta(a + bi\sqrt{6}) = |a + bi\sqrt{6}|^2 = a^2 + 6b^2$.

If $z, z' \in R$, we have

$$\delta(z.z') = |z.z'|^2 = (|z|.|z'|)^2 = |z|^2.|z'|^2 = \delta(z).\delta(z').$$

Let $z \in R^*$, so $z.z' = 1$ for some $z' \in R$. Thus we have $\delta(z.z') = \delta(z).\delta(z') = \delta(1) = 1$. Since $\delta(z), \delta(z') \in \mathbb{N}$, the only possibility is $\delta(z) = 1$.

Write $z = a + ib\sqrt{6}$. We then have $a^2 + 6b^2 = 1$. If $b \neq 0$, then $|b| \geq 1$, and we would get $a^2 + 6b^2 \geq 6b^2 \geq 6$. So $b = 0$, which implies $a^2 = 1$, that is $a = \pm 1$. Thus $z = \pm 1$. Conversely, ± 1 are units, so we get $R^* = \{\pm 1\}$.

Exercise: Show that $\mathbb{Z}[i]^* = \{\pm 1, \pm i\}, \mathbb{Z}[i\sqrt{2}]^* = \{\pm 1\}$ and $\mathbb{Z}[j]^* = \{\pm 1, \pm j, \pm(1 + j)\}$.

Notice that they are infinitely many units in $\mathbb{Z}[\sqrt{2}]$. Indeed, $\pm(1+\sqrt{2})^n$ is a unit for all $n \geq 0$.

Trick 2: To prove that some element is irreducible.

Suppose that you want to prove that $z \in R$ is irreducible. First you have to check that it is not a unit. Normally, you should have been asked to determine all the units, so you should have proved $z \in R^* \iff \delta(z) = 1$.

Now if $z = z_1.z_2$, then $\delta(z) = \delta(z_1).\delta(z_2)$, so $\delta(z_1)$ divides $\delta(z)$ in $\mathbb{N}$.

Therefore, to prove that we can proceed as follows:

- Find the list of ALL positive divisors m of $\delta(z)$.

- For all $m \neq 1, \delta(z)$, solve the equation $\delta(z_1) = m$ and check that no solution $z_1 \in R$ of this equation actually divides z.

- Then the only possibilities are $\delta(z_1) = 1$ or $\delta(z_1) = m$, which means $\delta(z_1) = 1$ or $\delta(z_2) = 1$. By the previous trick, if means that $z_1 \in R^*$ or $z_2 \in R^*$, so z is irreducible.

This is exactly the method we used to prove that 2 was irreducible in $\mathbb{Z}[i\sqrt{6}]$.

Notice that if $\delta(z)$ is a prime number, then z is irreducible (do you see why?).

Be careful, it could happen though that $\delta(z)$ is not prime, although z is irreducible in $\mathbb{Z}[i]$, e.g., 3 is irreducible in $\mathbb{Z}[i]$, but $\delta(3) = 9$. Also $i\sqrt{6}$ is irreducible in $\mathbb{Z}[i\sqrt{6}]$ but $\delta(z) = 6$.

Example: Let $R = \mathbb{Z}[i\sqrt{6}]$, and let's prove that $i\sqrt{6}$ is irreducible. This is not a unit, since it is different from ± 1 (recall that $R^* = \{\pm 1\}$ in this particular case).

Remember that we proved that $\delta(z) = 1 \Rightarrow z = \pm 1$.

Assume that $i\sqrt{6} = z_1.z_2$ for some $z_1, z_2 \in R$. We have $\delta(i\sqrt{6}) = 6 = \delta(z_1).\delta(z_2)$ in $\mathbb{N}$.

Since $\delta(z_1), \delta(z_2) \in \mathbb{N}$, it implies that $\delta(z_1)$ divides 6, so $\delta(z_1) = 1, 2, 3$ or 6.

Assume that $\delta(z_1) = 2$, and write $z_1 = a + bi\sqrt{6}$. We then have $a^2 + 6b^2 = 2$.

If $b \neq 0$, then $|b| \geq 1$, and we would get $a^2 + 6b^2 \geq 6b^2 \geq 6$. So $b = 0$, which implies $a^2 = 2$, which not possible since 3 is not a square in $\mathbb{Z}$.

Assume now that $\delta(z_1) = 3$, and write $z_1 = a + bi\sqrt{6}$. We then have $a^2 + 6b^2 = 3$.

Write $z = a + ib\sqrt{6}$. We then have $a^2 + 6b^2 = 3$. If $b \neq 0$, then $|b| \geq 1$, and we would get $a^2 + 6b^2 \geq 6b^2 \geq 6$. So $b = 0$, which implies $a^2 = 3$, which not possible since 2 is not a square in $\mathbb{Z}$ either.

If $\delta(z_1) = 1$, we have $z_1 = \pm 1$, so z_1 is a unit.

If $\delta(z_1) = 6$, we have $\delta(z_2) = 1$, so $z_2 = \pm 1$, and z_1 is a unit.

Hence we proved that if $i\sqrt{6} = z_1.z_2$, then z_1 or z_2 is a unit, so $i\sqrt{6}$ is irreducible.

Trick 3: To factor elements.

If $z = a_1.a_2 \cdots a_r$, where the a_i's are irreducible elements, then $\delta(z) = \delta(a_1).\delta(a_2) \cdots \delta(a_r)$, so $\delta(a_i)$ divides $\delta(z)$ in $\mathbb{N}$.

Notice that $\delta(a_i) \neq 1$ since a_i is not a unit.

Therefore, to factor z, we can proceed as follows:

Step 1: Find the set S of ALL positive divisors m of $\delta(z), m \neq 1, \delta(z)$.

Step 2: Pick $m \in S$.

Step 3: Solve the equation $\delta(u) = m$ for $u \in R$.

If there is no solution, remove m form the set S and go to Step 2.

If there are some solutions, check if some of them divides z in R. If not, remove m from S and go to Step 2.

If some $u \in R, \delta(u) = m$ divides z in R, then divide by u, replace z by the quotient. If this new value of z is a unit, then goto Step 4. Otherwise proceed to Step 1 with this new value of z.

Step 4: Once you have a decomposition $z = u_1 \cdots u_r$, repeat process for each u_i, to get a new decomposition of z. If in this new decomposition each factor is irreducible then stop. Otherwise, repeat the process with each new factor until this is the case.

Example: Let $R = \mathbb{Z}[i]$ and let $z = 22 + 19i$. Then $\delta(z) = 22^2 + 19^2 = 845 = 5 \times 13^2$. Thus if z has any irreducible factors, they must have norms $5, 13, 65$ or 169.

Let's solve $\delta(u) = 5$, $u = a + bi \in \mathbb{Z}[i]$. We have $a^2 + b^2 = 5$. If $|b| \geq 3$, then $b^2 \geq 9$ and so $a^2 + b^2 \geq 9$ as well.

So $|b| \leq 2$, and therefore $b = 0, \pm 1, \pm 2$.

If $b = 0$, we get $a^2 = 5$, which is not possible since $a \in \mathbb{Z}$.

If $b = \pm 1$, we get $a^2 = 4$, that is $a = \pm 2$.

if $b = \pm 2$, we get $a^2 = 1$, that is $a = \pm 1$.

Therefore $u = 2+i, 2-i, -1+2i, -1-2i$.

Let's figure out if $2+i$ divides $22+19i$ in $\mathbb{Z}[i]$. We have

$$\frac{22+19i}{2+i} = \frac{(22+19i)(2-i)}{5} = \frac{63+16i}{5},$$

which is not an element of $\mathbb{Z}[i]$, so $2+i$ does not divide $22+19i$.

Let's figure out if $2-i$ divides $22+19i$ in $\mathbb{Z}[i]$. We have

$$\frac{22+19i}{2-i} = \frac{(22+19i)(2+i)}{5} = 5+12i,$$

which is an element of $\mathbb{Z}[i]$, so $2+i$ does divide $22+19i$ and $22+9i = (2-i)(5+12i)$.

Now $\delta(5+12i) = 25+144 = 169 = 13^2$, so the possible irreducible divisors of $5+12i$ have norm 13.

Let's solve $\delta(u) = 13$, $u = a+bi \in \mathbb{Z}[i]$. We have $a^2+b^2 = 13$. If $|b| \geq 4$, then $b^2 \geq 16$ and so $a^2+b^2 \geq 16$ as well.

So $|b| \leq 3$, and therefore $b = 0, \pm 1, \pm 2, \pm 3$.

If $b = 0$, we get $a^2 = 3$, which is not possible since $a \in \mathbb{Z}$.

If $b = \pm 1$, we get $a^2 = 12$, which is not possible since $12 = 2^2.3$ is not a square.

If $b = \pm 2$, we get $a^2 = 9$, that is $a = \pm 3$.

If $b = \pm 3$, we get $a^2 = 4$, that is $a = \pm 2$.

Therefore $u = 2+3i, 2-3i, -2+3i, -2, -3i, 3+2i, 3-2i, -3+2i, -3-2i$.

One can check that $\dfrac{5+12i}{3+2i} = 3+2i$, so $5+12i = (3+2i)^2$.

We then get $z = (2-i)(3+2i)^2$.

Now since the norm of $2-i$ is 5, which is a prime number, then $2-i$ cannot be factored further. Since the norm of $3+2i$ is 13, which is a prime number, then $3+2i$ cannot be factored further.

So a decomposition of z into products of irreducible elements is given by $z = (2-i)(3+2i)^2$.

Remark: You really have to proceed to Step 4 ! Imagine that you started by investigating the equation $\delta(u) = 65$, and that you figured out that $u = 8+i$ has norm 65 and divides $22+19i$, so $22+19i = (3+2i)(8+i)$.

The process stops with the new quotient $3+2i$, because $3+2i$ cannot be factored, but the decomposition $z = (3+2i)(8+i)$ you obtained is NOT a decomposition into product of irreducible elements, since $8+i = (2-i)(3+2i)$ can be decomposed further.

You can avoid this situation by starting to deal with the elements m of S which are powers of a prime number.

Remark: If $\mathbb{Z}[\sqrt{d}]$ happens to be an ED, it is often Euclidean for its standard norm. If it happens to be the case, the arguments used to prove it are those used to prove that $\mathbb{Z}[\sqrt{2}]$ is an ED.

However, this is not always the case, since we saw that $\mathbb{Z}[i\sqrt{6}]$ is NOT a UFD, and therefore not an ED for any norm δ.

6. Polynomial Rings

In all this chapter, R will denote a commutative ring.

6.1. **Basic results.**

We recall the definition of the degree of a polynomial given in the first chapter.

Definition. Let $P \in R[X]$, $P \neq 0$. Since $P \neq 0$, we can write $P = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, with $a_i \in R$, $a_n \neq 0$. The integer $n \geq 0$ is called the *degree* of P, and is denoted by $\deg(P)$.

The coefficient a_n is called the *leading coefficient* of P.

We continue with a very important proposition, which generalizes the long division for polynomials that you use to know.

Theorem 6.1 (Generalized long division for polynomials)**.** *Let R be an integral domain. Let $P_1, P_2 \in R[X]$, $P_1 \neq 0, P_2 \neq 0$, $\deg(P_1) \geq \deg(P_2)$ and assume that the leading term of P_2 is a* **unit** *of R. Then there exists two polynomials $Q, S \in R[X]$, with $S = 0$ or $\deg(S) < \deg(P_2)$ such that $P_1 = Q.P_2 + S$. Moreover, Q and S are unique.*

Proof. We prove the existence of Q and S by induction on $\deg(P_1)$. More precisely, let (H_n) be the following property:

(H_n) : For all $P_1, P_2 \in R[X]$ non zero polynomials satisfying $\deg(P_2) \leq \deg(P_1) \leq n$, and such that the leading term of P_2 is a unit, there exist $Q, S \in R[X]$, such that $S = 0$ or $\deg(S) < \deg(P_2)$ such that $P_1 = P_2Q + S$.

If $\deg(P_1) = 0$, then $\deg(P_2) = 0$. Therefore it means that $P_1 = a, P_2 = b, a, b \in R, b \in R^*$. We can set $Q = b^{-1}a$ and $S = 0$ in this case. Thus (H_0) is true.

Now assume that (H_n) is true for some $n \geq 0$, and let's prove that (H_{n+1}) is true. If $\deg(P_1) \leq n$, then Q and S exist because (H_n) is true, so we can assume without loss of generality that $\deg(P_1) = n+1$.

Write $P_1 = a_{n+1}X^{n=1} + \cdots + a_1X + a_0, P_2 = b_mX^m + \cdots + b_1X + b_0$. Notice that by assumption we have $n + 1 \geq m$.

Set $Q_1 = b_m^{-1}a_mX^{n+1-m}$ and $S_1 := P_1 - b_m^{-1}a_mX^{n+1-m}P_2$. Notice that b_m^{-1} makes sense in R, because $b_m \in R^*$. Since $P_1, P_2 \in R[X]$, we also have $Q_1, S_1 \in R[X]$. Now by construction, the coefficient of X^{n+1} in S_1 is 0, so $\deg(S_1) \leq n$.

If $S_1 = 0$ or $\deg(S_1) < \deg(P_2)$, then we are done since we can write $P_1 = b_m^{-1}a_mX^{n+1-m}P_2 + S_1$, so we can set $S = S_1$ and $Q = b_m^{-1}a_mX^{n+1-m}$.

Now assume that $\deg(S_1) \geq \deg(P_2)$. Since $\deg(P_2) \leq \deg(S_1) \leq n$, we can apply (H_n), so there exist $Q_2, S_2 \in R[X]$ such that $S_1 = P_2Q_2 + S_2$, such that $S_2 = 0$ or $\deg(S_2) < \deg(P_2)$.

Hence we have $P_1 = (b_m^{-1}a_mX^{n+1-m} + Q_2)P_2 + S_2$, and $S_2 = 0$ or $\deg(S_2) < \deg(P_2)$. Now set $Q = b_m^{-1}a_mX^{n+1-m} + Q_2$ and $S = S_2$.

Therefore, (H_{n+1}) is true, and this concludes by induction. We now prove uniqueness of Q and S.

Assume that $P_1 = Q_1P_2 + S_1 = Q_2P_2 + S_2$, with $Q_i, S_i \in R[X]$ and $S_i = 0$ or $\deg(S_i) < \deg(P_2)$. Then we get $P_2(Q_1 - Q_2) = S_2 - S_1$.

If $S_2 - S_1 \neq 0$, then it implies that $Q_1 - Q_2 \neq 0$ either, hence we get $\deg(P_2(Q_1 - Q_2)) = \deg(P_2) + \deg(Q_1 - Q_2) \geq \deg(P_2)$ (the equality comes from the fact that R is an integral domain).

However, it is easy to see that $\deg(S_2 - S_1) < \deg(P_2)$, so we get a contradiction. Hence $S_1 = S_2$, so $P_2(Q_1 - Q_2) = 0$. Since R is an integral domain, so is $R[X]$, and since $P_2 \neq 0$, we get $Q_1 = Q_2$. □

In practice, the best way to find Q and S is to proceed exactly as in the case of polynomials with coefficients in a field.

