

ALGEBRA 411-412
PARTIAL NOTES

CONTENTS

1. Rings	2
1.1. Rings	2
1.2. Constructions of new rings from old	2
1.3. Subrings	3
1.4. Product or sum of rings	3
1.5. Isomorphisms of rings	4
1.6. Ideals and quotients of rings	4
1.7. Homomorphisms of rings	5
1.8. The group of invertible elements in a ring	6
1.9. Commutative rings and determinants over commutative rings	6
2. Zero divisors, integral domains and fields	8
2.1. Zero divisors	8
2.2. Fields and Integral domains	9
2.3. Division rings	10
2.4. Characteristic of a field	11
3. Homomorphisms and ideals	11
3.1. Kernels and images of homomorphisms	11
3.2. The first isomorphism theorem for rings	11
3.3. Homomorphisms “preserve”	14
3.4. Relations between different ideals	15
3.5. Maximal ideals and prime ideals	18
3.6. Calculation of $\text{Hom}(R, S)$	20
4. Rings of polynomials	20
4.1. Division of polynomials	28

Date: ?

4.2.	Some consequences of division of polynomials	31
5.	Fraction fields	33
5.1.	The problem	33
5.2.	The quotient construction of new sets	35
5.3.	Constructing Numbers: \mathbb{Z} from \mathbb{N}	38
5.4.	Construction of the fraction field of an integral domain	38
6.	Algebra and geometry	40
7.	Modules	41

In 411 we studied *Groups* and in 412 we are interested in *Rings*.

1. Rings

The set of natural numbers is defined as $\mathbb{N} \stackrel{\text{def}}{=} \{0, 1, 2, \dots\}$.

1.1. **Rings.** The notion of a ring is an abstraction of the notion of a *system of numbers*. For instance the basic systems of numbers we have encountered $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}_n$ are all rings (but \mathbb{N} does not make it). However, the notion of rings also allows more complicated objects such as matrices.

A *ring* $(R, +, \cdot)$ consists of a set R and two operations $+, \cdot$ called addition and multiplication, such that

- (1) $(R, +)$ is an abelian group,
- (2) \cdot is associative and has a neutral element (which we denote 1 and call *unity*).
- (3) Operations $+$ and \cdot are compatible in the sense that they satisfy two distributivity properties

$$(a + b)c = ac + bc \quad \text{and} \quad c(a + b) = ca + cb.$$

1.1.1. *Examples.* $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ and $M_n(\mathbb{R})$. \mathbb{N} is not a ring since $(\mathbb{N}, +)$ is not a group.

1.1.2. *Example.* Addition and multiplication modulo n give a ring: $(\mathbb{Z}_n, +_n, \cdot_n)$ is a ring for any $n = 0, 1, 2, \dots$. Here $r_n(k)$ denotes the remainder of k modulo n and

$$a +_n b \stackrel{\text{def}}{=} r_n(a + b) \quad \text{and} \quad a \cdot_n b \stackrel{\text{def}}{=} r_n(a \cdot b) \quad \text{for } a, b \in \mathbb{Z}_n.$$

1.2. Constructions of new rings from old.

1.2.1. *Motivation.* We will motivate these with analogous constructions for groups. Recall that we have one or more groups, we know several ways to construct new groups:

- (1) Take a subgroup. (In particular any homomorphism of groups $\phi : G \rightarrow H$ defines two subgroups $\text{Ker}(\phi) \subseteq G$ and $\text{Im}(\phi) \subseteq H$).
- (2) Product of groups (also called sum).
- (3) Quotient group G/N .
- (4) Stabilizer subgroups when a group acts on a set.
- (5) Centralizer subgroups.

Some of these will adapt to rings:

- (1) Take a subring of a given ring. (In particular any homomorphism of rings $\phi : R \rightarrow S$ defines a subring $\text{Im}(\phi) \subseteq S$).
- (2) Product of rings (also called sum).
- (3) Quotient ring R/I .
- (4) Matrix rings $M_n(R)$.

1.3. **Subrings.** A subset S of a ring $(R, +, \cdot)$ is a *subring* if

- S contains the unit 1 from R .
- operations $+, \cdot$ on R restrict to operations on S and make S into a ring,

In particular any subring has a structure of a ring. Here is a simple criterion for checking whether a subset is a subring:

Lemma. $S \subseteq R$ is a subring iff

- It is closed under difference and under multiplication, i.e., for any $u, v \in S$, both $u - v$ and uv are again in S .
- $S \ni 1_R$.

Examples. (a) All of inclusions $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ are subrings.

(b) $\mathbb{Z}[i] \stackrel{\text{def}}{=} \{a + bi; a, b \in \mathbb{Z}\}$ is a subring of \mathbb{C} .

(c) $\mathbb{Q}[\sqrt{2}] \stackrel{\text{def}}{=} \{a + b\sqrt{2}; a, b \in \mathbb{Q}\}$ is a subring of \mathbb{R} .

1.4. **Product or sum of rings.** We can combine two rings R, S to get a new ring which we call the *product* of R and S and denote $R \times S$. (We also call it the sum of rings and denote $R \oplus S$.)

Lemma. If $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$ are rings then $R \times S$ with operations

$$(r, s) + (r', s') \stackrel{\text{def}}{=} (r +_R r', s +_S s') \quad \text{and} \quad (r, s) \cdot (r', s') \stackrel{\text{def}}{=} (r \cdot_R r', s \cdot_S s')$$

is also a ring.

Example. $\mathbb{Z}_2 \times \mathbb{Z}_3$.

1.5. Isomorphisms of rings. These are the simplest relations between rings.

For two rings $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$, a function $f : R \rightarrow S$ is said to be an *isomorphism of rings* if

(1) It preserves sums and products:

$$f(a +_R b) = f(a) +_S f(b) \quad \text{and} \quad f(a \cdot_R b) = f(a) \cdot_S f(b) \quad \text{for } a, b \in R,$$

(2) it preserves units: $f(1_R) = 1_S$; and

(3) it is a bijection (i.e., a one-to-one correspondence).

The idea is that if we know an isomorphism f from R to S then for all practical purposes the two rings are “the same” – everything works the same in R and in S , just the names for elements are different (and f is the dictionary which translates one set of names into the other).

We say that rings R and S are isomorphic if there exists an isomorphism of rings $f : R \rightarrow S$. We denote “ R and S are isomorphic” by: $R \cong S$.

Example. The product of \mathbb{Z}_2 and \mathbb{Z}_3 is not a really new ring for us, it is just a new way of thinking about an old ring:

$$\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6.$$

Lemma. (a) If $R_1 \xrightarrow{\phi} R_2$ and $R_2 \xrightarrow{\psi} R_3$ are both isomorphisms of rings then the composition $R_1 \xrightarrow{\psi \circ \phi} R_3$ is also an isomorphism of rings.

(b) If $R \xrightarrow{\phi} RS$ is an isomorphism of rings then the inverse function $\phi^{-1} : S \rightarrow R$ exists and it is also an isomorphism of rings.

Corollary. Relation \cong of isomorphism of rings is an *equivalence relation*, i.e., it satisfies

- (1) If $R_1 \cong R_2$ and $R_2 \cong R_3$ then $R_1 \cong R_3$.
- (2) If $R \cong S$ then $S \cong R$.

1.6. Ideals and quotients of rings. A subset I of a ring R is said to be an *ideal* if

- (1) it is a subgroup of $(R, +)$ (i.e., it is nonempty and if $x, y \in I$ then $x - y \in I$), and
- (2) it is closed under multiplication with elements of R , i.e.,

$$x \in I \text{ and } a \in R \Rightarrow ax, xa \in I.$$

Lemma. If I is an ideal in R then the set of cosets $R/I = \{r + I; r \in R\}$ is a ring for the operations

$$(a + I) +_{R/I} (b + I) \stackrel{\text{def}}{=} (a + b) + I \quad \text{and} \quad (a + I) \cdot_{R/I} (b + I) \stackrel{\text{def}}{=} ab + I.$$

Theorem. (a) Ideals in the ring \mathbb{Z} are the same as subgroups of $(\mathbb{Z}, +)$. They are all of the form $n\mathbb{Z}$ for some $n \in \mathbb{N}$.

(b) Rings \mathbb{Z}_n and $\mathbb{Z}/n\mathbb{Z}$ are canonically isomorphic. The two natural isomorphisms (in two directions) are

- (1) $\phi : \mathbb{Z}_n \xrightarrow{\cong} \mathbb{Z}/n\mathbb{Z}$, $\phi(k) \stackrel{\text{def}}{=} k + n\mathbb{Z}$ for $k \in \mathbb{Z}_n$.
- (2) $\psi : \mathbb{Z}/n\mathbb{Z} \xrightarrow{\cong} \mathbb{Z}_n$, $\psi(k + n\mathbb{Z}) \stackrel{\text{def}}{=} r_n(k)$ for $k \in \mathbb{Z}$.