WARNING: This is not true for general polynomials of $R[X]$. For example, $X, 2X \in \mathbb{Z}[X]$, but the quotient of X by $2X$ in $\frac{1}{2}$, which is not an element of $\mathbb{Z}[X]$.

Corollary 6.2. *If K is a field, then $K[X]$ is an Euclidean domain.*

Proof. Take δ to be the degree of a non-zero polynomial in $K[X]$. Since K is a field, any non-zero element is a unit, so we can proceed to long division of arbitrary non-zero polynomials by the previous result. □

Definition. Let R be a commutative ring, and let $P(X) \in R[X]$. We say that $\alpha \in R$ is a root of P if $P(\alpha) = 0$.

Proposition 6.3 (Remainder theorem)**.** *Let R be an integral domain, and let $P \in R[X]$. Then $\alpha \in R$ is a root of P if and only if $(X - \alpha)$ divides P.*

Proof. By the generalized long division algorithm, $P(X) = (X-\alpha)P_1 + P_2$, where either $P_2 = 0$ or P_2 has degree zero (in which case P_2 is a non-zero constant). So we can write $P_2 = r$, for some $r \in R$. Then $P(\alpha) = r$, so $P(\alpha) = 0$ if and only if $r = P_2 = 0$. □

Proposition 6.4. *Let R be an integral domain, and $P \in R[X]$ be a polynomial of degree $n \geq 0, P \neq 0$. Then P has at most n roots in R.*

Proof. We proceed by induction. If $n = 0$, then P is a nonzero constant polynomial, hence P has no roots. Now assume that we proved the result for all polynomials of degree $n \geq 0$, and let P be a polynomial of degree $n+1$. If P has no roots then we are done. Otherwise, there exists $\alpha \in R$ which is a root of P. But then $P(X) = (X-\alpha)Q$, by the previous result, for some $Q \in R[X]$. Notice that, since R is an integral doamin, we have

$$\deg(P) = \deg(X-\alpha) + \deg(Q),$$

that is $\deg(Q) = n$.

Now if $\beta \in R$ is any root of P, $0 = P(\beta) = (\beta-\alpha)Q(\beta)$, so either $\beta = \alpha$ or β is a root of Q, since R is an integral domain. By induction, Q has at most n roots, so P has at most $n+1$ roots and we are done. □

When we study fields, the irreducibility of polynomials will be important, so we need first to determine the units of $R[X]$.

Recall a result we proved earlier:

Lemma 6.5. *Let R be an integral domain. Then $R[X]^* = R^*$.*

Remark 6.6. If K is a field, then the notion of irreducibility in $K[X]$ coincides with the usual one.

Indeed, let $P \in K[X]$ be an irreducible element of $K[X]$ (in the general sense). Then P is irreducible $\iff$ $P \neq 0, P \notin K[X]^*$ and $P = P_1P_2 \in K[X] \Rightarrow P_1 \in K[X]^*$ or $P_2 \in K[X]^*$. But here $K[X]^* = K^* = K - \{0\}$, and in particular $P \neq 0$ and $P \notin K[X]^*$ is equivalent to say that $\deg(P) \geq 1$, so we recover the usual definition of an irreducible polynomial, that is $\deg(P) \geq 1$ and P can be written as a product of two non constant polynomials.

Example. If K is a field, then any polynomial of degree 1 is irreducible. Indeed, let $P \in K[X]$, $\deg(P) = 1$. Assume that $P = P_1P_2$. Then $\deg(P) = 1 = \deg(P_1) + \deg(P_2)$, so necessarily $\deg(P_1) = 0$ or $\deg(P_2) = 0$, which means that P_1 or P_2 is a non zero constant polynomial, that is a unit of $K[X]$.

The following theorem solves the question of which polynomials are irreducible over $\mathbb{C}$:

Theorem 6.7 (Fundamental theorem of algebra)**.** *Any non-constant polynomial in $\mathbb{C}[X]$ has a root in $\mathbb{C}$.*

We shall not prove this, although there is a proof in Allenby, section 4.8. One reason why we won't prove this is that despite of its title, it isn't a completely algebraic result. The proof given in Allenby uses no analytic facts except that a polynomial of odd degree in $\mathbb{R}[X]$ has a root in $\mathbb{R}$.

A field K such that every polynomial over $K[X]$ has a root in K, or equivalently, has all its roots in K, is called *algebraically closed.* Thus this theorem says that $\mathbb{C}$ is algebraically closed.

Proposition 6.8. *The only irreducible polynomials in $\mathbb{C}[X]$ are polynomials of degree one, i.e., up to multiplication by a unit they are polynomials of the form $X - \alpha$.*

Proof. Any polynomial in $\mathbb{C}[X]$ of degree greater than one is not irreducible. Also any polynomial of degree zero is a unit. □

Proposition 6.9. *The only irreducible polynomials in $\mathbb{R}[X]$ are polynomials of degree one, and polynomials $aX^2 + bX + c$ of degree two such that $4ac > b^2$.*

Proof. View $P(X) \in \mathbb{R}[X]$ as an element of $\mathbb{C}[X]$. Then $P = \lambda \prod_{i=1}^{n}(X - \alpha_i)$. Since the coefficients of P are real, $P(\overline{\alpha}_i) = \overline{P(\alpha_i)} = 0$, so the non-real roots occur in conjugate pairs. Now if α is not real, $(X - \alpha)(X - \overline{\alpha})$ is a quadratic polynomial with no real roots, so is as described above (with $a = 1$). Such a quadratic is irreducible in $\mathbb{R}[X]$ because factors would have degree one, so would lead to roots. □

The problem of determining irreducibility in $\mathbb{Q}[X]$ is harder.

6.2. **Factorisation in $R[X]$.** We have shown that $K[X]$ is a unique factorization domain for any field K. We'll look at the same problem for R an integral domain with field of fractions $F = K_R$. Clearly, if R is not a UFD, then $R[X]$ cannot be, because even the degree zero polynomials won't factor uniquely. This turns out to be the only restriction however.

In this section, R is a UFD and F denotes its field of fractions. Recall now that in a UFD, prime and irreducible elements coincide, so we will talk about prime elements of R and decomposition into prime factors.

Definition. A polynomial $f \in R[X]$ is said to be *primitive* if a h.c.f. of all the coefficients is a unit of R.

Examples.

- If $f \in R[X]$ is a **monic** polynomial, then f is primitive.
- If $f \in R[X]$ is primitive and $c \in R^*$, then cf is primitive.

We start with a little lemma:

Lemma 6.10. *Let $f \in F[X], f \neq 0$. Then there exists $c \in F - \{0\}$ and $f_1 \in R[X]$ primitive such that $f = c.f_1$ Moreover, c and f_1 are unique up to multiplication by a unit of R.*

Proof. Let $b \in R$ be a common denominator for all the coefficients of f. Then $f_2 = b.f \in R[X]$. Now let d be a h.c.f. of the coefficients of f_2. We then have $f_2 = d.f_1$, where f_1 satisfies the required conditions, and $f = (b.d).f_1$.

To prove the second part, assume that $f = c.f_1 = c'.f_1'$, where f_1, f_1' satisfy the two conditions. Write $c = \frac{a}{b}, c' = \frac{a'}{b'}$, so we have $ab'f_1 = a'bf_1'$. Set $u = ab', u' = ba'$, so $uf_1 = u'f_1'$. If u and u' are units, then $\frac{c}{c'} = \frac{u}{u'}$ is a unit, and we are done. So assume that u is not a unit for example. Let π be a prime element dividing u. Write $f_1' = \sum \alpha_i X^i, \alpha_i \in R$. By assumption on f_1', π does not divide some α_i. But we have $\pi | u$, so it divides all the coefficients of $uf_1 = u'f_1'$. In particular $\pi | u'\alpha_i$, so $\pi | u'$ since $\pi \not| \alpha_i$ and π is prime. Since this is true for all the prime elements dividing u, we get $u | u'$. But now u' is therefore not a unit, and we can do the same reasoning to prove that $u' | u$; therefore u and u' are associate and thus differ by a unit in R (since R is an integrla domain), so the same is true for c and c'. □

Lemma 6.11. *If $f \in R[X]$, then c is a h.c.f. of the coefficients of f. In particular $c \in R$.*

Proof. Let c' be a h.c.f. of the coefficients of f. Then $f = c'.g$, where $g \in R[X]$ is primitive. Now we also have $f = c.f_1$. By the uniqueness part of Lemma 6.10, c and c' differ by a unit of R. In particular, c is also a h.c.f. of the coefficients of f. □

Lemma 6.12 (Gauss Lemma)**.** *Let $f, g \in R[X]$. If f and g are primitive, then fg is primitive.*

Proof. We have to prove that a given prime element π does not divide all the coefficients of fg. Consider $\overline{R} := R/(\pi)$. Since π is a prime element, (π) is a prime ideal and thus $\overline{R}$ is an integral domain. Now by assumption $\overline{f}$ and $\overline{g}$ are non zero polynomials of $\overline{R}[X]$ (otherwise π would divide all the coefficients of f and g, and f and g would not be primitive). Therefore $\overline{f}\overline{g} = \overline{fg} \neq \overline{0}$, so π does not divide all the coefficients of fg. □

Lemma 6.13. *Let $f, g \in R[X]$ and $h \in F[X]$ such that $f = g.h., g, h, \in F[X]$. If f and g are primitive, then h is a primitive polynomial of $R[X]$.*

Proof. Write $h = ch_1$, where $c \in F^*$ and $h_1 \in R[X]$ is primitive, so we have $f = cgh_1$. By Gauss Lemma, gh_1 is primitive. By the uniqueness part of Lemma 6.10, it implies that $c \in R^*$ (since f is primitive). Hence $h = ch_1$ is a polynomial of $R[X]$ and is primitive as well. □

Before continuing, let's give an example which point out some difficulties concerning the relation between irreducibility in $R[X]$ and irreducibility in $F[X]$.

First of all, let's recall that, if R is an integral domain, then $R[X]^* = R^*$.

Now consider $P(X) = 2X - 2 \in \mathbb{Z}[X]$ for example. This polynomial is irreducible in $\mathbb{Q}[X]$, since it has degree 1. However, it is not irreducible in $\mathbb{Z}[X]$!!! Indeed, $P(X) = 2.(X-1)$. But $\mathbb{Z}[X]^* = \mathbb{Z}^* = \{\pm 1\}$, so 2 and $X - 1$ are not units, and therefore P is not irreducible.

More generally, if $P \in R[X]$ is not primitive, then P is not irreducible in $R[X]$, although it could be irreducible in $F[X]$. Indeed, by assumption every coefficient of P are divisible by a same irreducible element π, and therefore we obtain a non-trivial decomposition $P = \pi.Q, Q \in R[X]$.

The next result says that is essentially the only nasty thing which could happen:

Theorem 6.14 (Gauss theorem). *Let R be a UFD with field of fractions F. Then $R[X]$ is a UFD. Its irreducible elements are described as follows:*

1) *The prime elements of R*

2) *The primitive polynomials of $R[X]$ of degree ≥ 1 which are irreducible in $F[X]$.*

Proof. We first describe the irreducible elements of $R[X]$. Let $f \in R[X]$.

Assume $\deg(f) = 0$, then $f = a \in R$. Notice that if $f = g.h \in R[X]$, then $\deg(g) = \deg(h) = 0$, so $g, h \in R$. It is therefore immediate that f is irreducible in $R[X]$ if and only if it is irreducible in R.

Now assume that $\deg(f) \geq 1$. If f is not irreducible, then let $\pi \in R$ be a prime element dividing all the coefficients of f. Then $f = \pi.g$ for some $g \in R[X]$ of degree ≥ 1. Since π and g are not units of $R[X]$ (recall that we have $R[X]^* = R^*$, we obtain a non trivial decomposition of f and thus f is not irreducible. Hence if f is irreducible, it is necessarily primitive.

Now assume that f is primitive. We now show that f is irreducible in $R[X]$ if and only if it is irreducible in $F[X]$, which will conclude the proof of the second part of the theorem.

Assume first that f is irreducible in $R[X]$, and assume that f is not irreducible in $F[X]$, that is $f = g.h, g, h \in F[X], \deg(g) \geq 1, \deg(h) \geq 1$. Write $g = cg_1$, where $c \in F^*$ and $g_1 \in R[X]$ is primitive. We then have $f = g_1.(c.h)$. By Lemma 6.13 $h' := c.h \in R[X]$ (and is even primitive), so we obtain a non trivial decomposition in $R[X]$, which is a contradiction. Hence f is also irreducible in $F[X]$.

Finally, assume that f is not irreducible in $R[X]$, so $f = g.h$ for $g, h \in R[X], g, h \notin R[X] = R^*$. Assume first that $\deg(g) = 0$, so $g \in R$. Since $g \notin R^*$, then it is divisible by a prime element of R, and therefore $f = g.h$ is not primitive, which is a contradiction, so $\deg(g) \geq 1$.

Similarly, $\deg(h) \geq 1$. Hence the decomposition $f = g.h$ is also non-trivial in $F[X]$, so f is not irreducible in $F[X]$.

This concludes the proof of the second part of the theorem.

Now let's prove the existence of the decomposition.

Let $f \in R[X], f \neq 0$. Write $f = P_1 \cdots P_r$,where the P_i's are irreducible in $K[X]$. Write $P_i = c_i f_i$, where $f_i \in R[X]$ is primitive and $c_i \in F^*$. Then $f = c' f_1 \cdots f_r$, where $c' = c_1 \cdots c_r \in F^*$. Since all the f_i's are primitive, then $f_1 \cdots f_r$ is primitive as well by Gauss Lemma (applied several times). Write $f = cf_1$, where $c \in R$ is a h.c.f of the coefficients of f and f_1 is primitive. By the uniqueness part of Lemma 6.10, c and c' differs by a unit of R, and in particular $c' \in R$. Since we know that each f_i is an irreducible element of $R[X]$, and that irreducible elements of R are also irreducible in $R[X]$, it suffices to decompose c' in R to obtain a suitable decomposition in $R[X]$.

Now assume that $f = \pi_1 \cdots \pi_m f_1 \cdots f_r = \pi'_1 \cdots \pi'_n g_1 \cdots g_s$, where f_i, g_j are primitive and π_i, π'_j are prime elements in R. By uniqueness of the decomposition in $F[X]$, we get that $r = s$ and that each f_i is associate to some g_j. After renumbering, one can assume that $g_i = \lambda_i f_i$, $\lambda_i \in F^*$. Since $g_i, f_i \in R[X]$ are primitive, we have $\lambda \in R^*$, so f_i and g_i are associate in $R[X]$. Therefore, since R is an integral domain, we can simplify by each f_i on both sides, and we get that $\pi_1 \cdots \pi_m = u.\pi'_1 \cdots \pi'_n$ for some $u \in R^*$. We conclude using the uniqueness of a decomposition in R. □

Important remark: Gauss Theorem has an important consequence: if $P \in R[X]$ is a **primitive** polynomial, then P is irreducible in $R[X]$ if and only if it is irreducible in $F[X]$.

Corollary 6.15. *The ring $K[X_1, \ldots, X_n]$ is a UFD for any field K and n, and $\mathbb{Z}[X_1, \ldots, X_n]$ is a UFD.*

We would like to continue now by giving some irreducibility criterions. Let's start with some criterion which works on any field.