These are mutually inverse isomorphisms, i.e., $\psi = \phi^{-1}$.

Remark. One can prove part (b) without knowing in advance that $(\mathbb{Z}_n, +_n, \cdot_n)$ is a ring:

- We know from (a) that $\mathbb{Z}/n\mathbb{Z}$ is a ring since $n\mathbb{Z}$ is an ideal.
- It is easy to see that the functions ϕ, ψ are inverse bijections.
- We can use ϕ, ψ to move operations $+_{\mathbb{Z}/n\mathbb{Z}}, \cdot_{\mathbb{Z}/n\mathbb{Z}}$ from $\mathbb{Z}/n\mathbb{Z}$ to \mathbb{Z}_n , this gives operations on \mathbb{Z}_n

$$a +_{new} b \stackrel{\text{def}}{=} \phi^{-1}[\phi(a) +_{\mathbb{Z}/n\mathbb{Z}} \phi(b)] \quad \text{and} \quad a \cdot_{new} b \stackrel{\text{def}}{=} \phi^{-1}[\phi(a) \cdot_{\mathbb{Z}/n\mathbb{Z}} \phi(b)].$$

- Since everything works the same for $(\mathbb{Z}_n, +_{new}, \cdot_{new})$ as for the ring $(\mathbb{Z}/n\mathbb{Z}, +_{\mathbb{Z}/n\mathbb{Z}}, \cdot_{\mathbb{Z}/n\mathbb{Z}})$, the new operations on \mathbb{Z}_n also satisfy the properties needed to be a ring. So, $(\mathbb{Z}_n, +_{new}, \cdot_{new})$ is a ring and ϕ, ψ are isomorphisms of rings $(\mathbb{Z}_n, +_{new}, \cdot_{new})$ and $(\mathbb{Z}/n\mathbb{Z}, +_{\mathbb{Z}/n\mathbb{Z}}, \cdot_{\mathbb{Z}/n\mathbb{Z}})$.
- One now calculates what operations $+_{new}, \cdot_{new}$ do and finds that they are the same as $+_n, \cdot_n$. Therefore, $(\mathbb{Z}_n, +_n, \cdot_n)$ is a ring and ϕ, ψ are isomorphisms of rings $(\mathbb{Z}_n, +_n, \cdot_n)$ and $(\mathbb{Z}/n\mathbb{Z}, +_{\mathbb{Z}/n\mathbb{Z}}, \cdot_{\mathbb{Z}/n\mathbb{Z}})$.

1.7. Homomorphisms of rings. Homomorphisms of rings are basic ways of relating two rings.

For two rings $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$, a function $f : R \rightarrow S$ is said to be an *homomorphism of rings* if

- (1) It preserves sums and products:

$$f(a +_R b) = f(a) +_S f(b) \quad \text{and} \quad f(a \cdot_R b) = f(a) \cdot_S f(b) \quad \text{for } a, b \in R,$$

- (2) It preserves units: $f(1_R) = 1_S$.

(One often says “morphism” instead of the more traditional “homomorphism”.)

Lemma. (a) If R is a ring and $S \subseteq R$ is a subring then the inclusion map

$$i : S \rightarrow R, i(s) = s \quad \text{for } s \in S;$$

is a homomorphism of rings.

(b) If I is an ideal in a ring R then the quotient map

$$q: R \rightarrow R/I, \quad q(r) \stackrel{\text{def}}{=} r + I \quad \text{for } r \in R;$$

is a morphism of rings.

Remarks. (0) By definitions, a map f between two rings is an isomorphism iff it is a homomorphism and a bijection. So, isomorphisms are examples of homomorphism.

(1) Homomorphisms are ways to naturally relate two rings. The simplest homomorphisms are isomorphisms, they tell us that “everything is the same” in two rings. However, as we see in the lemma when we “travel down a homomorphism” information can be created or lost.

1.7.1. Rings of matrices.

Lemma. For any ring R , the set $M_n(R)$ of $n \times n$ matrices with entries in the ring R , is again a ring for the usual operations of addition and multiplication of matrices.

1.8. The group of invertible elements in a ring. We say that an element a of a ring R is *invertible* if there exists an element $b \in R$ such that $ab = 1 = ba$.

Lemma. If such b exists it is unique.

So, we can denote it a^{-1} and call it the *inverse of a* .

Proposition. The subset $R^* \subseteq R$ of all invertible elements of R is a group for multiplication operation inherited from R .

Example. (a) $\mathbb{Z}^* = \{\pm 1\}$ while $\mathbb{Q}^* = \mathbb{Q} - \{0\}$, $\mathbb{R}^* = \mathbb{R} - \{0\}$ and $\mathbb{C}^* = \mathbb{C} - \{0\}$.

(b) We will check later that for quaternions one also has

$$\mathbb{H}^* = \mathbb{H} - \{0\}.$$

(c) From linear algebra we recall that $A \in M_n(\mathbb{R})$ is invertible iff $\det(A) \neq 0$.

Lemma. \mathbb{Z}_n^* consists precisely of all $k \in \mathbb{Z}_n$ which are relatively prime to n .

Remark. So, group \mathbb{Z}_n^* is the group we denoted $U(n)$ (elements of \mathbb{Z}_n relatively prime to n with operation \cdot_n of multiplication modulo n).

Book denotes R^* by $U(R)$.

1.9. Commutative rings and determinants over commutative rings.

1.9.1. *Commutative rings.* A ring R is said to be commutative if the multiplication operation is commutative.

Example. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}_n$ are commutative but matrices $M_n(\mathbb{R})$ are not commutative for $n > 1$. Neither are the quaternions commutative.

1.9.2. *Linear algebra over commutative rings.* Calculations with matrices work “the same” for matrices $M_n(R)$ with entries in a commutative ring R , as they do in the case $R = \mathbb{R}$. Here, “the same” is not literally true, as we will see the difference appears when we want to invert a matrix. The cause of this is that in \mathbb{R} any nonzero element is invertible and this is no longer true for all commutative rings, for instance for $R = \mathbb{Z}$ only ± 1 are invertible.

However as long as we just use sums of matrices $A + B$, products of matrices AB and determinants, there is no difference. For instance:

Lemma. (a) If R is commutative, the determinant $\det : M_N(R) \rightarrow R$ is defined in the same way as for $R = \mathbb{R}$ and all the standard tricks for computing determinants work as well.

(b) $\det(AB) = \det(A) \cdot \det(B)$.

(c) For a square matrix $A \in M_n(R)$ over a commutative ring R . define the *adjoint matrix* \tilde{A} so that the entry at position (i, j) is

$$\tilde{A}_{ij} \stackrel{\text{def}}{=} (-1)^{i+j} \alpha_{ji},$$

where α_{ji} is the determinant of the matrix obtained by removing from A both the j^{th} row and the i^{th} column.

Then the product of A and its adjoint matrix in either order is a multiple of the unit matrix 1_n

$$A \cdot \tilde{A} = \det(A) \cdot 1_n = \tilde{A} \cdot A.$$

(Here, 1_n denotes the $n \times n$ matrix with 1 on the main diagonal and zeroes elsewhere.)

“*Proof.*” (a) and (b) have the same proofs as over real numbers, because all calculations that you did in Linear Algebra to define and calculate determinants used only the properties of numbers that are true for any commutative ring: $a + b, ab$ are defined, $b + a = a + b$, $ba = ab$, associativity of both $+$ and \cdot , distributivity of \cdot over $+$.

(c) is a standard formula from linear algebra. We may come back to it later. Now we just check this claim for $n = 2$

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, you can easily calculate that

- the adjoint matrix is $\tilde{A} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, and then
- $A \cdot \tilde{A} = (ad - bc) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\tilde{A} \cdot A = (ad - bc) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

1.9.3. *General Linear groups.* We denote by $GL_n(R)$ the group $M_n(R)^*$ of invertible elements of the ring $M_n(R)$. We know that this is group. It is called the *general linear group* of $n \times n$ matrices with entries in R . We will be interested in $GL_n(R)$ when R is a commutative ring (this is complicated enough!).

Theorem. Let R be a commutative ring. A matrix $A \in M_n(R)$ has an inverse in $M_n(R)$ iff its determinant $\det(A) \in R$ has inverse in R .

Proof. If A has an inverse $B \in M_n(R)$ then $AB = 1_n$ and therefore

$$1 = \det(1_n) = \det(AB) = \det(A) \cdot \det(B)$$

Examples. (0) Invertible elements of \mathbb{R} are all numbers except zero, so $A \in M_n(\mathbb{R})$ is invertible in $M_n(\mathbb{R})$ iff $\det(A) \neq 0$.