Proposition 6.16. *Let K be any field, and let $P \in K[X]$, $\deg(P) = 2$ or 3. Then P is irreducible in $K[X]$ if and only if P has no roots in K.*

Proof. If P has a root $\alpha \in K$, then $P(X) = (X - \alpha).Q(X), Q \in K[X]$. Since we have $\deg(Q) = \deg(P) - 1 \geq 1$, we obtain a non-trivial decomposition and P is therefore not irreducible. Now assume that P is not irreducible, so $P = P_1.P_2, P_i \in K[X]$, $\deg(P_i) \geq 1$.

Since $\deg(P) = 2$ or 3, it is easy to see that necessarily one of the polynomials P_i has degree 1, say P_1. Then $P_1(X) = aX + b, a \neq 0$, and $\alpha := -b.a^{-1} \in K$ is a root of P_1, hence a root of P. □

Warning: This is not true for polynomials of degree ≥ 4 or polynomials with coefficients in an arbitrary ring R. Indeed, $2X^2+2 = 2.(X^2+1)$ is not irreducible in $\mathbb{Z}[X]$ (it is not primitive) but has no root in $\mathbb{Z}$. Moreover $X^4+2X^2+1 \in \mathbb{R}[X]$ has no root in $\mathbb{R}$, but is not irreducible, since $X^4+2X^2+1 = (X^2+1)^2 = (X^2+1)(X^2+1)$.

Now let's give a trick to decide whether or not a given polynomial $P \in \mathbb{Q}[X]$ has a root in $\mathbb{Q}$.

We can always write $P = r.Q$, where $r \in \mathbb{Q}$ and $Q \in \mathbb{Z}[X]$, and P has a root if and only if Q as a root. So we can assume without loss of generality that $P \in \mathbb{Z}[X]$

Write $P = a_nX^n+\cdots+a_1X+a_0, a_i \in \mathbb{Z}$. Assume that $r = \dfrac{p}{q}, h.c.f(p,q) = 1$ is a root of P.

Then we have

$$a_n\frac{p^n}{q^n} + \cdots + a_1\frac{p}{q} + a_0 = 0,$$

so

$$a_np^n + a_{n-1}p^{n-1}q + \cdots + a_1pq^{n-1} + a_0q^n = 0.$$

Therefore,

$$a_np^n = -(a_{n-1}p^{n-1}q + \cdots + a_1pq^{n-1} + a_0q^n).$$

Since the right -hand side is divisible by q is $\mathbb{Z}$, so is a_np^n. But $h.c.f(p,q) = 1$, so $q|a_n$.

A similar proof shows that $p|a_0$.

Hence, if $\dfrac{p}{q}, h.c.f.(p,q) = 1$ is a root of $P = a_nX^n+\cdots+a_1X+a_0, a_i \in \mathbb{Z}$, then $q|a_n$ and $p|a_0$.

Therefore, we only have finitely many possibilities to try.

We now come to two powerful irreducibility criterions:

Theorem 6.17 (Eisenstein's irreducibility criterion)**.** *Let R be a UFD with quotient field F, and let $f \in R[X]$ be a* **primitive** *polynomial, $f = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1x + a_0$. Assume that there is a prime element $\pi \in R$ such that*

1) $\pi \not| a_n$

2) $\pi | a_i$ *for all* $i = 0, \cdots, n-1$

3) $\pi^2 \not| a_0$.

Then f is irreducible in $R[X]$ and $F[X]$.

Proof. Since f is primitive, f is irreducible in $F[X]$ if and only if f is irreducible in $R[X]$. Clearly f is not zero nor a unit. Suppose that $f = gh$ in $R[X]$, with $g, h \notin R[X]^* = R^*$.

We have $\overline{f} = \overline{gh}$ in $\overline{R}[X]$, where

$$g(X) = b_m X^m + b_{m-1} X^{m-1} + \cdots + b_1 x + b_0,$$

$$h(X) = c_{n-m} X^{n-m} + \cdots + c_1 x + c_0.$$

Let $\overline{R} = R/(\pi)$.

We have $\overline{f} = \overline{gh}$ in $\overline{R}[X]$. Since $\overline{f} = \overline{a_n} X^n$, with $\overline{a_n} \neq \overline{0}$, we have $\overline{a}_n = \overline{b}_m \overline{c}_{n-m} \neq \overline{0}$.

Therefore $\deg(\overline{g}) = m, \deg(\overline{h}) = n - m$.

Moreover $m \neq 0$ and $n \neq m$. Indeed, if $m = 0$ for example, then $g = b_0 \in R - 0$. But since $f = g.h$ and f is primitive, it would imply that $g = b_0 \in R^*$, which is a contradiction. Similarly $m \neq n$.

Therefore $X^n = \overline{g}.\overline{h}$, with $1 \leq \deg(\overline{g}), \deg(\overline{h}) \leq n - 1$ in $\overline{R}[X]$. Comparing the constant terms, we get $\overline{b}_0 \overline{c}_0 = \overline{0}$, so $\overline{b}_0 = \overline{0}$ for example since $\overline{R}$ is an integral domain, that is $X | \overline{g}$ in $\overline{R}[X]$. Hence we can write $\overline{g} = X^k \overline{g'}$ for some $1 \leq k \leq m \leq n-1$ and $\overline{g'}$ with a non zero constant coefficient $d_0 \in \overline{R}$. So we get $X^n = X^k \overline{g'}\overline{h}$. Comparing the coefficient of X^k, we get $\overline{d}_0 \overline{c}_0 = \overline{0}$, so $\overline{c}_0 = \overline{0}$, and $\pi | c_0$. Therefore π^2 divides $b_0.c_0 = a_0$, a contradiction. □

This criterion is really useful when $R = \mathbb{Z}$, and $\pi = p$ is a prime number. For example, it follows immediately that $X^n - 2 \in \mathbb{Z}[X]$ is irreducible in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$ for all $n \geq 1$. In particular, there exist irreducible polynomials of $\mathbb{Q}[X]$ of arbitrary large degree.

We end this section by another irreducibility criterion.

Theorem 6.18 (Reduction irreducibility criterion)**.** *Let R be a UFD with quotient field F, and let $f = a_n X^n + \cdots + a_1 X + a_0 \in R[X]$ be a* **primitive** *polynomial.*

Assume that there exists a prime element $\pi \in R$ such that:

1) $\pi \not| a_n$.

2) *The reduction $\overline{f}$ of f modulo π is irreducible in $\overline{R}[X]$, where $\overline{R} = R/(\pi)$.*

Then f is irreducible in $R[X]$ and $F[X]$.

Proof. Since f is primitive, f is irreducible in $F[X]$ if and only if f is irreducible in $R[X]$. Let's prove that f is irreducible in $R[X]$. Clearly f is not zero nor a unit. Assume that $f = g.h, g, h \in R[X]$. We have $\overline{f} = \overline{g}.\overline{h}$. Since $\pi \not| a_n$, we have $\deg(\overline{g}) = \deg(g)$ and $\deg(\overline{h}) = \deg(h)$, as in the previous proof. Now $\overline{f}$ is irreducible, hence $\overline{g}$ or $\overline{h}$ is a unit of $\overline{R}[X]$, that is an element of $\overline{R}^*$, since $\overline{R}$ is an integral domain. Assume for example that $\overline{g} \in \overline{R}^*$, so $\deg(\overline{g}) = 0 = \deg(g)$. Then $g = a \in R$.

Since $f = g.h$ and f is primitive, then necessarily $g = a \in R^*$. Hence f is irreducible in $R[X]$. □

Once again, this is very useful in the case $R = \mathbb{Z}$.

Example:

$f = X^3 - 2398563495689866X - 2876387683712763767468761 \in \mathbb{Q}[X]$

is irreducible.

We could of course try to see if such polynomial as a root. To do so, we will have factor the last coefficient, list all the set of divisors, and try all the possibilities. Without a computer or a calculator, it could be quite long and painful.

Let's try to apply the previous criterion. First of all, this polynomial is primitive (it is monic !!!), so it is irreducible in $\mathbb{Q}[X]$ if and only if it is irreducible in $\mathbb{Z}[X]$.

Now let's reduce modulo 3. We have $\overline{f} = X^3 - X - 1 \in \mathbb{F}_3[X]$. One easily check that this polynomial has no root in $\mathbb{F}_3$, and since it has degree 3, it is sufficient to prove it is irreducible in $\mathbb{F}_3[X]$. Now apply the previous criterion.

We finish with a criterion which may be useful sometimes, and which works over any commutative ring R.

Proposition 6.19. *Let R be a commutative ring and let $P \in R[X]$. For all $a \in R$, we have*

$$P(X) \textit{ is irreducible in } R[X] \iff P(X-a) \textit{ is irreducible in } R[X]$$

Proof. See exercise sheets. □

Example: let $f = X^4 + 4X^3 + 6X^2 + 4 - 1 \in \mathbb{Z}[X]$. We have

$$f(X-1) = X^4 - 2 \in \mathbb{Z}[X],$$

which is irreducible in $\mathbb{Z}[X]$, by Einsenstein's criterion, so f is irreducible as well.

7. Digression: Things you should know about vector spaces

Let F be a field. A *vector space* over F is a set V which is an abelian group with group composition written as $+$, identity element $\mathbf{0}$ and inverse operation written as $-$ (think of the group operation as addition of vectors) and is also equipped with a function from $F \times V$ to V (think of multiplication by scalars) such that for all $\lambda, \mu \in F$ and $\mathbf{v}, \mathbf{w} \in V$,

$$\lambda(\mathbf{v}+\mathbf{w}) = \lambda\mathbf{v}+\lambda\mathbf{w}, (\lambda+\mu)\mathbf{v} = \lambda\mathbf{v}+\mu\mathbf{v}, \lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}, 1\mathbf{v} = \mathbf{v}.$$

Recall that saying that V is an abelian group as above means that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$:

$$\mathbf{0}+\mathbf{v} = \mathbf{v}, \mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}, \mathbf{v}+(-\mathbf{v}) = \mathbf{0}, \mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}.$$

A finite list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of vectors is *independent* if for all $\lambda_1, \ldots, \lambda_n \in F$,

$$\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \mathbf{0} \text{ implies that for all } i, \lambda_i = 0.$$

Note that a list containing the zero vector can never be independent, nor can a list containing the same vector twice. An infinite list of vectors is independent if each finite sublist is independent. A set of vectors *spans* V if every vector in V can be written as a finite sum of scalar multiples of the vectors in the set. A *basis* for V is a set of vectors that spans V and is independent. It may be shown (using Zorn's lemma) that any vector space has a basis, and that any two bases for the same vector space contain the same number of elements. The *dimension* of a vector space is the size of a basis for it. We denote this by $\dim_F V$, or $\dim V$ if there is no ambiguity.

A *subspace*, W, of V is a subset such that for all $\lambda \in F$ and $\mathbf{v}, \mathbf{w} \in W$, we have $\lambda\mathbf{v} \in W$ and $\mathbf{v}+\mathbf{w} \in W$. This is intrinsically a vector space, and $\dim W \leq \dim V$.

Examples: 1. The space F^n of row vectors consisting of ordered n-tuples $(\lambda_1, \ldots, \lambda_n)$ is a vector space of dimension n. The unit vectors $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ form a basis.

In the first year, you have studied vector subspaces of the vector space $\mathbb{R}^n$, and proved that any vector subspace of $\mathbb{R}^n$ has a basis, and that the numbers of elements in any two bases of the same subspace are equal. So hopefully the step to abstract vector spaces (as opposed to subspaces of F^n) won't be too daunting.

Indeed, if V has finite dimension, n, then V is is isomorphic *as a vector space* to F^n.

2. The polynomial ring $F[X]$ is a vector space over F, with basis the monomials $1, X, X^2, X^3, \ldots$.

3. If K is a subfield of L, we may view L as a vector space over K. For example, $\mathbb{C}$ is a 2-dimensional vector space over $\mathbb{R}$, since 1, i form a basis. As another example, $\mathbb{Q}[i]$ is a 2-dimensional vector space over $\mathbb{Q}$, again with basis 1, i.

8. Field extensions

8.1. Basic definitions.

Definition. Let F be a field. If $K \subseteq F$ is a subfield of F, we call F a *field extension of* K, and we denote it by F/K. Then F can be viewed as a vector space over K, and the *degree of* F *over* K is defined to be the dimension of F as a K-vector space. Write $[F : K]$ for the degree of F over K. We say that F/K is *finite* if $[F : K]$ is finite, and *infinite* otherwise.

Examples.

- $\mathbb{C}\,/\,\mathbb{R}$ is a finite field extension of degree 2.
- $\mathbb{R}\,/\,\mathbb{Q}$ is an infinite field extension.
- $\mathbb{C}(X)/\,\mathbb{C}$ is an infinite field extension.

Lemma 8.1 (Tower degree formula). *Let L/F and F/K two field extensions (i.e. $K \subset F \subset L$). Then L/K is finite if and only if L/F and F/K are finite, and in this case, we have:*

$$[L : K] = [L : F].[F : K].$$

Proof. Let $(e_i)_{i\in I}$ be a F-basis of L and let $(f_j)_{j\in J}$ be a K-basis of F. We proceed to show that $(e_i.f_j)_{(i,j)\in I\times J}$ is a K-basis of L, which will show everything at once.

First we show that this family spans L as a K-vector space.

Let $x \in L$. Since $(e_i)_{i\in I}$ is a F-basis of L, we have $x = \displaystyle\sum_i \lambda_i e_i$ for some $\lambda_i \in F$. Now each λ_i may be expressed as $\lambda_i = \sum_j \mu_{ij} f_j$ for some $\mu_{ij} \in K$, and so

$$x = \sum_{i,j} \mu_{ij} e_i f_j.$$

Now we have to show that for any finite subset S of $I \times J$, the elements $(e_i.f_j)_{(i,j)\in S}$ are linearly independent over K. Adding some elements if necessary, one can always assume that $S = I' \times J'$, where $I' \subset I$ and $J' \subset J$ are finite. Suppose that $\displaystyle\sum_{(i,j)\in I'\times J'} \mu_{ij} e_i f_j = 0$ for some $\mu_{ij} \in K$. Letting $\lambda_i \in F$ be defined by $\lambda_i = \displaystyle\sum_{j\in J'} \mu_{ij} f_j$, we obtain $0 = \displaystyle\sum_{i\in I'} \lambda_i e_i$. Now by linear independence of the $e_i, i \in I'$, we deduce that for each i, $\lambda_i = 0$. Then by linear independence of the $f_j, j \in J'$, we deduce from the equation $0 = \lambda_i = \displaystyle\sum_{j\in J'} \mu_{ij} f_j$ that $\mu_{ij} = 0$ for all i and j. □

Definition. Let F/K be a field extension, and let $\alpha_1, \cdots, \alpha_n \in F$. We denote by $K(\alpha_1, \cdots, \alpha_n)$ the smallest subfield of F containing K and $\alpha_1, \cdots, \alpha_n$, and call it *the subextension of F/K generated by $\alpha_1, \cdots, \alpha_n$.*

We say that F/K is a *simple extension* of K if there exists $\alpha \in K$ such that $F = K(\alpha)$.