(1) Invertible elements of \mathbb{Z} are ± 1 , so $A \in M_n(\mathbb{Z})$ is invertible in $M_n(\mathbb{Z})$ iff $\det(A) = \pm 1$.

Remark. The proof shows that the formula for A^{-1} is the usual one

$$A^{-1} = \det(A)^{-1} \cdot \tilde{A}.$$

So, for any invertible 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ over a commutative ring R the inverse is

$$A^{-1} = (ad - bc)^{-1} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(You can multiply to check that the product is 1_2 .)

2. Zero divisors, integral domains and fields

2.1. **Zero divisors.** We say that an element A of a ring R is a 0-divisor if $a \neq 0$ and it divides zero, i.e., there exists some $b \neq 0$ such that $ab = 0$ or $ba = 0$.

Examples. (0) In \mathbb{Z}_6 we have zero divisors 2, 3, 4 (since $2 \cdot_6 3 = 4 \cdot_6 3 = 0$).

(1) There are no 0-divisors in \mathbb{Z} .

(2) In $Z \times Z = \mathbb{Z} \oplus \mathbb{Z}$, element $(1, 0)$ is a zero divisor since $(1, 0) \cdot (0, 1) = (0, 0) = 0_{\mathbb{Z} \times \mathbb{Z}}$.

Lemma. (a) Invertible elements are not zero divisors.

(b) If $0 \neq a \in R$ and a is not a zero divisor then one can cancel a from equalities, i.e.,

$$\text{If } ax = ay \text{ for some } x, y \in R, \text{ then actually } x = y.$$

Theorem. In a finite ring any element which is not 0 nor a 0-divisor is invertible.

Corollary. For a finite ring R (with at least two elements), any element a satisfies precisely one of the following three possibilities:

- (1) $a = 0$,
- (2) a is a zero divisor,
- (3) a is invertible.

Proof.

Example. For an element $k \neq 0$ in \mathbb{Z}_n :

- (a) k is a zero divisor iff it is not relatively prime to n ;
- (b) k is invertible iff it is relatively prime to n .

Proof. (1) First we notice that if k is not relatively prime to n then k is a 0-divisor:

If $d > 1$ is a common divisor then $k = dq$ and $n = dQ$ for some integers q, Q . Then $0 \neq Q = n/d < n$, hence $Q \in \mathbb{Z}_n$ and therefore $Q \neq 0$ in \mathbb{Z}_n . Now $k \cdot_n Q = r_n(dqQ) = r_n(qn) = 0$.

(2) Now notice that if k is relatively prime to n then k is not 0-divisor:

If $l \in \mathbb{Z}_n$ and $k \cdot_n l = 0$, then $k \cdot_n l = r_n(kl)$, so n divides kl . Now, since n has no common factors with k and n divides kl , we see that n divides l . Of course if $l \in \mathbb{Z}_n$ and l is a multiple of n then $l = 0$.

(3) So, we have proved (a). Then (b) follows because by the above corollary, k is invertible iff it is not a zero divisor, i.e., iff it is relatively prime to n .

2.2. Fields and Integral domains.

2.2.1. *Fields.* A *field* is a commutative ring F such that any non-zero element is invertible, i.e., $F - \{0\} \subseteq F^*$.

Examples. (a) It is clear that $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields but \mathbb{Z} is not.

(b) \mathbb{Z}_n is a field iff n is a prime.

Proof. The non-invertible elements of a finite ring \mathbb{Z}_n are 0 and the 0-divisors, i.e., elements which are not prime to n . Therefore, \mathbb{Z}_n is a field iff all non zero elements $1, \dots, n - 1$ in \mathbb{Z}_n are relatively prime to n . This is the same as asking that n is a prime.

2.2.2. *Integral domains.* An *integral domain* is a commutative ring with no zero divisors. The second requirement really says that if $a, b \neq 0$ then $ab \neq 0$.

Examples. (a) Any field is an integral domain.

(b) The simplest example of an integral domain which is not a field is \mathbb{Z} . Actually, the word “integral” in “integral domain” is supposed to remind us of integers, so roughly, integral domains are commutative rings which are “as good as integers”.

Lemma. (a) In an integral domain one can cancel any non zero element from equalities, i.e., $ax = ay$ and $a \neq 0$ implies that $x = y$.

(b) Any subring S of an integral domain R is again an integral domain.

Example. In particular any subring of a field is an integral domain, for instance $\mathbb{Z}[\sqrt{2}] \subseteq \mathbb{R}$, $\mathbb{Z}[i] \subseteq \mathbb{C}$.

Proposition. Any finite integral domain is a field!

Proof. In a finite ring any non-zero element which is not a 0-divisor is invertible!

2.2.3. Subfields. We say that a subset S of a field F is a *subfield* if it satisfies either of the following equivalent conditions (a), (b) or (c) :

Lemma. For a subset S of a field F , the following are equivalent

- (a)
 - (1) $S \ni 1_F$,
 - (2) S inherits operations $+$, \cdot from F , and
 - (3) S is a field for these operations.
- (b) S is a subring and for any $0 \neq a \in S$ its inverse a^{-1} in F lies in S .
- (c)
 - (1) $S \ni 1_F$,
 - (2) $a, b \in S$ implies $a - b, ab \in S$.
 - (3) $0 \neq a \in S$ implies $a^{-1} \in S$.

Examples. (a) $\mathbb{Q}[i] \subseteq \mathbb{C}$ is a subfield.

(b) $\mathbb{Q}[\sqrt{2}] \subseteq \mathbb{R}$ is a subfield.

(c) $\mathbb{Z} \subseteq \mathbb{Q}$ and $\mathbb{Z}[i] \subseteq \mathbb{C}$ are not subfields.

2.3. Division rings. A *division ring* is a ring D such that any non-zero element is invertible, i.e., $D - \{0\} \subseteq F^*$.

So it is like a field except that we do not ask for commutativity, i.e., a field is the same as a division ring which is commutative.

2.3.1. Examples. (i) Fields are division rings. (ii) The simplest division ring which is not a field (i.e., not commutative) is the ring of quaternions \mathbb{H} .

Lemma. Finite division rings are fields.

2.4. Characteristic of a field.

Lemma. For a field F , and an integer n the following are equivalent

- (1) $n \cdot 1_F = 0$,
- (2) $n \cdot F = 0$.

Definition. (i) We say that the *characteristic* of a field F is zero if for any positive integer n , $n \cdot 1_F \neq 0$.

(ii) We say that the *characteristic* of a field F is *finite* if there is a positive integer n such that $n \cdot 1_F = 0$. Then the least such n is called the *characteristic* of F .

We denote the characteristic of F by $\text{char}(F)$.

Lemma. If the characteristic of F is finite then $\text{char}(F)$ is a prime.

Example. $\text{char}(F) = 0$ for $F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, while for a finite field \mathbb{Z}_p we have $\text{char}(\mathbb{Z}_p) = p$.

Theorem. (a) If the characteristic of a field F is finite then F contains \mathbb{Z}_p .

(b) If the characteristic of F is zero then F contains \mathbb{Q} .

Remark. The book defines characteristic for all rings.

3. Homomorphisms and ideals

3.1. Kernels and images of homomorphisms.

Lemma. Let $f : R \rightarrow S$ be a morphism of rings, then

- (a) The *image* $f(R) \stackrel{\text{def}}{=} \{f(r); r \in R\} \subseteq S$ of f is a subring of S .
- (b) The kernel $\text{Ker}(f) \stackrel{\text{def}}{=} \{r \in R; f(r) = 0\} \subseteq R$ is an ideal in R .

3.2. The first isomorphism theorem for rings. We will now see that any morphism of rings leads to an isomorphism of related rings. Later we will use this observation to construct many interesting isomorphisms between rings.

Theorem. For any morphism of rings $f : R \rightarrow S$, there is canonical isomorphism of rings

$$\bar{f} : R/\text{Ker}(f) \xrightarrow{\cong} f(R)$$

given by the formula

$$\bar{f}(r + \text{Ker}(f)) \stackrel{\text{def}}{=} f(r), \quad r \in R.$$

3.2.1. *Graphic rendition of morphisms and of the first isomorphism theorem.* The structure of a situation which involves several maps is sometimes indicated graphically by a diagram of maps, say

$$\begin{array}{ccccc} A & & B & & \\ \rho \downarrow & & \sigma \downarrow & & \\ C & \xrightarrow{g} & D & \xrightarrow{f} & E \\ & & \alpha \downarrow & & \beta \downarrow \\ & & X & \xleftarrow{z} & Y \end{array}$$

One can also indicate graphically how some functions work

$$\begin{array}{ccc} \mathbb{Z} \xrightarrow{u} \mathbb{Q} & n \mapsto \frac{2}{3}n^2 - 2n + 1 & b^2/2 + 1 \\ x \downarrow & \downarrow & \text{and} & y \uparrow & ; \\ \mathbb{R} \xleftarrow{v} \mathbb{Q} & \sqrt{n} & & 2b \leftarrow b \end{array}$$

means that $u(n) = \frac{2}{3}n^2 - 2n + 1$, $y(b) = b^2/2 + 1$ etc.