Lemma 8.2. *Let F/K be a field extension, let $n, m \geq 0$ be two non-negative integers and let $\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m \in F$. Then we have*

$$K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m) = K(\alpha_1, \cdots, \alpha_n)(\beta_1, \cdots, \beta_m)$$

Proof. By definition, $K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m)$ is a subfield of F containing K and the α_i's and β_j's . In particular, it contains K and the α_i's. Since $K(\alpha_1, \cdots, \alpha_n)$ is the smallest subfield of F satisfying these properties, we have

$$K(\alpha_1, \cdots, \alpha_n) \subseteq K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m)$$

Since $K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m)$ contains also the coordinates of the points of $\mathcal{A}'$, we get for the same kind of reason

$$K(\alpha_1, \cdots, \alpha_n)(\beta_1, \cdots, \beta_m) \subseteq K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m)$$

Moreover $K(\alpha_1, \cdots, \alpha_n)(\beta_1, \cdots, \beta_m)$ is a subfield of F which contains $K(\alpha_1, \cdots, \alpha_n)$ and the β_j's. Hence it contains K, the α_i's and the β_j's. Since $K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m)$ is the smallest subfield of F with these properties, we get

$$K(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m) \subseteq K(\alpha_1, \cdots, \alpha_n)(\beta_1, \cdots, \beta_m)$$

□

In fact, one can give a precise description of $K(\alpha_1, \cdots, \alpha_n)$:

Exercise: Show that $K(\alpha_1, \cdots, \alpha_n)$ is equal to

$$\{\frac{P(\alpha_1, \cdots, \alpha_n)}{Q(\alpha_1, \cdots, \alpha_n)}, P, Q \in K[X_1, \cdots, X_n], Q(\alpha_1, \cdots, \alpha_n) \neq 0\}$$

We end this section by studying field extensions of small degree. We start with a simple remark.

Remark 8.3. Let F/K be a field extension and $\alpha \in F$. Then we have

$$[K(\alpha) : K] = 1 \iff K(\alpha) = K \iff \alpha \in K$$

Indeed, if $[K(\alpha) : K] = 1$, then $K(\alpha)$ is a K-vector space of dimension 1 over K, so $K(\alpha) = K$. Hence $\alpha \in K$ since $K(\alpha)$ contains α. Now if

$\alpha \in K$, then K is a subfield of F containing K and α, so $K(\alpha) \subset K$. Now by definition $K \subset K(\alpha)$ so we are done.

Notation: Let F/K be a field extension. If $d \in K$, we denote by $\sqrt{d}$ any element $\alpha \in F$ satisfying $\alpha^2 = d$. Notice that α is only determined up to a sign.

Lemma 8.4. *Let F/K be a field extension of K. If $d \in K$, then*

$$K(\sqrt{d}) = \{a + b\sqrt{d}, a, b \in K\}$$

Moreover, $[K(\sqrt{d}) : K] = 1$ if d is a square in K, and $[K(\sqrt{d}) : K] = 2$ otherwise.

Proof. If d is a square, this is obvious, since both sets are equal to K, since $\sqrt{d} \in K$ is this case. So we can assume that d is not a square in K.

Since $K(\sqrt{d})$ contains K and $\sqrt{d}$ and is closed under addition and multiplication, we get $K(\sqrt{d}) \supset \{a + b\sqrt{d}, a, b \in K\}$. To prove the reverse inclusion, we have to check that $\{a+b\sqrt{d}, a, b \in K\}$ is a subfield of F containing K and $\sqrt{d}$. The proof works exactly as for the case $K = \mathbb{Q}$ (see solutions of exercise sheet 1 for a proof in this case), so we are done.

From the previous point, 1 and $\sqrt{d}$ spans $K(\sqrt{d})$ as a K-vector space, so $[K(\sqrt{d}) : K] \leq 2$. By a previous remark, $[K(\sqrt{d}) : K] = 1 \iff \sqrt{d} \in K$, which means that d is a square in K. Hence, $[K(\sqrt{d}) : K] \geq 2$ since d is not a square in K, so we are done. □

We are now ready to prove the structure theorem of field extensions of degree 2.

Theorem 8.5. *Let F/K be a field extension. Assume that* $\mathrm{char}(K) \neq 2$. *Then $[F : K] = 2 \iff F = K(\sqrt{d})$ for some $d \in K$, such that d is not a square in K.*

Proof. Let $\beta \in F, \beta \notin K$ (such a β exists since the assumption implies that $F \neq K$). The two elements 1 and β are easily seen to be linearly independent over K, since $\beta \notin K$. Since F has dimension 2 over K, we get that $1, \beta$ is a K-basis of F. Hence $F = \{x + y\beta, x, y \in K\}$.

Since $K(\beta)$ contains K and β, and is closed under addition and multiplication, we get that $F \subseteq K(\beta)$. Now by definition $K(\beta)$ is a subfield of F, so $K(\beta) \subseteq F$, hence $F = K(\beta)$.

Since F has dimension 2 over K, the elements $1, \beta, \beta^2$ are necessarily dependent. Hence there exists $a, b, c \in K$ not all zero such that

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

We necessarily have $a \neq 0$, otherwise we would have $\beta = \frac{-c}{b} \in K$. Hence $\beta = \frac{-b\pm\sqrt{d}}{2a}$, where $d = b^2 - 4ac$ (here we use the fact that $\mathrm{char}(K) \neq 2$, since we divide by 2). Notice that d is not a square since $\beta \notin K$.

Since $F = \{x + y\beta, x, y \in K\}$, easy computations show that $F = \{u + v\sqrt{d}, u, v \in K\} = K(\sqrt{d})$. □

Remark 8.6. If $\mathrm{char}(K) = 2$, the result of 2) is not true anymore.

Side remark: Let F/K be a field extension, and let $\alpha, \beta \in F$. It is useful sometimes to be able to compare $K(\alpha)$ and $K(\beta)$. In fact we have

$$K(\alpha) = K(\beta) \iff \alpha \in K(\beta) \text{ and } \beta \in K(\alpha)$$

If $K(\alpha) = K(\beta)$, then $\alpha \in K(\beta)$, since $K(\alpha)$ contains α. Similarly $\beta \in K(\alpha)$. Conversely, assume that $\alpha \in K(\beta)$ and $\beta \in K(\alpha)$. In this case, $K(\beta)$ is a subfield of F containing K and α. Since $K(\alpha)$ is the smallest subfield of F with these properties, we get $K(\alpha) \subseteq K(\beta)$. The other inclusion is proved similarly using the fact that$\beta \in K(\alpha)$.

Example. We have

$$\mathbb{Q}(\frac{3+\sqrt{242}}{8}) = \mathbb{Q}(\sqrt{2})$$

Indeed $\dfrac{3+\sqrt{242}}{8} = \dfrac{3+11\sqrt{2}}{8} = \dfrac{3}{8} + \dfrac{11}{8}\sqrt{2}$. Since $\mathbb{Q}(\sqrt{2})$ contains $\mathbb{Q}$ and $\sqrt{2}$, and is closed under addition and multiplication, it follows that$\frac{3+\sqrt{242}}{8} \in \mathbb{Q}(\sqrt{2})$. Hence we get

$$\mathbb{Q}(\frac{3+\sqrt{242}}{8}) \subseteq \mathbb{Q}(\sqrt{2})$$

Now we have $\sqrt{2} = -\dfrac{3}{11} + \dfrac{8}{11} \cdot \dfrac{3+\sqrt{242}}{8}$. Since $\mathbb{Q}(\frac{3+\sqrt{242}}{8})$ contains $\mathbb{Q}$ and $\frac{3+\sqrt{242}}{8}$, the previous equality show that $\sqrt{2} \in \mathbb{Q}(\frac{3+\sqrt{242}}{8})$, hence we get

$$\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\frac{3+\sqrt{242}}{8})$$

8.2. **Algebraic elements and minimal polynomial.**

Definition. If F/K is a field extension and $\alpha \in F$, we say that α is *algebraic* over K if there is a **non-zero** polynomial $f \in K[X]$ such that $f(\alpha) = 0$. We say that α is *transcendental* over K otherwise.

Examples.

- Any $\alpha \in K$ is algebraic over K, since it is a root of $f(X) = X - \alpha$.

- $i \in \mathbb{C}$ is algebraic over $\mathbb{Q}$, since it is a root of $X^2 + 1 \in \mathbb{Q}[X]$.

- Any $\alpha \in \mathbb{C}$ is algebraic over $\mathbb{R}$, as it is a root of
$f(X) = (X - \alpha)(X - \bar{\alpha}) = X^2 - (\alpha + \bar{\alpha})X + \alpha\bar{\alpha} \in \mathbb{R}[X]$.

- The real numbers e and π are transcendental over $\mathbb{Q}$ (difficult!).

Proposition 8.7. *Let F/K be a field extension, and let $\alpha \in F$. The set*

$$I_\alpha := \{P \in K[X] | P(\alpha) = 0\}$$

is an ideal of K[X]. It is a non-zero ideal if and only if α is algebraic over K. In this case, there exists a unique **monic irreducible** *polynomial $\mu_{\alpha,K}$ such that*

$$I_\alpha = (\mu_\alpha)$$

Proof. The fact that I_α is an ideal comes from the fact that it is the kernel of the ring homomorphism

$$h_\alpha : P \in K[X] \mapsto P(\alpha) \in F$$

or by a direct proof. Now by definition, I_α is not zero if and only if α is algebraic. Since $K[X]$ is a PID, then $I = (P_0)$ for some $P_0 \in K[X]$. Assume that $I_\alpha \neq (0)$, so $P_0 \neq 0$. Notice also that P_0 is not a constant polynomial, since α is a root of P_0. Since non-zero elements of K are units, we have $(P_0) = (cP_0)$ for all $c \in K^*$, so we can assume that P_0 is monic after multiplying it by a suitable non-zero element of K.

Let us prove that P_0 is irreducible. Since $P_0 \neq 0$ is not constant, it is not a unit. Now assume that $P_0 = P_1P_2, P_i \in K[X]$. Then $P_0(\alpha) = 0 = P_1(\alpha)P_2(\alpha)$. Hence $P_1(\alpha) = 0$ or $P_2(\alpha) = 0$. Assume that $P_1(\alpha) = 0$ for example, the second case being similar. Then $P_1 \in I_\alpha = (P_0)$, so $P_0|P_1$. Since $P_1|P_0$ as well, we get that $P_0 = cP_1$ for some $c \in K, c \neq 0$. Hence we get that $P_2 = c$, which is a unit.

Now if $I_\alpha = (P_0) = (Q_0)$ with $Q_0 \in K[X]$ monic, then $P_0|Q_0$ and $Q_0|P_0$, so $Q_0 = cP_0$ for some $c \in K$. But $c = 1$ since both polynomials are monic. □

Definition. The polynomial $\mu_{\alpha,K}$ is called the *minimal polynomial of α over K.*

Practical remark:

- From the definition of the minimal polynomial, it follows that if $P \in K[X]$ satisfies $P(\alpha) = 0$, then $\mu_{\alpha,K}|P$; if moreover P is monic and irreducible then $P = \mu_{\alpha,K}$.

Therefore, to compute the minimal polynomial of a given $\alpha \in F$, one may proceed as follows:

1) Find a monic polynomial $P \in K[X]$ satisfying $P(\alpha) = 0$.

2) If P is irreducible, then $P = \mu_{\alpha,K}$ and we are done. If P is not irreducible, then decompose P as a product of monic irreducible factors. Then $\mu_{\alpha,K}$ will be the unique monic irreducible factor of P for which α is a root.

Examples.

- Let us compute $\mu_{i,\mathbb{R}}$. We know that $i^2 = -1$, so $P := X^2 + 1 \in \mathbb{R}[X]$ is a monic polynomial of $\mathbb{R}[X]$ satisfying $P(i) = 0$. Now $X^2 + 1$ is irreducible in $\mathbb{R}[X]$, since it has degree 2 and has no roots in $\mathbb{R}$. Hence $\mu_{i,\mathbb{R}} = X^2 + 1$.

- Let us compute $\mu_{\alpha,\mathbb{Q}}$, where $\alpha := \dfrac{1 + i\sqrt{19}}{2}$. We would like to find a polynomial $P \in \mathbb{Q}[X]$ such that $P(\alpha) = 0$. If we take a look to the definition of α, the only thing which is not in $\mathbb{Q}$ is $i\sqrt{19}$, so we should get rid of it, the best way being squaring it. So we write $2\alpha - 1 = i\sqrt{19}$, and thus we have $(2\alpha - 1)^2 = -19$. Hene we get $4\alpha^2 - 4\alpha + 20 = 0$. Since we want P to be monic, we divide everything by 4, and we get $P(\alpha) = 0$, for $P = X^2 - X + 5$. This polynomial P is irreducible (check it; there is a lot of ways to proceed), hence $\mu_{\alpha,\mathbb{Q}} = X^2 - X + 5$.

- Let us compute $\mu_{j,\mathbb{Q}}$. We know that $j^3 = 1$, so $P := X^3 - 1 \in \mathbb{Q}[X]$ satisfies $P(j) = 0$ and is monic. However, P is not irreducible, since $P(1) = 0$, so P has a root. So we have to factor P into a product of monic irreducible polynomials in $\mathbb{Q}[X]$. We get easily $P = (X - 1)(X^2 + X + 1)$, and each factor is monic and irreducible (check it), so $\mu_{j,\mathbb{Q}}$ should be one of them. We have $j - 1 \neq 0$, so $\mu_{j,\mathbb{Q}} \neq X - 1$, since j must be a root of its minimal polynomial. Hence $\mu_{j,\mathbb{Q}} = X^2 + X + 1$, which is consistent with the fact that we know the relation $j^2 + j + 1 = 0$. One could have also use this relation to compute $\mu_{j,\mathbb{Q}}$ (it would have been quicker).

Theorem 8.8. *Let F/K be a field extension. Then $\alpha \in F$ is algebraic over K if and only if $K(\alpha)/K$ is has finite degree.*

In this case, a K-basis of $K(\alpha)$ is given by $1, \alpha, \cdots, \alpha^{d-1}$, where $d = \deg(\mu_{\alpha,K})$. In particular, we have the equality

$$[K(\alpha) : K] = \deg(\mu_{\alpha,K})$$

Proof. Assume first that α is transcendental over K. Then for all $n \geq 1$, the elements $1, \alpha, \cdots, \alpha^{n-1}, i \geq 0$ are linearly independent over K. Indeed, assume that $a_0, \cdots, a_{n-1} \in K$ are satisfying

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0.$$

Then the polynomial $P = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} \in K[X]$ satisfies $P(\alpha) = 0$, so $P = 0$ by assumption. Hence $a_0 = a_1 = \cdots = a_{n-1} = 0$.

Hence the dimension of $K(\alpha)$ over K is necessarily infinite. Otherwise, the $\dim_K K(\alpha) + 1$ elements $1, \alpha, \cdots, \alpha^{\dim_K K(\alpha)}$ would be linearly dependent over K.

Now assume that α is algebraic over K, and let d be the degree of its minimal polynomial over K.

We first proceed to show that $K(\alpha) = \{P(\alpha), P \in K[X]\}$. Since α and K are contained in $K(\alpha)$, and since $K(\alpha)$ is closed under multiplication and addition, it follows that $K(\alpha) \supset \{P(\alpha), P \in K[X]\}$.

Now to prove the other inclusion, it is enough to show that $\{P(\alpha), P \in K[X]\}$ is a subfield of F containg K and α. The only non obvious thing to prove is that if $P(\alpha) \neq 0$, then $P(\alpha)^{-1} = U(\alpha)$ for some $U \in K[X]$.

Since $P(\alpha) \neq 0$, P is not divisible by $\mu_{\alpha,K}$. Since $\mu_{\alpha,K}$ is irreducible, it implies that $h.c.f(P, \alpha_{\mu,K}) = 1$, so there exist $U, V \in K[X]$ such that $U(X)P(X) + V(X)\mu_{\alpha,K}(X) = 1$. We then have $U(\alpha)P(\alpha) = 1$ since $\mu_{K,\alpha}(\alpha) = 0$. This proves that $P(\alpha)^{-1} = U(\alpha)$ for some $U \in K[X]$.

We are now able to finish the proof of the theorem. For $P \in K[X]$, write

$$P = \mu_{\alpha,K}Q_1 + Q_2, Q_2 = 0 \text{ or } \deg(Q_2) < d$$

We then get $P(\alpha) = Q_2(\alpha)$. Since $Q_2 = 0$ or $\deg(Q_2) < d$, it follows that

$$P(\alpha) = a_{d-1}\alpha^{d-1} + \cdots + a_1\alpha + a_0$$

for some $a_i \in K$. Using the previous point, it shows that

$$K(\alpha) = \{a_{d-1}\alpha^{d-1} + \cdots + a_1\alpha + a_0, a_i \in K\}$$

It just remains to show that $1, \alpha, \cdots, \alpha^{d-1}$ are linearly independent over K. Assume that $a_0, \cdots, a_{d-1} \in K$ are satisfying

$$a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1} = 0.$$

Then the polynomial $P = a_0 + a_1X + \cdots + a_{d-1}X^{d-1} \in K[X]$ satisfies $P(\alpha) = 0$, so $P \in I_\alpha$ by assumption. Hence $\mu_{\alpha,K}$ divides P. If $P \neq 0$, it implies that $\deg(P) \geq d$, which is a contradiction. Hence $P = 0$ and so $a_0 = a_1 = \cdots = a_{d-1} = 0$. □

Example. Let $\alpha = \sqrt[3]{2} \in \mathbb{C}$. Let us compute $[\mathbb{Q}(\alpha) : \mathbb{Q}]$.

We have to find the minimal polynomial of α over $\mathbb{Q}$. We have $\alpha^3 - 2 = 0$, so $P(X) = X^3 - 2 \in \mathbb{Q}[X]$ satisfies $P(\alpha) = 0$. It is monic, and irreducible by Einsenstein criterion. Hence $\mu_{\alpha,\mathbb{Q}} = X^3 - 2$, and $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$.

9. Ruler and compass constructions

We are going to apply the the formalism of field extensions to solve geometric problems going back to the Antiquity.

Among other problems, the Greeks asked the following geometric problems:

- Is the trissection of a given angle θ is always possible just using ruler and compass?

- Can we construct a square with same area as the unit circle using ruler and compass?

- Can we construct a cube whose volume is twice the volume of the unit cube using ruler and compass ?

- Can we construct a regular n-gon using ruler and compass?

They were unable to answer these questions. We propose to study the question of constructibility by ruler and compass in full generality and to solve the previous problems.

9.1. **Definitions and first results.**

Definition. 1) Let $\mathcal{A}_0$ be a set of points in the real plane containing the points $(0,0),(1,0),(0,1)$.

We say that a point P of $\mathbb{R}^2$ is *constructible by ruler and compass* (CRC) *from* $\mathcal{A}_0$ if there exist of a sequence of points $P_1,\cdots,P_n=P$ satisfying the following property:

For $1\leq i\leq n$, P_i is constructed from the points of $\mathcal{A}_{i-1}:=\mathcal{A}_0\cup\{P_1,\cdots,P_{i-1}\}$ by intersection any two of the following geometric objects:

- a line passing through two points of $\mathcal{A}_{i-1}$

- a circle centered on a point of $\mathcal{A}_{i-1}$ and whose radius is the distance between two points of $\mathcal{A}_{i-1}$

2) We say that $\alpha=x+iy\in\mathbb{C}$ is CRC from $\mathcal{A}_0$ if (x,y) is CRC from $\mathcal{A}_0$.

3) We say that $P=(x,y)$ (resp. $\alpha\in\mathbb{C}$) is *absolutely CRC* if it is CRC from $\{(0,0),(1,0),(0,1)\}$.

Remark 9.1. Recall that using a ruler and a compass, we can construct perpendiculars and parallels passing through constructible points, and in particular projections of a constructible point on the axes of the real plane. Hence (x,y) is CRC from $\mathcal{A}_0$ if and only if $(x,0)$ and $(0,y)$ are. But $(0,y)$ is CRC if and only if $(y,0)$ is, so (x,y) is CRC from $\mathcal{A}_0$ if and only if the real numbers x,y are. It follows also that $x+iy\in\mathbb{C}$ is CRC from $\mathcal{A}_0$ if and only if x,y are.

The previous remark shows that it is enough to investigate which real numbers are CRC from $\mathcal{A}_0$.

It is time to give examples of constructible real numbers.

Lemma 9.2. *Every $\alpha \in \mathbb{Q}$ is constructible from any set $\mathcal{A}_0$.*

Proof. Indeed, one can assume that $\alpha > 0$, since if α is constructible, so is $-\alpha$ (use the compass to construct $(-\alpha, 0)$ from $(\alpha, 0)$).

Write $\alpha = \frac{p}{q}, p, q$. Since $(1,0),(0,1) \in \mathcal{A}_0$, one can construct the points $P = (0,p), Q = (q,0), Q' = (q+1,0)$. Now let M be the intersection of the line (PQ) with the line parallel to the $y-axis$ and passing through Q'. The use of Thales theorem shows that $M' = (q+1, \frac{p}{q})$. Hence $\frac{p}{q}$ is constructible by the previous remark. □

Proposition 9.3. *If $\alpha, \beta \in \mathbb{R}$ are CRC from $\mathcal{A}_0$, so are $\alpha \pm \beta, \alpha\beta$ and α^{-1}.*

Proof. For $\alpha \pm \beta$, it is obvious. For the other cases, one can assume that $\alpha, \beta > 0$.

For $\alpha\beta$, let $P = (\alpha, 0), Q = (\alpha+1, 0), Q' = (\alpha+1, -\beta)$ (Q' is easily seen to be constructible). Thales theorem show that the intersection of the line (QQ') with the y-axis is the point $(0, \alpha\beta)$.

For α^{-1}, let $P = (0,1)$ and $Q = (\alpha, 0)$. From these points, one can construct $Q' = (\alpha+1, 0)$. Now let M be the intersection of the line (PQ) with the line parallel to the $y-axis$ and passing through Q'. The use of Thales theorem shows that $M' = (\alpha+1, \frac{1}{\alpha})$. Hence α^{-1} is constructible.

□

Remark 9.4. One can show that this result is still true if $\alpha, \beta \in \mathbb{C}$.

Proposition 9.5. *If $d \in \mathbb{R}, d \geq 0$ is CRC from $\mathcal{A}_0$, then $\sqrt{d}$ is constructible from $\mathcal{A}_0$.*

Proof. If $d = 0$, there is nothing to prove, so we can assume $d \neq 0$.

Assume first that $d > 1$. Since d is constructible, so are $d-1$ and $d+1$. Now $\frac{1}{2}$ is constructible as well, so $\frac{d-1}{2}$ and $\frac{d+1}{2}$ are constructible. Now let us consider the circle centered in $P = (\frac{d-1}{2}, 0)$ with radius $\frac{d+1}{2}$, and let us take the intersection $Q = (0, y)$ with the y-axis.

Then OPQ is a rectangle triangle, so we have $y^2 + (\frac{d-1}{2})^2 = (\frac{d-1}{2})^2$. One can easily deduce from this equation that $y = \sqrt{d}$. Hence $(0, \sqrt{d})$ is constructible, so $\sqrt{d}$ is constructible.

Now if $0 < d < 1$, then $\frac{1}{d} > 1$, so $\sqrt{\dfrac{1}{d}} = \dfrac{1}{\sqrt{d}}$ is constructible, and thus its inverse $\sqrt{d}$ is constructible.

□

Remark 9.6. One can show that this result is still true if $d \in \mathbb{C}$.

Since a complex number $\alpha = x + iy$ is CRC from $\mathcal{A}_0$ if and only if x, y are, it is enough to investigate which real numbers are CRC from $\mathcal{A}_0$.

Theorem 9.7. *Let K be a subfield of $\mathbb{R}$, and assume that every element of K is CRC from $\mathcal{A}_0$.*

1) *Assume that $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$ are CRC from $\mathcal{A}_0$. Then every element of $K(\alpha_1, \cdots, \alpha_n)$ is CRC from $\mathcal{A}_0$.*

2) *Assume that $K \subseteq F \subseteq \mathbb{R}$. If $[F : K] \leq 2$, then every element of F is CRC from $\mathcal{A}_0$.*

Proof. 1) Since every element of $K(\alpha_1, \cdots, \alpha_n)$ is a rational fraction in $\alpha_1, \cdots, \alpha_n$ with coefficients in K, it is enough to show that if $\alpha, \beta \in K(\alpha_1, \cdots, \alpha_n)$ are CRC from $\mathcal{A}_0$, then $\alpha + \beta, \alpha\beta$ and α^{-1} are CRC from $\mathcal{A}_0$. This has been proved in the previous section.

2) If $[F : K] = 1$, then $F = K$ and there is nothing to prove. If $[F : K] = 2$, we have $F = K(\sqrt{d})$ for some d which is not a square in K. Since $F \subset \mathbb{R}$, we necessarily have $d > 0$. By the previous point, it is enough to show that $\sqrt{d}$ is constructible, which has been already done. □

This result initiates a link between constructibility by ruler and compass and field extensions. The goal of the next section is to investigate further this relation.

9.2. **Ruler and compass constructions and field extensions.** We now start to investigate the following question:

Let $\mathcal{A}_0$ be a finite set of points of $\mathbb{R}^2$ containing $(0,0), (1,0)$ and $(0,1)$. What are the reals numbers which are CRC from $\mathcal{A}_0$?

Definition. If K be any subfield of $\mathbb{R}$ and $\mathcal{A} = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ is a finite subset of $\mathbb{R}^2$, we set

$$K(\mathcal{A}) := K(x_1, y_1, \cdots, x_n, y_n) \subseteq \mathbb{R}$$

Remark 9.8. By a previous result, every element of $\mathbb{Q}(\mathcal{A}_0)$ is CRC from $\mathcal{A}_0$.

The following result gives a first link between our geometric problem and field extensions (even it has a limited interest):

Proposition 9.9. *Let $\alpha \in \mathbb{R}$. Then α is CRC from $\mathcal{A}_0$ if and only if every element of $\mathbb{Q}(\mathcal{A}_0)(\alpha)$ is CRC from $\mathcal{A}_0$.*

Proof. Assume that α is CRC from $\mathcal{A}_0$. Since every element of $\mathbb{Q}(\mathcal{A}_0)$ is CRC from $\mathcal{A}_0$, so is every element of $\mathbb{Q}(\mathcal{A}_0)(\alpha)$. The other implication is obvious since $\mathbb{Q}(\mathcal{A}_0)(\alpha)$ contains α. □

We now investigate the structure of the extension $\mathbb{Q}(\mathcal{A}_0)(\alpha)/\,\mathbb{Q}(\mathcal{A}_0)$.

Lemma 9.10. *Let $\mathcal{B}$ be a finite set of points of $\mathbb{R}^2$, and let P be obtained by intersecting by intersection any two of the following geometric objects:*

- a line passing through two points of $\mathcal{B}$

- a circle centered on a point of $\mathcal{B}$ and whose radius is the distance between two points of $\mathcal{B}$. Then $\mathbb{Q}(\mathcal{B}) \subseteq \mathbb{Q}(\mathcal{B} \cup \{P\})$ and we have

$$[\mathbb{Q}(\mathcal{B} \cup \{P\}) : \mathbb{Q}(\mathcal{B})] \leq 2$$

Proof. If $P = (x, y)$, it follows from the properties of field extensions and the definitions that we have

$$\mathbb{Q}(\mathcal{B} \cup \{P\}) = \mathbb{Q}(\mathcal{B})(x, y)$$

In particular $\mathbb{Q}(\mathcal{B}) \subseteq \mathbb{Q}(\mathcal{B} \cup \{P\})$.

For convenience, we will say the lines and circles described above are constructed on $\mathcal{B}$.

We first take a closer look to the equations of the lines and circles constructed on $\mathcal{B}$.

Let $P_0 = (x_0, y_0), P_1 = (x_1, y_1), P_2 = (x_2, y_2)$ be 3 points of $\mathcal{B}$. Suppose that P_0 and P_1 are distinct.

- The equation of the line (P_0P_1) is given by

$$(x_1 - x_0)(y - y_0) - (y_1 - y_0)(x - x_0) = 0$$

Hence the equation has the form

$$ax + by + c = 0 \text{ for some } a, b, c \in \mathbb{Q}(\mathcal{B})$$

- The equation of the circle centered in P_3 with radius P_0P_1 is given by

$$(x - x_3)^2 + (y - y_3)^2 = (x_1 - x_0)^2 + (y_1 - y_0)^2$$

hence it has the form

$$x^2 + y^2 + ax + by + c = 0 \text{ for some } a, b, c \in \mathbb{Q}(\mathcal{B})$$

Now assume that $P = (x, y)$ is obtained by intersecting two lines constructed on $\mathcal{B}$. Hence (x, y) is the solution of a linear system of the form

$$\begin{cases} ax + by + c = 0 \\ a'x + b'y + c' = 0 \end{cases}$$

for some $a, b, c, a', b,', c' \in \mathbb{Q}(\mathcal{B})$.

Using Cramer's Rule for example (or by solving the system directly), it is easy to see that $x, y \in \mathbb{Q}(\mathcal{B})$. In this case we have

$$\mathbb{Q}(\mathcal{B})(x, y) = \mathbb{Q}(\mathcal{B})(x)(y) = \mathbb{Q}(\mathcal{B})(y) = \mathbb{Q}(\mathcal{B})$$

In particular, $[\mathbb{Q}(\mathcal{B} \cup \{P\}) : \mathbb{Q}(\mathcal{B})] = 1$.

Assume now that $P = (x, y)$ is the intersection of a line and a circle constructed on $\mathcal{B}$. Hence (x, y) is the solution of a linear system of the form

$$\begin{cases} ax + by + c = 0 \\ x^2 + y^2 + a'x + b'y + c' = 0 \end{cases}$$

for some $a, b, a', b', c' \in \mathbb{Q}(\mathcal{B})$.