We say that a *diagram* \mathcal{D} of maps *commutes* if for any two sets A, B in the diagram \mathcal{D} , all ways of going from A to B coincide. For instance, commutativity of the diagrams

$$\begin{array}{ccc} A \xrightarrow{f} B & & A \xrightarrow{u} B \\ \alpha \downarrow & \beta \downarrow & \text{and} & x \downarrow & y \uparrow \\ C \xrightarrow{g} D & & C \xrightarrow{v} D \end{array}$$

means that

$$\beta \circ f = g \circ \alpha \quad \text{and} \quad y \circ v \circ x = u.$$

This is often denoted by drawing a directed circle inside triangles, squares and other polygons in \mathcal{D} such that their edges split into two composable groups and the two compositions coincide. (In our example a circle would be drawn inside the two squares above.)

Corollary. For any morphism of rings $f : R \rightarrow S$, there is a commutative diagram

$$\begin{array}{ccc} R & \xrightarrow{f} & S \\ p \downarrow & & i \uparrow \\ R/\text{Ker}(f) & \xrightarrow{\bar{f}} & f(R) \end{array} \quad \text{with} \quad \begin{array}{ccc} r & \mapsto & f(r) \\ \downarrow & & \uparrow \\ r + \text{Ker}(f) & \mapsto & f(r) \end{array}.$$

Here, $p : R \rightarrow R/\text{Ker}(f)$ is the quotient map $p(r) = r + \text{Ker}(f)$ and $i : f(R) \rightarrow$ is the inclusion map $i(x) = x$.

Remark. Commutativity means that

$$f = i \circ \bar{f} \circ p.$$

So we say that f *factors* through \bar{f} , or that \bar{f} is a *factorization* of f .

3.2.2. *The content of the first isomorphism theorem.* For any ideal I in a ring R , the ring R/I is smaller than R . The reason is that it consists of cosets $r + I$ for the ideal I – any $r \in R$ gives one such coset $r + I$, but often two different elements r_1, r_2 of R give the same coset. For such elements r_1, r_2 , the distinction between r_1 and r_2 is lost when we pass from R to R/I .

For any homomorphism $f : R \rightarrow S$ the subring $f(R)$ of S is also smaller than R . The reason is that it consists of values $f(r)$ on elements $r \in R$ – any $r \in R$ gives one such element $f(r)$ of $f(R)$, but often two different elements r_1, r_2 of R have the same f -image, i.e., $f(r_1) = f(r_2)$. Again, for such elements r_1, r_2 , the distinction between r_1 and r_2 is lost when we pass from R to the image $f(R)$ of the morphism f .

Now, if we start with a morphism $f : R \rightarrow S$ and then choose ideal I in R as the kernel $\text{Ker}(f)$, it turns out that the passage from R to $f(R)$ and the passage from R to $R/I = R/\text{Ker}(f)$ lose the same information! The reason is that

$$f(r_1) = f(r_2) \Leftrightarrow f(r_1 - r_2) = 0 \Leftrightarrow r_1 - r_2 \in \text{Ker}(f) \Leftrightarrow r_1 + \text{Ker}(f) = r_2 + \text{Ker}(f).$$

Therefore, the two objects $f(R)$ and $R/\text{Ker}(f)$ that we got from R and f should really be the same! Moreover, the way to identify them should be that $f(r) \in f(R)$ should correspond to $r + \text{Ker}(f) \in R/\text{Ker}(f)$.

One can also summarize this by:

- $f(R)$ and $R/\text{Ker}(f)$ are both obtained from the ring R by “squeezing the ideal $\text{Ker}(f)$ into a point”.

(The first time this happens because f takes $\text{Ker}(f)$ into 0_S and the second time by the definition of the quotient ring $R/\text{Ker}(f)$.)

Remark. Now we have seen why the first isomorphism theorem is true. But what do we gain from this theorem? The beauty of the isomorphism $R/\text{Ker}(f) \xrightarrow{\cong} f(R)$ is that the first object is intimately related to R and the second to S .

So, the theorem says that once we have a homomorphism between two rings R and S some “parts” of the two rings are forced to be the same. (Here the vague word “part” appears in two different meanings – once it is a quotient of the ring and the other time is a subring.)

3.2.3. *An example of the first isomorphism theorem: comparison of rings \mathbb{Z}_n and $\mathbb{Z}/n\mathbb{Z}$.*

Theorem. The morphism of rings

$$r_n : \mathbb{Z} \rightarrow \mathbb{Z}_n$$

factors to an isomorphism

$$\iota : \mathbb{Z}/n\mathbb{Z} \xrightarrow{\cong} \mathbb{Z}_n, \quad \iota(k + n\mathbb{Z}) = r_n(k) \text{ for } k \in \mathbb{Z};$$

The inverse isomorphism is

$$\iota^{-1} : \mathbb{Z}_n \xrightarrow{\cong} \mathbb{Z}/n\mathbb{Z}, \quad \iota^{-1}(x) = x + n\mathbb{Z}, \quad x \in \mathbb{Z}_n.$$

Proof. The word “factors” suggests that we will use the first isomorphism theorem to construct ι as the factorization \bar{r}_n of the homomorphism r_n to isomorphism of rings

$$\bar{r}_n : \mathbb{Z}/\text{Ker}(r_n) \xrightarrow{\cong} r_n(\mathbb{Z}).$$

Indeed, we take $\iota = \bar{r}_n$ then we know that ι is an isomorphism but we need to check that the kernel $\text{Ker}(r_n)$ is $n\mathbb{Z}$ and the image $r_n(\mathbb{Z}) \subseteq \mathbb{Z}_n$ is all of \mathbb{Z}_n . Both of these claims are obvious.

Finally, recall that the isomorphism

$$\bar{r}_n : \mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/\text{Ker}(r_n) \xrightarrow{\cong} r_n(\mathbb{Z}) = \mathbb{Z}_n$$

is given by the formula $\bar{r}_n(k + \text{Ker}(r_n)) = r_n(k)$. So, $\iota(k + n\mathbb{Z}) \stackrel{\text{def}}{=} \bar{r}_n(k + \text{Ker}(r_n)) = r_n(k)$.

In the opposite direction, function $f : \mathbb{Z}_n \xrightarrow{\cong} \mathbb{Z}/n\mathbb{Z}$ defined by $f(x) = x + n\mathbb{Z}$ for $x \in \mathbb{Z}_n$; is clearly the inverse of the function $\iota(k + n\mathbb{Z}) = r_n(k)$. Therefore, $f = \iota^{-1}$ and this implies that f is also an isomorphism.

3.3. Homomorphisms “preserve”.

3.3.1. Homomorphisms preserve polynomials.

Lemma. (1) Let P be a polynomial $P(X) = a_0 + a_1X + \cdots + a_nX^n$ with integer coefficients $a_0, \dots, a_n \in \mathbb{Z}$. For any ring R , one can evaluate P at any element r of R , i.e.,

$$P(r) \stackrel{\text{def}}{=} a_0 + a_1r + \cdots + a_nr^n$$

is a well defined element of R .

So, such P defines a function $\tilde{P} : R \rightarrow R$, $\tilde{P}(r) \stackrel{\text{def}}{=} P(r)$. (We will often denote the function \tilde{P} simply by P if it does not cause confusion.)

(2) Any homomorphism of rings $f : R \rightarrow S$ satisfies

- (a) $f(a \cdot x) = a \cdot f(x)$ for any $x \in R$ and $a \in \mathbb{Z}$.
- (b) $f(a \cdot x + b \cdot y) = a \cdot f(x) + b \cdot f(y)$ for $x, y \in R$ and $a, b \in \mathbb{Z}$.

- (c) For any polynomial $P(X) = a_0 + a_1X + \cdots + a_nX^n$ with integer coefficients $a_0, \dots, a_n \in \mathbb{Z}$,

$$f(P(x)) = P(f(x)), \quad (\forall x \in R).$$

In, other words

$$f(a_0 + a_1 \cdot x + \cdots + a_n \cdot x^n) = a_0 + a_1 \cdot f(x) + \cdots + a_n \cdot f(x)^n, \quad (\forall x \in R).$$

3.3.2. Homomorphisms preserve (some) properties of rings and their elements.

Lemma. Let $f : R \rightarrow S$ be a homomorphism of rings.