Notice that we have by construction $(a, b) \neq (0, 0)$. Say for example that $b \neq 0$, so $y = -\dfrac{ax + c}{b}$. It follows that $y \in \mathbb{Q}(\mathcal{B})(x)$. Hence we have

$$\mathbb{Q}(\mathcal{B})(x, y) = \mathbb{Q}(\mathcal{B})(x)(y) = \mathbb{Q}(\mathcal{B})(x)$$

Now plugging the expression of y is the second equation, we get

$$ux^2 + vx + w = 0 \text{ for some } u, v, w \in \mathbb{Q}(\mathcal{B})$$

Hence $\mu_{x,\mathbb{Q}(\mathcal{B})} | uX^2 + vX + w$, and we have

$$[\mathbb{Q}(\mathcal{B})(x) : \mathbb{Q}(\mathcal{B})] = \deg(\mu_{x,\mathbb{Q}(\mathcal{B})}) \leq 2$$

Finally assume that $P = (x, y)$ is the intersection of two circles constructed on $\mathcal{B}$. Hence (x, y) is the solution of a linear system of the form

$$\begin{cases} x^2 + y^2 + ax + by + c = 0 \\ x^2 + y^2 + a'x + by' + c' = 0 \end{cases}$$

for some $a, b, c, a', b', c' \in \mathbb{Q}(\mathcal{B})$.

By construction, we have $(a, b) \neq (a', b')$ (otherwise, the two circles are either eqaul or do not intersect), $(a-a', b-b') \neq (0, 0)$. By substracting the two equations, we go back to the previous case.

□

Definition. A field extension F/K is said to be 2-*decomposable* if there exists a tower of subfields

$$K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m = F$$

satisfying $[K_i : K_{i-1}] \leq 2$ for $1 \leq i \leq m$.

Remark 9.11. It follows from the definition and the Tower Formula that if F/K is 2-decomposable, then $[F : K]$ is a power of 2.

Examples.

- The extension $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is 2-decomposable. Indeed, we have

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2})$$

We know that $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$. Morever it is easy to check that $[\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}] = 4$, so by the Tower formula $[\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}(\sqrt{2})] = 2$, so we are done.

- The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ has degree 3, so it is not 2-decomposable.

Lemma 9.12. *Assume that $K \subseteq F \subseteq \mathbb{R}$, and that F/K is a 2-decomposable field extension. If every element of K is CRC from $\mathcal{A}_0$, then so is every element of F.*

Proof. Consider a tower of subfields

$$K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m = F$$

satisfying $[K_i : K_{i-1}] \leq 2$ for $1 \leq i \leq m-1$. Since $F \subseteq \mathbb{R}$, then $K_i \subseteq \mathbb{R}$ for all i. By assumption every element of $K_0 = K$ is CRC from $\mathcal{A}_0$. Since $[K_1 : K_0] \leq 2$, then so is every element of K_1. It implies in turns that it is true for every element of K_2, and by induction we obtain the result. □

We are now ready to prove the main theorem of this section.

Theorem 9.13. *Let $\alpha \in \mathbb{R}$. Then α is CRC from $\mathcal{A}_0$ if and only if there exists a field $F \subseteq \mathbb{R}$ such that $\mathbb{Q}(\mathcal{A}_0)(\alpha) \subseteq F$ and $F/\mathbb{Q}(\mathcal{A}_0)$ is 2-decomposable.*

Proof. Assume first that is α is CRC from $\mathcal{A}_0$, that is $P = (\alpha, 0)$ is CRC from $\mathcal{A}_0$. Let $P_1, \cdots, P_m = P$ points as in the definition of a constructible set, and let $\mathcal{A}_i = \mathcal{A}_{i-1} \cup \{P_i\}$ for all i.

As observed before, we have $\mathbb{Q}(\mathcal{A}_{i-1}) \subseteq \mathbb{Q}(\mathcal{A}_i)$ for all i, and then $\mathbb{Q}(\mathcal{A}_0) \subseteq \mathbb{Q}(\mathcal{A}_i)$.

We have $[\mathbb{Q}(\mathcal{A}_i) : \mathbb{Q}(\mathcal{A}_{i-1})] \leq 2$ by a previous result, hence the extension $\mathbb{Q}(\mathcal{A}_m)/\mathbb{Q}(\mathcal{A}_0)$ is 2-decomposable. Now by construction $\mathcal{A}_m$ contains the coordinates of the point P, so it constains α. Hence $\alpha \in \mathbb{Q}(\mathcal{A}_m)$. Since $\mathbb{Q}(\mathcal{A}_0) \subseteq \mathbb{Q}(\mathcal{A}_m)$, we get $\mathbb{Q}(\mathcal{A}_0)(\alpha) \subseteq \mathbb{Q}(\mathcal{A}_m)$.

Conversely, assume that $\mathbb{Q}(\mathcal{A}_0)(\alpha) \subseteq F \subseteq \mathbb{R}$, where $F/\mathbb{Q}(\mathcal{A}_0)$ is 2-decomposable. To prove that α is CRC from $\mathcal{A}_0$, it is enough to prove that every element of $\mathbb{Q}(\mathcal{A}_0)(\alpha)$ is CRC from $\mathcal{A}_0$. For, it is sufficient to prove that every element of F is CRC from $\mathcal{A}_0$.

But since $F/\mathbb{Q}(\mathcal{A}_0)$ is 2-decomposable and $F \subset \mathbb{R}$, then every element of F is CRC from $\mathcal{A}_0$ by the previous lemma, since we know that every element of $\mathbb{Q}(\mathcal{A}_0)$ is CRC from $\mathcal{A}_0$.

□

Corollary 9.14. *If $\alpha \in \mathbb{R}$ is CRC from $\mathcal{A}_0$ then α is algebraic over $\mathbb{Q}(\mathcal{A}_0)$ and $[\mathbb{Q}(\mathcal{A}_0)(\alpha) : \mathbb{Q}(\mathcal{A}_0)]$ is a power of 2.*

Proof. Since α is CRC from $\mathcal{A}_0$, it follows from the previous theorem that $\mathbb{Q}(\mathcal{A}_0)(\alpha)$ is contained in a 2-decomposable extension $F/\mathbb{Q}(\mathcal{A}_0)$. Then the Tower Degree Formula implies that $[\mathbb{Q}(\mathcal{A}_0)(\alpha) : \mathbb{Q}(\mathcal{A}_0)]$ divides $[F : \mathbb{Q}(\mathcal{A}_0)]$, which is a power of 2 since $F/\mathbb{Q}(\mathcal{A}_0)$ is a 2-decomposable field extension. Hence $[\mathbb{Q}(\mathcal{A}_0)(\alpha) : \mathbb{Q}(\mathcal{A}_0)]$ is a power of 2 as well.

□

Observe that if $\mathcal{A}_0 = \{(0,0),(1,0),(0,1)\}$, then $\mathbb{Q}(\mathcal{A}_0) = \mathbb{Q}$. Hence we get the following results:

Corollary 9.15. *Let $\alpha \in \mathbb{R}$. Then α is absolutely CRC if and only if there exists a field $F \subseteq \mathbb{R}$ such that $\mathbb{Q}(\alpha) \subseteq F$ and $F/\mathbb{Q}$ is 2-decomposable.*

Corollary 9.16 (Wantzel's Theorem)**.** *If $\alpha \in \mathbb{R}$ is absolutely CRC then α is algebraic over $\mathbb{Q}$ and $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ is a power of 2.*

Warning: One can construct infinitely many $\alpha \in \mathbb{R}$ which satisfy $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 4$ but such that α is not absolutely CRC (see exercise sheet), so the condition on the degree is not necessary but not sufficient.

9.3. **Applications.**

9.3.1. *Duplicating the cube.* We are interested here in the following question: can we contruct a cube with volume equal to 2 units ?

Clearly this is equivalent to determine if $\sqrt[3]{2}$ is absolutely constructible. But we know that $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, which is not a power of 2, so $\sqrt[3]{2}$ is not absolutely constructible, and the question has a negative answer.

9.3.2. *Circle quadrature.* The circle quadrature is the following problem: can we construct of square with same area than the unit circle?

Here we want to determine if $\sqrt{\pi}$ is absolutely CRC. Observe that $\mathbb{Q} \subseteq \mathbb{Q}(\pi) \subseteq \mathbb{Q}(\sqrt{\pi})$. Since π is transcendental, then $[\mathbb{Q}(\pi) : \mathbb{Q}]$ is infinite, and so is $[\mathbb{Q}(\sqrt{\pi}) : \mathbb{Q}]$. Hence π is not absolutely constructible, and the question has a negative answer.

9.3.3. *Trissection of angles.* It is well-known that it is always possible to construct $\theta/2$ from an angle θ using ruler and compass. But what about trissection ? Here we consider the following problem:

Given an angle θ, can we construct $\theta/3$ using ruler and compass ?

One can reformulate the problem as follows:

given the set of points $\mathcal{A}_0 := \{(0,0),(0,1),(1,0),(\cos(\theta),\sin(\theta))\}$, can we construct $P = (\cos(\theta/3),\sin(\theta/3))$ from $\mathcal{A}_0$?

We start by simplifying the problem:

set $\mathcal{A}_0' = \{(0,0),(0,1),(1,0),(\cos(\theta),0)\}$. We claim that P is CRC from $\mathcal{A}_0$ if and only if it is CRC from $\mathcal{A}_0'$.

From the definitions, one can see that it is enough to prove that $(\cos(\theta),\sin(\theta))$ is constructible from $\mathcal{A}_0'$, and that $(\cos(\theta),0)$ is constructible from $\mathcal{A}_0$ (a picture may be useful to help you understanding why).

The proof of this fact is very easy: starting from $\mathcal{A}_0'$, we can construct the point $Q = (\cos(\theta),\sin(\theta))$, since Q is one of the intersection points of the line $x = \cos(\theta)$ with the unit circle, and starting from $\mathcal{A}_0$, we can construct the point $Q' = (\cos(\theta),0)$ by projection on the x-axis.

Let us continue to simplify the question a bit:

reasoning as previously, we see that the point $(\cos(\theta/3),\sin(\theta/3))$ is CRC from $\mathcal{A}_0'$ if and only if $(\cos(\theta/3),0)$ is.

Indeed, if $(\cos(\theta/3),\sin(\theta/3))$ is constructible, then we know that the point $(\cos(\theta/3),0)$ is constructible, and conversely if $(\cos(\theta/3),0)$ is constructible, then $(\cos(\theta/3),\sin(\theta/3))$ is one of the intersection points of the line $x = \cos(\theta/3)$ with the unit circle.

Hence the question becomes:

is $\cos(\theta/3)$ CRC from $\mathcal{A}_0' := \{(0,0),(0,1),(1,0),(\cos(\theta),0)\}$?

First observe that $\mathbb{Q}(\mathcal{A}_0') = \mathbb{Q}(\cos(\theta))$, so we need to determine on wich conditions the extension $\mathbb{Q}(\cos(\theta))(\cos(\theta/3))/\,\mathbb{Q}(\cos(\theta))$ is contained in a subfield F over $\mathbb{R}$ such that $F/\,\mathbb{Q}(\cos(\theta))$ is 2-decomposable.

The formula

$$\cos(\theta) = 4\cos(\theta/3)^3 - 3\cos(\theta/3)$$

implies that $[\mathbb{Q}(\cos(\theta))(\cos(\theta/3)) : \mathbb{Q}(\cos(\theta))] \leq 3$.

Indeed, the polynomial $P = 4X^3 - 3X - \cos(\theta) \in \mathbb{Q}(\cos(\theta))[X]$ satisfies $P(\cos(\theta/3)) = 0$. Hence $\mu_{\cos(\theta/3),\mathbb{Q}(\cos(\theta))}|P$, so we get

$$[\mathbb{Q}(\cos(\theta))(\cos(\theta/3)) : \mathbb{Q}(\cos(\theta))] = \deg(\mu_{\cos(\theta/3),\mathbb{Q}(\cos(\theta))}) \leq 3$$

If $[\mathbb{Q}(\cos(\theta))(\cos(\theta/3)) : \mathbb{Q}(\cos(\theta))] = 3$, it implies that $\cos(\theta/3)$ is not CRC from $\mathcal{A}_0'$ by a previous result.

If $[\mathbb{Q}(\cos(\theta))(\cos(\theta/3)) : \mathbb{Q}(\cos(\theta))] \leq 2$, then we know that every element of $\mathbb{Q}(\cos(\theta))(\cos(\theta/3))$ is CRC from $\mathcal{A}_0'$; in particular so is $\cos(\theta/3)$.

Hence we have

$$\cos(\theta/3) \textit{ is CRC from } \mathcal{A}_0' \iff [\mathbb{Q}(\cos(\theta))(\cos(\theta/3)) : \mathbb{Q}(\cos(\theta))] \leq 2$$

This is equivalent to say that $\deg(\mu_{\cos(\theta/3)}) < 3$. Since $\deg(\mu_{\cos(\theta/3)})|P$, this is also equivalent to say that P is not irreducible in $\mathbb{Q}(\cos(\theta))[X]$. Since P has degree 3, this is equivalent to say that P has a root in $\mathbb{Q}(\cos(\theta))[X]$.

Putting things together, we proved the following theorem:

Theorem 9.17. *The trissection of the angle θ using ruler and compass is possible if and only if the polynomial $4X^3 - 3X - \cos(\theta)$ has a root in $\mathbb{Q}(\cos(\theta))$.*

Examples.

- Let $\theta = 2\pi/3$. Then $\cos(\theta) = -1/2$ and so $\mathbb{Q}(\cos(\theta)) = \mathbb{Q}$. We then have to decide if the polynomial $4X^3 - 3X + 1/2$ has a root in $\mathbb{Q}$. This is equivalent to check for the roots of $8X^3 - 6X + 1$. Here the possible roots are $\pm 1, \pm 1/2, \pm 1/4, \pm 1/8$, and one can check that none of these rational numbers are roots of $8X^3 - 6X + 1$.

Hence $2\pi/9$ cannot be constructed from $2\pi/3$ using ruler and compass.

- Let $\theta = \pi/4$. Then $\cos(\theta) = 1/\sqrt{2}$, an so $\mathbb{Q}(\cos(\theta)) = \mathbb{Q}(\sqrt{2})$. We then have to decide if the polynomial $4X^3 - 3X + 1/\sqrt{2}$ has a root in $\mathbb{Q}(\sqrt{2})$. One can check that $-1/\sqrt{2}$ is effectively a root of this polynomial.

Hence $\pi/12$ can be constructed from $\pi/4$ using ruler and compass.

There is an easier way to see it here. Since $\cos(\pi/6) = \sqrt{3}/2$, we can construct $\cos(\pi/6)$ and then the angle $\pi/6$ using ruler and compass. Then we construct the angle $\pi/12$ from the angle $\pi/6$, since bissection of angles is always possible.

9.3.4. *Construction of regular n-gons.* Here we are interested in constructing the regular n-gon whose points lie on the unit circle. Clearly, it is equivalent to construct the point $P = (\cos(2\pi/n), \sin(2\pi/n))$, since the other points of the n-gon can be obtained from P and $(1, 0)$ using the compass. Once again, the construction of P is equivalent to the construction of $\cos(2\pi/n)$. Hence the question is equivalent to:

is $\cos(2\pi/n)$ absolutely CRC ?

We don't have enough material to answer fully this question, but we can give few examples.

If $n = 2^m$, then the answer is positive, since starting from the angle π, one can construct $2\pi/n$ by successive bissections in this case.

The case of odd n is funnier. If $n = 3$, then $\cos(2\pi/3) = -1/2$ is constructible. If $n = 9$, we saw in the previous paragraph that $[\mathbb{Q}(\cos(2\pi/9)) : \mathbb{Q}] = 3$, so $2\pi/9$ is not constructible.

We can also show that the heptagon ($n = 7$) is not constructible. However, the pentagon ($n = 5$) and the heptakaidecagon ($n = 17$) are constructible. See the exercise sheet for the case $n = 5$.

10. Symmetric polynomials

Let R be a commutative ring, and let $P \in R[X_1, \dots, X_n]$ be a polynomial in n variables.

Remember that such P can be written uniquely under the form

$$P = \sum a_{m_1,m_2,\dots,m_n} X_1^{m_1} \cdots X_n^{m_n}, \qquad a_{m_1,m_2,\dots,m_n} \in R \text{ almost all zero}$$

Example: $n = 3, f = X_1^2 + 3X_2^3 - 8X_2X_3 + 45$.

Definition. We say that f is symmetric if it is unchanged by any permutation π of the n variables. This means

$$f(X_1, \dots, X_n) = f(X_{\pi(1)}, \dots, X_{\pi(n)}) \text{ for all permutations } \pi \in S_n.$$

Examples:

- If $n = 3$, $X_1 + X_2 + X_3$ is symmetric.

- If $n = 4$, $X_1 + X_2 + X_3$ is **NOT** symmetric. Indeed, if we apply the permutation which exchanges 1 and 4, and leaves 2 and 3 invariant, we get a different polynomial.

- If $n = 3$, $X_1X_2X_3$ is symmetric.

Elements of R (viewed as polynomials of degree zero) are symmetric, so in particular 1 and -1 are. Moreover, if P_1 and P_2 are symmetric, then so are $P_1 + P_2$ and P_1P_2 (it follows directly from the definition). It follows that the symmetric polynomials form a subring of $R[X_1, \cdots, X_n]$.

Definition. The *elementary symmetric polynomials* $\sigma_1, \dots, \sigma_n \in R[X_1, \dots, X_n]$ are the polynomials

$$\sigma_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} X_{i_1} X_{i_2} \cdots X_{i_k}.$$

For example, $\sigma_1 = X_1 + X_2 + \cdots X_n$, $\sigma_n = X_1X_2 \cdots X_n$.

These polynomials are symmetric.

Proposition 10.1. *Let K be a field and let $P = a_nT^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0 \in K[T], \deg(P) = n \geq 1$. Let $\alpha_1, \cdots, \alpha_n$ the (not necessarily distinct) roots of P in a suitable field extension L/K. Then we have :*

$$\textit{For all } 1 \le k \le n, \sigma_k(\alpha_1, \cdots, \alpha_n) = (-1)^k \frac{a_{n-k}}{a_n}.$$

Proof. We have $P = a_n(T - \alpha_1) \cdots (T - \alpha_n)$. To get the coefficient of T^{n-k}, we have to choose the term "T" in $n-k$ factors in every possible ways and choose a term of the form "$-\alpha_i$" in the remaining factors, and the choices are made in an ordered sequence (we choose first T or $-\alpha_1$ in the first factor, then T or $-\alpha_2$ in the second factor...).

Hence the coefficient of T^{n-k} is $a_n(-1)^k \sum_{1\leq i_1<i_2<\cdots<i_k\leq n} \alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_k}$, that is $a_n(-1)^k\sigma_l(\alpha_1,\cdots,\alpha_n)$. But this coefficient is also equal to a_{n-k}, hence the formula. □

Example: If $P = a_2X^2 + a_1X + a_0$, and if α_1, α_2 are the two roots of P, then $\alpha_1 + \alpha_2 = -\frac{a_1}{a_2}$ and $\alpha_1\alpha_2 = \frac{a_0}{a_2}$.

Let's check it: we have for example

$$\alpha_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_0a_2}}{2a_2}, \alpha_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_0a_2}}{2a_2}, \text{ so}$$

$$\alpha_1 + \alpha_2 = -\frac{2a_1}{2a_2} = -\frac{a_1}{a_2}, \text{ and}$$

$$\alpha_1\alpha_2 = \frac{(-a_1)^2 - (a_1^2 - 4a_0a_2)}{(2a_2)^2} = \frac{4a_0a_2}{4a_2^2} = \frac{a_0}{a_2}.$$

Definition. A *monomial* is an element of $R[X_1,\cdots,X_n]$ of the form $cX_1^{m_1}\cdots X_n^{m_n}, c \in R$.

The *degree* of a monomial $cX_1^{m_1}\cdots X_n^{m_n}, c \neq 0$ is $m_1 + \cdots + m_n$. The *degree* of a polynomial P is the maximum of degrees of the non-zero monomials appearing in P.

We say that a polynomial $P \in R[X_1,\cdots,X_n]$ is *homogeneous of degree* $d \geq 0$ if it is a sum of monomials of same degree d.

Each polynomial $P \in R[X_1,\ldots,X_n]$ of degree can be written uniquely as a sum of homogeneous polynomials (just collect the monomials of same degree appearing in P).

Example: the polynomial σ_k is homogeneous of degree k.

Remark 10.2. It follows from the definition that if $P \in R[X_1,\ldots,X_n]$ is homogeneous of degree d, then for all $r \in R$, we have

$$P(rX_1,\ldots,rX_n) = r^dP(X_1,\ldots,X_n),$$

which explains the name of "homogeneous polynomial".

Theorem 10.3 (Fundamental theorem on symmetric polynomials)**.** *Let R be a commutative ring. For any* **symmetric** *polynomial $P \in R[X_1,\ldots,X_n]$, there exists a* **unique** *polynomial Q with coefficients in R such that $P = Q(\sigma_1,\cdots,\sigma_n)$.*

Proof. We first prove that if such a Q exists, then it is unique. Assume that $P = Q_1(\sigma_1,\cdots,\sigma_n) = Q_2(\sigma_1,\cdots,\sigma_n)$, so $(Q_1-Q_2)(\sigma_1,\cdots,\sigma_n) = 0$.

We will prove that $S(\sigma_1,\cdots,\sigma_n) = 0$ for some polynomial S, then $S = 0$. Uniqueness will follow.

We will do it by induction on n. For $n = 1$, this is clear.

Now assume that the property is true for polynomials in n variables, and assume that the property is NOT true for polynomials in $n + 1$ variables,so there exists $S \neq 0$ of minimal degree m such that

$$S(\sigma_1, \cdots, \sigma_{n+1}) = 0(*)$$

We will denote by $\Sigma_1, \cdots, \Sigma_n$ the elementary symmetric polynomials in the variables $X_1 \cdots, X_n$. Notice that

$$\sigma_k(X_1, \cdots, X_n, 0) = \Sigma_k(**)$$

We can write $S = S_m X_{n+1}^m + \cdots + S_1 X_n + S_0, S_0, \cdots, S_m \in R[X_1, \cdots, X_n]$.

Assume first $S_0 = 0$, then we can write $S = X_n U$, with $U \neq 0$ of degree $m - 1$. By assumption on S, we get $\sigma_{n+1} U(\sigma_1, \cdots, \sigma_{n+1}) = 0$, and so $U(\sigma_1, \cdots, \sigma_{n+1}) = 0$, but this contradicts the minimality of m.

Therefore $S_0 \neq 0$, and replacing X_{n+1} by 0 in $(*)$ and using $(**)$, we get $S_0(\Sigma_1, \Sigma_n) = 0$. By induction, $S_0 = 0$, contradiction.

Now to prove the existence, we will give an effective algorithm.

First write $P = P_{i_1} + \cdots + P_{i_r}$, where each P_{i_j} is homogeneous of degree i_j. Now since this decomposition is unique, one can see that each P_{i_j} is symmetric (**Check it** !!!)

Therefore it is sufficient to do the case of homogeneous polynomials. Now assume that P is homogeneous and symmetric.

We order $\mathbb{N}^n$ as follows: we say that $(m_1, \cdots, m_r) \geq (m'_1, \cdots, m'_n)$ if either $m_1 > m'_1$, either there exists k such that $m_1 = m'_1, \cdots m_{k-1} = m'_{k-1}$ and $m_k > m'_k$.

Now take $(m_1, \cdots, m_n)$ maximal among the non-zero monomials $a_{m_1, \cdots, m_n} X_1^{m_1} \cdots X_n^{m_n}$ of P. We denote it by $\mathcal{D}(P)$.

By choice of $\mathcal{D}(P)$, we have $m_1 \geq m_2$. If $m_2 < m_3$, then applying a permutation exchanging X_2 and X_3 and fixing 1, we find a non-zero monomial such that the sequence of powers is $(m_1, m_3, m_2, \cdots)$. But this is greater than $\mathcal{D}(P)$.

Repeating the argument, we see that $\mathcal{D}(P) = (m_1, \cdots, m_n)$ with $m_1 \geq m_2 \geq \cdots \geq m_n$. Therefore $\mathcal{D}(P)$ is also the unique non-decreasing sequence such that m_1 is maximal among the non-zero monomials $a_{m_1, \cdots, m_n} X_1^{m_1} \cdots X_n^{m_n}$ of P.

Here is the algorithm for P homogeneous and symmetric:

1) Among the non-zero monomials $a_{m_1, \cdots, m_n} X_1^{m_1} \cdots X_n^{m_n}$ of P, find $(m_1, \cdots, m_n)$, with m_1 maximal and $m_1 \geq m_2 \geq \cdots \geq m_n$.

2) Set $P_1 = P - a_{m_1, \cdots, m_n} \sigma_1^{m_1 - m_2} \sigma_2^{m_2 - m_3} \cdots \sigma_{n-1}^{m_{n-1} - m_n} \sigma_n^{m_n}$.

One can check that P_1 is symmetric, and that $P_1 = 0$ or $\mathcal{D}(P_1) < \mathcal{D}(P)$.

3) If $P_1 = 0$, then stop. Otherwise, replace P by P_1 and goto Step 1).

The algorithm stops necessarily. □

Example: Let $P = X_1^2X_3 + X_1^2X_2 + X_2^2X_1 + X_2^2X_3 + X_3^2X_1 + X_3^2X_2$. This is an homogeneous polynomial of degree 3. Here $\mathcal{D}(P) = (2,1,0)$.

So we set $P_1 = P - \sigma_1^{2-1}\sigma_2^{1-0}\sigma_3^0 = P - \sigma_1\sigma_2$. One can check that $P_1 = -3X_1X_2X_3$, and therefore $P = \sigma_1\sigma_2 - 3\sigma_3$.

Example: Express the symmetric polynomial

$$X_1^2(X_2X_3 + X_2X_4 + X_3X_4) + X_2^2(X_1X_3 + X_1X_4 + X_3X_4)+ \\ X_3^2(X_1X_2 + X_1X_4 + X_2X_4) + X_4^2(X_1X_2 + X_1X_3 + X_2X_3)$$

in terms of the σ_i.

This is an homogeneous polynomial (of degree 4). We have $\mathcal{D}(P) = (2,1,1)$.

We then set $P_1 = P - 1.\sigma_1^{2-1}\sigma_2^{1-1}\sigma_3^1 = P_1 - \sigma_1\sigma_3$.

But

$$\sigma_1\sigma_3 = (X_1+X_2+X_3+X_4)(X_1X_2X_3+X_1X_2X_4+X_1X_3X_4+X_2X_3X_4)$$

and so

$$P_1 = -4X_1X_2X_3X_4 = -4\sigma_4.$$

Therefore $P = \sigma_1\sigma_3 - 4\sigma_4$.

Theorem 10.4. *Let K be a field. Let $P \in K[T]$ be a polynomial of degree $n \geq 1$, and let $\alpha_1, \ldots, \alpha_n$ be the (non necessarily distinct) roots of P in a larger field L containing K. Then for every* **symmetric** *polynomial $f \in K[X_1, \ldots, X_n]$, we have*

$$f(\alpha_1, \ldots, \alpha_n) \in K.$$

In other words, every symmetric polynomial expression in the roots of P is an element of K.

More precisely, if $P = a_nT^n + \ldots + a_1T + a_0$ and f has degree d, then there exists a polynomial $h \in K[Y_0, \ldots, Y_{n-1}, Y_n]$ such that

$$f(\alpha_1, \ldots, \alpha_n) = \frac{h(a_0, \ldots, a_n)}{a_n^d}.$$

In particular, if P is monic, every symmetric polynomial expression in the roots of P is a polynomial expression in the coefficients of P.

Proof. Let us decompose f into a sum of homogeneous polynomials

$$f = f_0 + f_1 + \ldots + f_d,$$

wheere each f_i is homogeneous of degree i.

By the fundamental theorem on symmetric polynomials, we can write

$$f_i(X_1,\ldots,X_n)=Q_i(\sigma_1(X_1,\ldots,X_n),\ldots,\sigma_n(X_1,\ldots,X_n)),$$

for some polynomial $Q_i\in K[T_1,\ldots,T_n]$. Since f_i is homogeneous of degre i, we have

$$f_i(a_nX_1,\ldots,a_nX_n)=a_n^i f_i(X_1,\ldots,X_n).$$

We then obtain

$$a_n^d f_i(X_1,\ldots,X_n)=a_n^{d-i}a_i^i f_i(X_1,\ldots,X_n)=a_n^{d-i}f_i(a_nX_1,\ldots,a_nX_n).$$

But since σ_k is homogeneous of degree k, we also have

$$\sigma_k(a_nX_1,\ldots,a_nX_n)=a_n^k\sigma_k,$$

and therefore

$$Q_i(\sigma_1(a_nX_1,\ldots,a_nX_n),\ldots,\sigma_n(a_nX_1,\ldots,a_nX_n))=Q_i(a_n\sigma_1,\ldots,a_n^n\sigma_n).$$

We then finally get

$$a_n^d f_i(X_1,\ldots,X_n)=a_n^{d-i}Q_i(a_n\sigma_1,\ldots,a_n^n\sigma_n).$$

By Proposition 10.1, we get

$$a_n^d f_i(\alpha_1,\ldots,\alpha_n)=a_n^{d-i}Q_i(-a_{n-1},a_{n-2}a_n,\ldots,(-1)^n a_0a_n^{n-1}).$$

Now summing over i gives the result. □

Example. Let $P=a_2X^2+a_1X+a_0\in K[X]$, and let α_1,α_2 be the two roots of P (in a suitable field). Then from the previous result, we should have

$$\alpha_1^2+\alpha_2^2\in K,$$

and in fact $a_2^2(\alpha_1^2+\alpha_2^2)$ should be a polynomial expression in a_0,a_1,a_2.

Let us check it. We have

$$\alpha_1^2+\alpha_2^2=(\alpha_1+\alpha_2)^2-2\alpha_1\alpha_2=\sigma_1(\alpha_1,\alpha_2)^2-2\sigma_2(\alpha_1,\alpha_2).$$

By Proposition 10.1, we get

$$\alpha_1^2+\alpha_2^2=\frac{a_1^2}{a_2^2}-2\frac{a_0}{a_2}=\frac{a_1^2-2a_0a_2}{a_2^2}.$$

Proposition 10.5. *Let K be a field. Let $P=a_nT^n+\ldots+a_1T+a_0\in K[T]$ be a polynomial of degree $n\geq 1$, and let $\alpha_1,\ldots,\alpha_n$ be the (non necessarily distinct) roots of P in a larger field L containing K. Then the element*

$$\mathrm{disc}(P)=a_n^{n(n-1)}\Delta(\alpha_1,\ldots,\alpha_n)=a_n^{n(n-1)}\prod_{1\leq i<j\leq n}(\alpha_i-\alpha_j)^2$$

is an element of K, which is even a polynomial expression into the coefficients of P.