(a) If an element r of R is killed by a polynomial $P(X) = a_0 + a_1X + \cdots + a_nX^n$ with integer coefficients $a_0, \dots, a_n \in \mathbb{Z}$, the same is true for the image $f(r) \in S$, i.e.,

$$a_0 + a_1fr + \cdots + a_nr^n = 0 \Rightarrow a_0 + a_1f(r) + \cdots + a_nf(r)^n = 0.$$

If we denote by $Sol_R(P)$ the set of all solutions of $P(X) = 0$ in the ring R , then the claim is that $f : R \rightarrow S$ sends $Sol_R(P)$ to $Sol_S(P)$.

(b) If an element r of R is killed by an integer a (in the sense of $a \cdot r = 0$), then $f(r) \in S$ is also killed by a (i.e., $a \cdot f(r) = 0$).

(c) If $r \in R$ is invertible in R then $f(r)$ is invertible in S and $f(r)^{-1} = f(r^{-1})$.

(d) If elements u, v of R commute then $f(u)$ and $f(v)$ commute in S .

We will also state this in terms of preservation of properties of rings:

Corollary. Let $f : R \rightarrow S$ be a homomorphism of rings.

(a) If $P(X) = 0$ has a solution in R then it also has a solution in S .

(b) If an integer a kills all elements of R then it also kills all elements of the subring $f(R) \subseteq S$.

(c) If R is a field then the subring $f(R) \subseteq S$ is also a field.

(d) If R is commutative then the subring $f(R) \subseteq S$ is also commutative.

3.3.3. Non-isomorphic rings. Let A and B be two rings. If A and B were isomorphic then “everything would work the same” in A and B . So, if we want to show that A and B are *not isomorphic*, we need to find some difference between R and S , i.e., some property \mathcal{P} that holds for one of them but not for the other.

Here are some examples of properties \mathcal{P} of a ring that one could use that two rings are not isomorphic: (i) R is commutative, (ii) R has 7 elements, (iii) R is (in)finite, (iv) equation $7x = 0$ has 7 solutions in R , (v) equation $X^7 - 1 = 0$ has 7 solutions in R , (vi) etc.

3.4. Relations between different ideals.

3.4.1. *Ideal $\langle Z \rangle$ generated by a subset Z of a ring.* For subsets X, Y of a ring R we denote

- (i) $X + Y \stackrel{\text{def}}{=} \{x + y; x \in X \text{ and } y \in Y\}$,
- (ii) $XY \stackrel{\text{def}}{=} \{xy; x \in X \text{ and } y \in Y\}$,
- (iii) $X * Y$ is the subset of R consisting of all *finite sums of products of elements from X and from Y* , i.e.,

$$X * Y \stackrel{\text{def}}{=} \{x_1 y_1 + \cdots + x_p y_p; p \in \mathbb{N}, x_i \in X \text{ and } y_i \in Y\}.$$

Remark. Actually, most of then notation (ii) is not used and then our $X * Y$ is denoted simply by $X \cdot Y$ or XY .

Lemma. If I_σ , $\sigma \in \mathcal{S}$, is a family of ideals in a ring R then

- (1) The intersection $\bigcap_{\sigma \in \mathcal{S}} I_\sigma$ is an ideal in R .
- (2) Let us define the sum $\sum_{\sigma \in \mathcal{S}} I_\sigma$ as the set of all finite sums $\sum_{\sigma \in \mathcal{S}} x_\sigma$ where each summand x_σ lies in the corresponding ideal I_σ . (Finiteness of the sum means that we ask that for all but finitely many indices the summand x_σ is zero.) Then the sum $\sum_{\sigma \in \mathcal{S}} I_\sigma$ is again an ideal in R .

Remark. Notice that if the family of ideals has two elements $I_1 = I$ and $I_2 = J$ then $\sum_{\sigma \in \{1,2\}} I_\sigma$ is just the sum $I + J$ defined above.

Corollary. (a) For any $a \in R$, the set $R * a * R$ is an ideal. It consists of all finite sums $\sum_1^n x_i a y_i$ with x_i, y_i in R .

(b) For any subset Z of R there exists the least ideal that contains Z . We denote it by $\langle Z \rangle$. It can be described either as

- (1) intersection of all ideals that contain Z or
- (2) the sum of all ideals $R * z * R$ for $z \in Z$.

Remark. We say that $\langle Z \rangle$ is the ideal generated by Z . For instance The ideal generated by one element $a \in R$ is $\langle a \rangle \stackrel{\text{def}}{=} \langle \{a\} \rangle = R * a * R$.

3.4.2. *Ideals in commutative rings.*

Remark. Ideals of the form $I = aA$ for some element $a \in A$, are called *principal* ideals. One also often denotes aA by (a) . For any $a_1, \dots, a_n \in A$ the ideal R .

Lemma. Let A be a commutative ring.

(a) For any $b \in A$, the subset $bA \stackrel{\text{def}}{=} \{bx; x \in A\}$ is an ideal. (This is the same as the ideal $\langle b \rangle \stackrel{\text{def}}{=} \langle \{b\} \rangle = A * b * A$ generated by a .)

(b) For any ideals I, J in A , the subset $I * J$ of A (consisting of all finite sums of products of elements from I and from J) is again an ideal in A .

(c) For any ideals I, J in A ,

$$I * J \subseteq I \cap J \subseteq I \subseteq I + J.$$

Theorem. In the commutative ring \mathbb{Z}

(1) All ideals are principal, i.e., of the form $(n) = n\mathbb{Z}$ for some $n \in \mathbb{N}$.

(2) $n\mathbb{Z} * m\mathbb{Z} = nm\mathbb{Z}$.

(3) Denote by $lcd(m, n)$ the least common divisor of m, n , then

$$n\mathbb{Z} \cap m\mathbb{Z} = lcd(n, m) \cdot \mathbb{Z}.$$

(4) Denote by $gcd(m, n)$ the greatest common divisor of m, n , then

$$n\mathbb{Z} + m\mathbb{Z} = gcd(n, m) \cdot \mathbb{Z}.$$

3.4.3. *Relation of ideals and quotients.* Recall that for any ideal I of a ring R the quotient map

$$p_I : R \rightarrow R/I, \quad p_I(r) \stackrel{\text{def}}{=} r + I \text{ for } r \in R;$$

is a morphism of rings.

Proposition. Let I, J be ideals in R such that $I \subseteq J$.

(a) There is a canonical morphism of rings

$$p_J^I : R/I \rightarrow R/J, \quad p_J^I(r + I) \stackrel{\text{def}}{=} r + J \text{ for } r \in R.$$

(b) $I = \{0_R\}$ is an ideal in R called the *zero ideal*. (It is often denoted simply by 0 .) The quotient $R/0$ is just R itself.

(c) Map $p_I : R \rightarrow R/I$ is a special case $p_I^0 : R/0 \rightarrow R/I$ of the general construction p_J^I .

(d) If $I \subseteq J \subseteq K$ then

$$p_K^J \circ p_J^I = p_K^I, \text{ i.e., the following diagram commutes: } \begin{array}{ccc} R/I & \xrightarrow{=} & R/I \\ p_J^I \downarrow & & p_K^I \downarrow \\ R/J & \xrightarrow{p_K^J} & R/K \end{array} .$$

In particular, p_J^I is compatible with the quotient maps from R to R/I and R/J in the sense that $p_J^I \circ p_I = p_J$.

Lemma. (a) Map $p_J^I : R/I \rightarrow R/J$ is surjective.

(b) Its kernel $\text{Ker}(p_J^I)$ consists of all cosets $j + I$ in R/I such that $j \in J$. We denote it by J/I .

Theorem. [The second isomorphism theorem.] Let I be an ideal in a ring R . Then:

(a) Any ideal J that contains I gives an ideal $I/J \stackrel{\text{def}}{=} \{j + I; j \in J\}$ in the quotient ring R/I .

(b) There is a canonical isomorphism of rings

$$\phi : (R/I)/(J/I) \rightarrow R/J.$$

For any $r \in R$ isomorphism ϕ sends the coset $(r + I) + I/J \in (R/I)/(J/I)$ to the coset $r + J \in R/J$.

3.4.4. Relation of ideals in R/I and in R .

Lemma. Let I be an ideal in a ring R and let $q : R \rightarrow R/I$ be the canonical quotient map.

(a) Any ideal J in R which is larger than I , i.e., $J \supseteq I$, gives rise to an ideal J/I in the quotient ring R/I . Here $J/I \subseteq R/I$ consists of all cosets in R/I with representative in J :

$$J/I \stackrel{\text{def}}{=} \{j + I; j \in J\}.$$

(b) For any ideal K in R/I , its inverse in R

$$q^{-1}K \stackrel{\text{def}}{=} \{r \in R; q(r) \in K\} = \subseteq \{r \in R; r + I \in K\} \subseteq R,$$

is an ideal in R .

(c) The procedures in (a) and (b) give two inverse bijections between

- ideals K in R/I and
- ideals J in R that contain I .

3.5. Maximal ideals and prime ideals. We consider ideals in a commutative ring A . We say that an ideal $I \subseteq A$ is

- *proper* if $I \neq A$;
- *principal* if it is generated by one element, i.e., if there exist some $b \in A$ such that $I = bA$, i.e., I consists of all multiples of b ;
- *maximal* if it is proper and and it is maximal among proper ideals, i.e., the only proper ideal that J that contains I is $J = I$.
- *prime* if whenever I contains some product ab of elements $a, b \in A$, one of the factors a, b lies in I .

3.5.1. *Principal ideal domains (PID)*. We say that a commutative ring A is a principal ideal domain if

- (i) A is an integral domain, (ii) any ideal in A is principal.

Examples. (a) \mathbb{Z} is a principal ideal domain (PID) since we know that any ideal I in \mathbb{Z} is of the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$.

(b) Later we will see that the ring of polynomials in one variable is a PID but the ring of polynomials in two variables is not a PID.

3.5.2. *Maximal and prime ideals in \mathbb{Z}* . In \mathbb{Z} any ideal is of principal, i.e., of the form $(n) = n\mathbb{Z}$ for some integer n . Moreover, such n can be chosen in \mathbb{N} (since $(-n)\mathbb{Z} = n\mathbb{Z}$), and actually the map

$$\mathbb{N} \ni n \mapsto n\mathbb{Z} \in \text{Ideals in } \mathbb{Z}$$

is a bijection.

Therefore, the relevant question is: for which $n \in \mathbb{N}$ is the ideal $n\mathbb{Z}$ maximal and for which n is it a prime ideal? the ideal $n\mathbb{Z}$ maximal

Lemma. Let $n \in \mathbb{N}$.

- (a) Ideal $n\mathbb{Z}$ is maximal iff n is a prime,
- (b) Ideal $n\mathbb{Z}$ is prime iff n is a prime or $n = 0$.

Proof.

Remark. In this example all maximal ideals are prime but notice that there are more prime ideals besides maximal ones.

3.5.3. *Example: zero ideal.*

Theorem. Consider the zero ideal $\{0\} \subseteq A$.

- (a) $\{0\}$ is a maximal ideal in A iff A/I is a field.
- (b) $\{0\}$ is a prime ideal in A iff A is an integral domain.

3.5.4. *Quotients by maximal and prime ideals.*

Theorem. (a) A proper ideal $I \subseteq A$ is maximal iff A/I is a field.

(b) A proper ideal $I \subseteq A$ is prime iff A/I is an integral domain.

Proof. Use lemmas 3.5.3 and 3.4.4.

Corollary. Any maximal ideal is a prime ideal.

Remark. We noticed above that the converse is not true for the ring \mathbb{Z} .

3.6. Calculation of $\text{Hom}(R, S)$. We denote by $\text{Hom}(R, S)$ the set of all homomorphism $f : R \rightarrow S$ from the ring R to the ring S .

Finding all homomorphisms between rings is often quite useful. This problem is the beginning of *Category Theory*, a way of thinking in mathematics where relations primary interest is not so much in mathematical objects but in relations between such objects.

Lemma. (a) For any ring R there is precisely one homomorphism $\phi : \mathbb{Z} \rightarrow R$. It takes $k \in \mathbb{Z}$ to the multiple $k \cdot 1_R$ of unity in R .

(b) For any n and any ring R , there is at most one homomorphism $\psi : \mathbb{Z}_n \rightarrow R$. It exists iff n kills the unity in R : $n \cdot 1_R = 0$. Then it takes $k \in \mathbb{Z}_n$ to the multiple $k \cdot 1_R$ of unity in R .

Corollary. For given n and m , there is at most one homomorphism

$$\phi : \mathbb{Z}_m \rightarrow \mathbb{Z}_n.$$

It exists iff m is a multiple of n , equivalently, iff the ideal $m\mathbb{Z}$ is contained in the ideal $n\mathbb{Z}$. Then

$$\phi(k) = r_n(k), \quad k \in \mathbb{Z}_m.$$

4. Rings of polynomials

4.0.1. Polynomial functions on real numbers. A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is a *polynomial* if there are numbers $a_0, \dots, a_n \in \mathbb{R}$ such that for all $x \in \mathbb{R}$ one has

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n.$$

Then the function f is denoted

$$f = a_0 + a_1X + a_2X^2 + \dots + a_nX^n = \sum_{i=0}^n a_iX^i.$$

(We can also denote it by $\sum_{i=0}^{\infty} a_iX^i$ if we put $a_i = 0$ for $i > n$.) Let us denote by \mathcal{P} the set of all polynomial functions on \mathbb{R} .

Here is some notation for sets and in particular sets of functions:

- For a set A we denote by $|A| \in \mathbb{N} \cup \{\infty\}$ the number of elements of A .
- For sets A, B we denote by A^B the set of all functions $f : B \rightarrow A$, from B to A .

Remark. The origin of this notation is the observation that if A and B are finite then the number of functions from B to A is $|B|^{|A|}$ (for each element a of A there are $|B|$ choices for (a)). So the new notation satisfies

$$|A^B| = |A|^{|B|}.$$

Lemma. (a) If R is a ring, for any set X the set R^X of functions from X to R has a natural structure of a ring, the sum and product of functions are defined *pointwise*, i.e.,

$$(f + g)(x) \stackrel{\text{def}}{=} f(x) + g(x) \quad \text{and} \quad (f \cdot g)(x) \stackrel{\text{def}}{=} f(x) \cdot g(x) \quad \text{for } x \in X \quad \text{and} \quad f, g \in R^X.$$

(b) The set \mathcal{P} of polynomial functions is a subring of the ring $\mathbb{R}^{\mathbb{R}}$ of functions from \mathbb{R} to \mathbb{R} .

Proof. (a) The needed properties of operations on functions all follow from the corresponding properties of operations in R , say $g + f = f + g$ since for any $x \in X$

$$(g + f)(x) = g(x) + f(x) = f(x) + g(x) = (f + g)(x).$$

Also, zero and unity in \mathbb{R}^X are constant functions $0_{R^X}(x) = 0_R$ and $1_{R^X}(x) = 1_R$ for all $x \in X$.

(b) If functions $f, g \in \mathbb{R}^{\mathbb{R}}$ are polynomial then we can denote them $f = \sum_i f_i X^i$ and $g = \sum_i g_i X^i$. Then, for any $a \in \mathbb{R}$

$$(f+g)(a) \stackrel{\text{def}}{=} f(a)+g(a) = [f_0+f_1a+f_2a^2+\dots] + [g_0+g_1a+g_2a^2+\dots] = (f_0+g_0)+(f_1+g_1)a+(f_2+g_2)a^2+\dots$$

So, $f + g = \sum (f_i + g_i)X^i$ so it is again a polynomial.

Similarly, , for any $a \in \mathbb{R}$

$$(f \cdot g)(a) \stackrel{\text{def}}{=} f(a) \cdot g(a) = [f_0+f_1a+f_2a^2+\dots] \cdot [g_0+g_1a+g_2a^2+\dots] = (f_0g_0)+a(f_0g_1+f_1g_0)+a^2(f_0g_2+f_1g_1+f_2g_0)+\dots$$

So, $f \cdot g = \sum (\sum_{i,j} \text{ with } i+j=n f_i g_j)X^i$ and therefore it is again a polynomial.

4.0.2. *Ring $R[X]$ of polynomials with coefficients in a ring R .* The idea is that just as one has a ring of polynomials with coefficients in \mathbb{R} , for any ring R there is a ring $R[X]$ of polynomials with coefficients in \mathbb{R} , Here, polynomials are defined as expressions

$$A = a_0 + a_1X + a_2X^2 + \dots + a_nX^n \quad \text{with} \quad a_i \in R.$$

We can write it as $A = \sum_{i=0}^n a_i X^i$.⁽¹⁾ If we do not need to keep in mind what is the highest power that occurs in A we can write it as

$$A = \sum_{i=0}^{\infty} a_i X^i = \sum_{i \in \mathbb{N}} a_i X^i$$

where we put $a_i = 0$ for $i > n$.

¹It does not matter how we name the index, i.e., $\sum_{i=0}^n a_i X^i$ and $\sum_{j=0}^n a_j X^j$ both mean the sum $a_0 + a_1X + a_2X^2 + \dots + a_nX^n$. We will use this freedom of choice of naming indices in computations.