Moreover, $\mathrm{disc}(P)\neq 0$ *if and only if P has no multiple roots in L.*

Proof. Set

$$\Delta = \prod_{1\leq i<j\leq n} (X_i - X_j)^2 \in K[X_1, \dots, X_n].$$

We claim that Δ is symmetric.

To see it, consider the Vandermonde matrix

$$M = (X_j^{i-1})_{1\leq i,j\leq n} \in \mathrm{M}_n(K(X_1, \dots, X_n)).$$

It is well-known that the determinant of this matrix is

$$\det(M) = \prod_{1\leq i<j\leq n} (X_i - X_j).$$

Hence $\Delta = \det(M)^2$. Now let $\tau \in S_n$ be a permutation. Then we have

$$\Delta(X_{\tau(1)}, \dots, X_{\tau(n)}) = \det(M_\tau)^2,$$

where

$$M_\tau = (X_{\tau(j)}^{i-1})_{1\leq i,j\leq n},$$

since a permutation of the variables the variables $X_1, \dots, X_n$ induces the same permutation of the columns of M. Since M_τ is just obtained by a permutation of the columns of M, we have

$$\det(M_\tau) = \pm\det(M),$$

and therefore

$$\Delta(X_{\tau(1)}, \dots, X_{\tau(n)}) = \det(M_\tau)^2 = \det(M)^2 = \Delta(X_1, \dots, X_n).$$

Hence Δ is a symmetric polynomial of degree $n(n-1)$. By the previous theorem, the element

$$\mathrm{disc}(P) = a_n^{n(n-1)}\Delta(\alpha_1, \dots, \alpha_n) = a_n^{n(n-1)} \prod_{1\leq i<j\leq n} (\alpha_i - \alpha_j)^2$$

is a polynomial expression into the coefficients of P, called the *discriminant* of P.

By definition, $\mathrm{disc}(P) \neq 0$ if and only if $\alpha_1, \dots, \alpha_n$ are distinct.

□

Definition. The element $\mathrm{disc}(P)$ is called the *discriminant* of P.

Example: if $P = a_2T^2 + a_1T + a_0 \in K[T]$, we have

$$\mathrm{disc}(P) = a_2^2(\alpha_1 - \alpha_2)^2 = a_2^2((\alpha_1 + \alpha_2)^2 - 4\alpha_1\alpha_2).$$

Since we have

$$\alpha_1 + \alpha_2 = \sigma_1(\alpha_1, \alpha_2) = -\frac{a_1}{a_2},$$

and

$$\alpha_1\alpha_2 = \sigma_2(\alpha_1, \alpha_2) = \frac{a_0}{a_2},$$

we get

$$\operatorname{disc}(P) = a_2^2(\frac{a_1^2}{a_2^2} - 4\frac{a_0}{a_2}) = a_1^2 - 4a_2a_0,$$

which is the classical definition of the discriminant of a polynomial of degree 2.

Example: Assume that $P = a_3T^3 + a_2T^2 + a_1T + a_0 \in K[T]$. Applying the algorithm given in the proof of the fondamental theorem on symmetric polynomials, one can show (using Maple, for example) that

$$\Delta = -4\sigma_1\sigma_3 + \sigma_1^2\sigma_2^2 - 4\sigma_2^3 + 18\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2.$$

In particular, if $P = T^3 + pT + q$, we have

$$\sigma_1(\alpha_1, \alpha_2, \alpha_3) = 0, \sigma_2(\alpha_1, \alpha_2, \alpha_3) = p, \sigma_3(\alpha_1, \alpha_2, \alpha_3) = -q,$$

and therefore

$$\operatorname{disc}(P) = -4p^3 - 27q^2,$$

which is the classical quantity to decide whether of not P has multiple roots.

11. Algebraic integers

Definition. A complex number $\alpha \in \mathbb{C}$ is said to be an *algebraic integer* if there exists a **monic** $f \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$.

Examples:

- Any $n \in \mathbb{Z}$ is an algebraic integer, because it is a root of the monic polynomial $X - n$.

- The rational $\frac{1}{2}$ is not an algebraic integer: it is a root of the polynomial $2X - 1$, but it can be shown that it cannot be a root of any monic polynomial in $\mathbb{Z}[X]$.

- $\sqrt{2}$ is a root of the monic polynomial $X^2 - 2$, so is an algebraic integer.

- i is an algebraic integer since it is a root of the monic polynomial $X^2 + 1$.

- $i + \sqrt{2}$ is an algebraic integer: To see this, consider the polynomial

$$(X - i - \sqrt{2})(X - i + \sqrt{2})(X + i - \sqrt{2})(X + i + \sqrt{2}) = X^4 - 2X^2 + 9.$$

- More easily, $i\sqrt{2}$ is an algebraic integer since it is a root of $X^2 + 2$.

Theorem 11.1. *The set of algebraic integers is a subring of* $\mathbb{C}$.

Proof. Clearly, 0 and 1 are algebraic integers.

Let α be a root of $f(x) \in \mathbb{Z}[X]$, a monic polynomial of degree m, and similarly let β be a root of $g(x) \in \mathbb{Z}[X]$ a monic polynomial of degree n. Over $\mathbb{C}$, these polynomials split (by the fundamental theorem of algebra):

$$f(X) = \prod_{i=1}^{m}(X - \alpha_i) = X^m + a_{m-1}X^{m-1} + \cdots + a_1 X + a_0,$$

$$g(x) = \prod_{i=1}^{m}(X - \beta_i) = X^n + b_{n-1}X^{n-1} + \cdots + b_1 X + b_0,$$

where $\alpha = \alpha_1$ and $\beta = \beta_1$ say.

The easy part of the proof is to note that $-\alpha$ is a root of

$$(-1)^m f(-X) = X^m - a_{m-1}X^{m-1} + a_{m-2}X^{m-1} + \cdots + (-1)^m a_0.$$

Now consider

$$F(X) = \prod_{i=1}^{m}\prod_{j=1}^{n}(X - \alpha_i - \beta_j).$$

This has $\alpha + \beta$ as a root, and is monic. Thus it suffices to show that $F(X) \in \mathbb{Z}[X]$. Let

$$F_1(X, Y_1, \dots, Y_n) = \prod_{i=1}^{m} \prod_{j=1}^{n} (X - \alpha_i - Y_j) \in \mathbb{C}[X, Y_1, \dots, Y_n].$$

Then

$$F_1(X, Y_1, \dots, Y_n) = \prod_{j=1}^{n} \prod_{i=1}^{m} (X - Y_j - \alpha_i) = \prod_{i=1}^{n} f(X - Y_j).$$

Therefore, $F_1 \in \mathbb{Z}[X, Y_1, \dots, Y_n]$. Moreover, F_1 is symmetric in $Y_1, \dots, Y_n$. Applying the fundamental theorem on symmetric polynomials in the case when $R = \mathbb{Z}[x]$, it follows that

$$F_1(X, Y_1, \dots, Y_n) = F_2(X, \sigma_1, \dots, \sigma_n) \in \mathbb{Z}[x, \sigma_1, \dots, \sigma_n]$$

for some polynomial F_2, where σ_j is the jth elementary symmetric function in the Y_j. Substituting $Y_j = \beta_j$ gives $\sigma_j(\beta_1, \cdots, \beta_n) = (-1)^j b_{n-j}$. Hence

$$F(X) = F_1(X, \beta_1, \dots, \beta_n) = F_2(X, -b_{n-1}, b_{n-2}, \dots, (-1)^n b_0) \in \mathbb{Z}[X].$$

Similarly, let

$$G(X) = \prod_{i=1}^{m} \prod_{j=1}^{n} (X - \alpha_i \beta_j),$$

and define

$$G_1(X, Y_1, \dots, Y_n) = \prod_{i=1}^{m} \prod_{j=1}^{n} (X - \alpha_i Y_j).$$

Then

$$G_1(X, Y_1, \dots, Y_n) = \prod_{j=1}^{n} Y_j^m f(X/Y_j),$$

where

$$Y_j^m f(X/Y_j) = X^m + a_{m-1} X^{m-1} Y_j + a_{m-2} X^{m-2} Y_j^2 + \cdots + a_1 X Y_j^{m-1} + a_0 Y_j^m.$$

Then $G_1 \in \mathbb{Z}[X, Y_1, \dots, Y_n]$ is symmetric, so

$$G_1(X, Y_1, \dots, Y_n) = G_2(X, \sigma_1, \dots, \sigma_n) \in \mathbb{Z}[X, \sigma_1, \dots, \sigma_n],$$

and hence

$$G(X) = G_1(X, \beta_1, \dots, \beta_n) = G_2(X, -b_{n-1}, b_{n-2}, \dots, (-1)^n b_0) \in \mathbb{Z}[X].$$

□

Corollary 11.2. *Let K be a subfield of $\mathbb{C}$. The set of elements of K which are algebraic integers is a subring of K.*

Definition. If K is a subfield of $\mathbb{C}$, we set $\mathcal{O}_K = \{\alpha \in K | \alpha$ is an algebraic integer $\}$. This is a ring, called *the ring of integers of K*. An element of $\mathcal{O}_K$ is called *an integer of K*.

Lemma 11.3. *Let $\alpha \in \mathbb{C}$. Assume that α is algebraic over $\mathbb{Q}$. Then*

$$\alpha \textit{ is an algebraic integer } \iff \mu_{\alpha,\mathbb{Q}} \in \mathbb{Z}[X].$$

Proof. If $\mu_{\alpha,\mathbb{Q}} \in \mathbb{Z}[X]$ then α is an algebraic integer by definition. Conversely assume that α is an algebraic integer. Then $f(\alpha) = 0$ for some monic polynomial $f \in \mathbb{Z}[X]$. We know that $f = Q.\mu_{\alpha,\mathbb{Q}}$ for some $Q \in \mathbb{Q}[X]$, since $f(\alpha) = 0$. Notice that $c(f) \sim 1$ since $f \in \mathbb{Z}[X]$ is monic. Since $c(f) \sim c(Q).c(\mu_{\alpha,\mathbb{Q}})$, we get that $c(\mu_{\alpha,\mathbb{Q}})$ is a unit of $\mathbb{Z}$, and therefore $\mu_{\alpha,\mathbb{Q}} \in \mathbb{Z}[X]$. □

Example: The integers of $\mathbb{Q}$ are the integers ! In other words, $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$. Indeed if $r \in \mathbb{Q}$, then $\mu_{r,\mathbb{Q}} = X - r$, which lies in $\mathbb{Z}[X]$ if and only if $r \in \mathbb{Z}$.

In general, $\mathcal{O}_K$ is very difficult to describe. However, the case where $K = \mathbb{Q}(\sqrt{d})$ is completely known:

Proposition 11.4. *Let $d \in \mathbb{Z}$ be a square-free integer, and let $K = \mathbb{Q}(\sqrt{d})$. Then we have:*

1) $\mathcal{O}_K = \mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d}, a, b \in \mathbb{Z}\}$ *if* $d \not\equiv 1[4]$

2) $\mathcal{O}_K = \mathbb{Z}[\dfrac{1+\sqrt{d}}{2}] = \{a + b.\dfrac{1+\sqrt{d}}{2}, a, b \in \mathbb{Z}\}$ *if* $d \equiv 1[4]$.

Proof. Since $\sqrt{d}$ is a root of $X^2 - d$, $\sqrt{d} \in \mathcal{O}_K$. Since $\mathcal{O}_K$ is a ring, we deduce that $\mathbb{Z}[\sqrt{d}] \subset \mathcal{O}_K$. Similarly, if $d \equiv 1[4]$, then $\frac{1+\sqrt{d}}{2}$ is a root of $X^2 - X - \frac{d-1}{4} \in \mathbb{Z}[X]$, so $\frac{1+\sqrt{d}}{2} \in \mathcal{O}_K$, and therefore $\mathbb{Z}[\frac{1+\sqrt{d}}{2}] \subset \mathcal{O}_K$.

Now assume that $z = a + b\sqrt{d} \in K$ is an algebraic integer, and let's prove the missing inclusions.

If $b = 0$, $z = a \in \mathbb{Q}$, and so we know that $z \in \mathbb{Z}$ and we are done.

Assume that $b \neq 0$.

Let $f = (X-(a+b\sqrt{d}))(X+(a-b\sqrt{d})) = X^2-2aX+(a^2-b^2d) \in \mathbb{Q}[X]$.

Since d is square-free,$\sqrt{d} \notin \mathbb{Q}$, so f has no roots in $\mathbb{Q}$ ($b \neq 0$!!) and is therefore irreducible. Thus $f = \mu_{z,\mathbb{Q}}$. By assumption, we get $2a \in \mathbb{Z}$ and $a^2 - b^2d \in \mathbb{Z}$.

Notice also we therefore have $a-b\sqrt{d}$ is an integer of K, and substracting with z gives that $2b\sqrt{d} \in \mathcal{O}_K$. Now since $\sqrt{d}$ is also an integer of K, we get that $2bd$ is also an integer of K. Since $2bd \in \mathbb{Q}$, it implies that $2bd \in \mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

To sum up, we have $2bd, 2a, a^2 - b^2d \in \mathbb{Z}$.

Case 1: $a\,\mathbb{Z}$. In this case, $b^2d \in \mathbb{Z}$. Write $b = \frac{u}{v}, h.c.f(u,v) = 1$. Then $v^2|u^2d$, but since $h.c.f(u^2, v^2) = 1$, then $v^2|d$, which is impossible since d is square-free, unless $v = \pm 1$, and so $b \in \mathbb{Z}$.

Therefore $z = a + b\sqrt{d}, a, b \in \mathbb{Z} \in \mathbb{Z}[\sqrt{d}]$. Notice also that $a + b\sqrt{d} = (a - b) + 2b.\frac{1+\sqrt{d}}{2}$, so we also have $z \in \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$, which covers all the cases.

Case 2: $a \in \frac{1}{2}\mathbb{Z}$. In this case $a = \frac{m}{2}, m \in \mathbb{Z}$, and we can assume that m is odd (otherwise we go back to Case 1).

Then $a^2 - b^2 d = \frac{m^2}{4} - b^2 d \in \mathbb{Z}$, so $m^2 - 4b^2 d \in 4\,\mathbb{Z}$, so $4b^2 d = (2b)^2 d \in \mathbb{Z}$. Reasoning as in Case 1 shows that $2b \in \mathbb{Z}$. Set $b = \frac{n}{2}, n \in \mathbb{Z}$.

We have $m^2 - 4bd = m^2 - n^2 d \in \mathbb{Z}$. Since m is odd, we get $m^2 \equiv 1[4]$, which implies that $n^2 d \equiv 1[4]$, so n is odd. This implies in turn that $d \equiv 1[4]$.

Now we have $z = a + b\sqrt{d} = \frac{m+n\sqrt{d}}{2}, m, n$ odd.

Therefore $m - n$ is even and thus $\frac{m-n}{2} \in \mathbb{Z}$.

Consequently $z = \frac{m-n}{2} + n.\frac{1+\sqrt{d}}{2} \in \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$.

□