The ring structure is given by operations

$$(a_0 + a_1X + a_2X^2 + \cdots) + (b_0 + b_1X + b_2X^2 + \cdots) \stackrel{\text{def}}{=} (a_0 + b_0) + (a_1 + b_1)X + (a_2 + b_2)X^2 + \cdots$$

and

$$(a_0 + a_1X + a_2X^2 + \cdots) \cdot (b_0 + b_1X + b_2X^2 + \cdots) \stackrel{\text{def}}{=} a_0b_0 + (a_0b_1 + a_1b_0)X + (a_2b_0 + a_1b_1 + a_0b_2)X^2 + (a_3b_0 + a_2b_1 + a_1b_2 + a_0b_3)X^3 + \cdots$$

With the \sum notation this is more compact:

$$\left(\sum_{i=0}^{\infty} a_iX^i\right) + \left(\sum_{i=0}^{\infty} b_iX^i\right) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} (a_i + b_i)X^i \quad \text{and} \quad \left(\sum_{i=0}^{\infty} a_iX^i\right) \cdot \left(\sum_{j=0}^{\infty} b_jX^j\right) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \left[\sum_{i+j=n} a_ib_j\right] X^n.$$

Now we should prove

Theorem. $(R[X], +, \cdot)$ is a ring. If R is commutative, so is $R[X]$.

4.0.3. *Formal power series.* The ring of formal power series $R[[X]]$ with coefficients in R , is defined the same as the ring of polynomials $R[X]$, except that we allow that all coefficients be non-zero. So, elements are expressions

$$A = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n + \cdots = \sum_0^{\infty} a_iX^i \quad \text{with} \quad a_i \in R, \quad i \in \mathbb{N}.$$

The operations are then defined by exactly the same formulas as above.

Remarks. (0) The word “formal” indicates that we are not asking this series to converge in any sense.

(1) At the first glance, $R[[X]]$ is more scary than $R[X]$ because it deals with infinitely many summands. However, actually $R[[X]]$ is in some sense simpler than $R[X]$ since we omit the additional requirement that only finitely many coefficients are not zero so there are fewer things to check.

Again we need to prove that

Theorem. $(R[[X]], +, \cdot)$ is a ring. If R is commutative, so is $R[[X]]$.

4.0.4. *What is meant by “expression”?* In order to prove that $R[X]$ or $R[[X]]$ is a ring, we may need to be sure of: what are the objects that we are dealing with, i.e., what is meant here by “expression”?

The information contained in $\sum_{i=0}^n a_iX^i$ is the same as a sequence (a_0, a_1, \dots) of elements of R . So, “expression” $a_0 + a_1X + a_2X^2 + \cdots + a_nX^n + \cdots$ is really just a symbol which packages the sequence (a_0, a_1, \dots) in a *suggestive* way. The suggestion is that these symbols should add up and multiply the same as the polynomial functions (or power series) of real numbers: the addition is defined so that the coefficients of the same powers add up and

multiplication is defined so that each term is multiplied with each term and the powers add up.

If we would want to restate the theorems above simply in terms of sequences then they would say

Theorem. (a) For a ring R let \mathcal{S} be the set of all infinite sequences $a = (a_0, a_1, a_2, \dots)$ of elements of R . Define operations $+$, \cdot on \mathcal{S} by

$$(a + b)_n = a_n + b_n \quad \text{and} \quad (a \cdot b)_n = \sum_{i+j=n} a_i b_j,$$

i.e.,

$$(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$$

$$(a_0, a_1, a_2, \dots) \cdot (b_0, b_1, b_2, \dots) = (a_0 b_0, a_0 b_1 + a_1 b_0, a_2 b_0 + a_1 b_1 + a_0 b_2, a_3 b_0 + a_2 b_1 + a_1 b_2 + a_0 b_3, \dots).$$

Then \mathcal{S} is a ring.

(b) The subset $\mathcal{S}_f \subseteq \mathcal{S}$ of all sequences $a = (a_0, a_1, a_2, \dots)$ with only finitely many non-zero terms, is a subring.

(c) If R is commutative, so are \mathcal{S} and \mathcal{S}_f .

4.0.5. *Rings of polynomials in several variables.* The ring $R[X_1, \dots, X_n]$ of polynomials in variables X_1, \dots, X_n with coefficients in a ring R , is defined similarly. Elements are expressions

$$A = a_{0, \dots, 0} + a_{1, 0, \dots, 0} X_1 + a_{0, 1, \dots, 0} X_2 + \dots + a_{0, \dots, 0, 1} X_n$$

$$+ a_{2, 0, \dots, 0} X_1^2 + a_{0, 2, \dots, 0} X_2^2 + \dots + a_{0, \dots, 0, 2} X_n^2$$

$$+ a_{1, 1, 0, \dots, 0} X_1 X_2 + a_{1, 0, 1, \dots, 0} X_1 X_3 + \dots + a_{0, \dots, 0, 1, 1} X_{n_1} X_n + \dots$$

with all coefficients in R and only finitely many non-zero coefficients. Fortunately, the sum notation is much simpler. The indices (i_1, \dots, i_n) are n -tuples of natural numbers:

$$A = \sum_{(i_1, \dots, i_n) \in \mathbb{N}^n} a_{(i_1, \dots, i_n)} X_1^{i_1} X_2^{i_2} \dots X_n^{i_n}.$$

If we moreover denote for index $I = (i_1, \dots, i_n)$ the monomial $X_1^{i_1} X_2^{i_2} \dots X_n^{i_n}$ simply by X^I , the notation becomes very much alike the notation for polynomials of one variable

$$A = \sum_{I \in \mathbb{N}^n} a_I X^I.$$

This simplicity of notation extends to operations:

$$\left(\sum_{I \in \mathbb{N}^n} a_I X^I \right) + \left(\sum_{I \in \mathbb{N}^n} b_I X^I \right) \stackrel{\text{def}}{=} \sum_{I \in \mathbb{N}^n} (a_I + b_I) X^I$$

and

$$\left(\sum_{I \in \mathbb{N}^n} a_I X^I \right) \cdot \left(\sum_{J \in \mathbb{N}^n} b_J X^J \right) \stackrel{\text{def}}{=} \sum_{K \in \mathbb{N}^n} \left[\sum_{I+J=K} a_I b_J \right] X^K.$$

Here, indices add up componentwise: $I+K = (i_1, \dots, i_n) + (k_1, \dots, k_n) \stackrel{\text{def}}{=} (i_1+k_1, \dots, i_n+k_n)$.

Theorem. $(R[X_1, \dots, X_n], +, \cdot)$ is a ring. If R is commutative, so is $R[X_1, \dots, X_n]$.

Proof. Because of a convenient choice of notation the proof for n variables is going to be the same as for one variable. The difference in proofs appears only in the last step (5) bellow, where case $n = 1$ is slightly simpler.

(1) $(R[X_1, \dots, X_n], +)$ is an abelian group. The proof is obvious because addition in $R[X_1, \dots, X_n]$ is index-wise, so it consists of many copies of addition in R .

For instance,

$$(A+B)+C = \left[\sum_I (a_I + b_I)X^I \right] + \sum_I c_I X^I = \sum_I [(a_I + b_I) + c_I]X^I$$

and similarly, $A+(B+C) = \sum_I [a_I + (b_I + c_I)]X^I$. So, associativity follows from associativity in R . The neutral element is $0_{R[X_1, \dots, X_n]} \stackrel{\text{def}}{=} \sum_I 0 \cdot X^I$ etc.

(2) Associativity of multiplication is interesting because the multiplication in $R[X_1, \dots, X_n]$ is something new: a smart combination of many additions and multiplications in R .

$$\begin{aligned} (A \cdot B) \cdot C &= \left[\sum_P \left(\sum_{I+J=P} a_I b_J \right) X^P \right] \cdot \sum_K c_K X^K \\ &= \sum_Q \left[\sum_{P+K=Q} \left(\sum_{I+J=P} a_I b_J \right) \cdot c_K \right] X^Q \\ &= \sum_Q \left[\sum_{P+K=Q} \sum_{I+J=P} (a_I b_J) c_K \right] X^Q \\ &= \sum_Q \left[\sum_{I+J+K=Q} (a_I b_J) c_K \right] X^Q. \end{aligned}$$

Similarly, $(AB)C = \sum_Q \left[\sum_{I+J+K=Q} a_I (b_J c_K) \right] X^Q$. So, associativity of multiplication in polynomials follows from several properties of operations in R .

(3) Distributivity.

$$\begin{aligned} (A+B) \cdot C &= \left[\sum_I (a_I + b_I)X^I \right] \cdot \sum_J c_J X^J \\ &= \sum_K \left[\sum_{I+J=K} (a_I + b_I) \cdot c_J \right] X^K = \sum_K \left[\sum_{I+J=K} a_I c_J + b_I c_J \right] X^K \\ &= \sum_K \left[\sum_{I+J=K} a_I c_J \right] X^K + \sum_K \left[\sum_{I+J=K} b_I c_J \right] X^K = AC + BC. \end{aligned}$$

(4) If R is commutative so is $R[X_1, \dots, X_n]$ as

$$AB = \sum_K \left(\sum_{I+J=K} a_I b_J \right) X^K \quad \text{and} \quad BA = \sum_K \left(\sum_{I+J=K} b_J a_I \right) X^K.$$

(5) The condition of “only finitely many coefficients are not zero” is preserved by addition and multiplication operations in $R[X_1, \dots, X_n]$.

For index $I = (i_1, \dots, i_n)$ we introduce a notion of size by $|I| = \max\{i_1, i_2, \dots, i_n\}$. Notice that $|I + J| \leq |I| + |J|$.

Let us say that the degree $\deg(A)$ of $A = \sum_I a_I X^I$ is

- if there are only finitely many non-zero coefficients this is the largest value of $|I|$ among all I such that $a_I \neq 0$;
- if there are infinitely many non-zero coefficients then $\deg(A) = \infty$.

So, A satisfies the condition of “only finitely many coefficients are not zero” iff $\deg(A) < \infty$.

Then it is easy to see that

$$\deg(A + B) \leq \max[\deg(A), \deg(B)] \quad \text{and} \quad \deg(AB) \leq \deg(A) + \deg(B).$$

(Here we define $\max(x, \infty) = \infty = x + \infty$ for any $x \leq \infty$.)

This implies the required finiteness property for $A + B$ and AB .

4.0.6. *Central subrings.* We say that a subring A of a ring R is central if it lies in the center of R , i.e., $A \subseteq Z[R]$.

Clearly any central subring is commutative.

Examples. (0) Of course the center $Z[R]$ is a central subring of R .

(1) For the ring of $n \times n$ matrices with real coefficients $M_n(\mathbb{R})$, the center $Z[M_n(\mathbb{R})]$ consists of all *scalar* matrices, i.e., matrices that are multiples $r \cdot I_n$, $r \in \mathbb{R}$, of the unity matrix 1_n . ($r \cdot I_n$ has all diagonal elements r and all other elements vanish.)

Moreover, $i : \mathbb{R} \rightarrow Z[M_n(\mathbb{R})]$ defined by $i(r) \stackrel{\text{def}}{=} r \cdot I_n$ is an isomorphism of rings. So we say (imprecisely) that the center of $M_n(\mathbb{R})$ is the ring \mathbb{R} .

4.0.7. *A-algebras.* There is an important notion, however we will not really need it. It is a slight generalization of the notion of central subrings.

Let A be a commutative ring. We say that a ring R is an *A-algebra* if we are given a homomorphism of rings from A to the center of R :

$$\iota : A \rightarrow Z[R].$$

Example. Central subrings $A \subseteq Z[R]$ are a special case of *A-algebra* when the homomorphism i is inclusion (or at least i is injective).

4.0.8. *Evaluation.* Let A be a commutative ring which is a central subring of a ring R . Any element r in R defines the *evaluation at r* function

$$ev_r : A[X] \rightarrow R$$

which sends a polynomial $P = \sum_i p_i X^i \in A[X]$ to

$$ev_r\left(\sum_i p_i X^i\right) \stackrel{\text{def}}{=} P(r) \stackrel{\text{def}}{=} \sum_i p_i r^i.$$

The result is an element of R obtained by taking powers of r , multiplying them by elements p_i of the subring $A \subseteq R$ and then adding these terms in R .

Lemma. Evaluation at r is a morphism of rings

$$ev_r : A[X] \rightarrow R.$$

In other words for any $P, Q \in A[X]$ one has $ev_r(P+Q) = ev_r(P) + ev_r(Q)$ and $ev_r(P \cdot Q) = ev_r(P) \cdot ev_r(Q)$, i.e.,

$$(P+Q)(r) = P(r) + Q(r) \quad \text{and} \quad (P \cdot Q)(r) = P(r) \cdot Q(r).$$

Proof. The claim for addition is clear:

$$\begin{aligned} ev_r(P+Q) &= ev_r\left(\sum_i (p_i + q_i) X^i\right) \\ &= \sum_i (p_i + q_i) r^i = \sum_i p_i r^i + \sum_i q_i r^i = \sum_i p_i r^i + \sum_i q_i r^i \\ &= ev_r(P) + ev_r(Q). \end{aligned}$$

For multiplication one uses the fact that an element r of R always commutes with coefficients $p_i, q_i \in A$, i.e., that A is central in R .

$$\begin{aligned} ev_r(PQ) &= ev_r\left(\sum_n \left(\sum_{i+j=n} p_i q_j\right) X^n\right) \\ &= \sum_n \left(\sum_{i+j=n} p_i q_j\right) r^n = \sum_{i,j \in \mathbb{N}} p_i q_j r^i r^j \\ &= \left(\sum_{i \in \mathbb{N}} p_i r^i\right) \cdot \left(\sum_{j \in \mathbb{N}} q_j r^j\right) = ev_r(P) \cdot ev_r(Q). \end{aligned}$$

4.0.9. *Constructing complex numbers \mathbb{C} from real numbers \mathbb{R} by using polynomials.* Recall that any element b of a commutative ring A defines an ideal $bA = \{ba; a \in A\}$ in A . We denote such ideals by $(b) \stackrel{\text{def}}{=} bA$ and we call them principal ideals.

Theorem. The quotient of the ring of polynomials $\mathbb{R}[X]$ by the principal ideal $(X^2 + 1) \stackrel{\text{def}}{=} (X^2 + 1)\mathbb{R}[X]$, is isomorphic to the ring of complex numbers:

$$\mathbb{R}[X]/(X^2 + 1) \cong \mathbb{C}.$$

Proof. We need to construct an isomorphism $F : \mathbb{R}[X]/(X^2 + 1) \cong \mathbb{C}$.

(0) Since \mathbb{R} is a central subring of \mathbb{C} , element $i \in \mathbb{C}$ defines evaluation homomorphism

$$\phi \stackrel{\text{def}}{=} ev_i : \mathbb{R}[X] \rightarrow \mathbb{C}, \phi(A) = A(i) \text{ for } A \in \mathbb{R}[X].$$

This will be our starting point.

(1) Any homomorphism of rings defines an isomorphism of rings, so $\phi : \mathbb{R}[X] \rightarrow \mathbb{C}$ gives isomorphism

$$\bar{\phi} : \mathbb{R}[X]/\text{Ker}(\phi) \xrightarrow{\cong} \text{Im}(\phi) = \phi(\mathbb{R}[X]).$$

We will check that the image of ϕ is all of \mathbb{C} and that the kernel is the principal ideal $(X^2 + 1)$. Then $F = \bar{\phi}$ will be the isomorphism that we are looking for.

(2) *The image of ϕ is all of \mathbb{C} .* For any complex number z we have $z = a + bi$. Consider a polynomial $a + bX$, then

$$\phi(a + bX) = ev_i(a + bX) = a + bi = z.$$

(3) *The kernel of ϕ is the principal ideal $(X^2 + 1)$.* For this we will calculate directly the kernel $\text{Ker}(\phi)$ and the ideal $(X^2 + 1)$.

(3a) Polynomial $A = a_0 + a_1X + a_2X^2 + \dots$ is in $\text{Ker}(\phi)$ if

$$\phi(A) = A(i) = a_0 + a_1i + a_2i^2 + a_3i^3 + a_4i^4 + \dots = (a_0 - a_2 + a_4 - a_6 + \dots) + i(a_1i - a_3 + a_5 - a_7 + \dots)$$

is zero, i.e., if the alternating sum of even coefficients is zero and the alternating sum of odd coefficients is zero.

(3b) Polynomial $A = a_0 + a_1X + a_2X^2 + \dots$ is in $(X^2 + 1) = (X^2 + 1)\mathbb{R}[X]$ if it can be written as a product $A = (X^2 + 1)T$ for some polynomial $T = t_0 + t_1X + t_2X^2 + \dots$. Such product are polynomials

$$(X^2 + 1)(t_0 + t_1X + t_2X^2 + \dots) = (t_0X^2 + t_1X^3 + t_2X^4 + \dots) + (t_0 + t_1X + t_2X^2 + \dots) = t_0 + X t_1 + X^2(t_0 + t_2) + \dots$$

So, A is in the principal ideal if one can find solutions t_0, t_1, t_2, \dots of a system of equations

- $a_0 = t_0$ and $a_1 = t_1$,
- $a_2 = t_0 + t_2$,
- $a_3 = t_1 + t_3$,
- $a_4 = t_2 + t_4$, etc.

(3c) Surprisingly, there is always a unique solution:

- $t_0 = a_0$ and $t_1 = a_1$,

- $t_2 = a_2 - t_0 = a_2 - a_0$,
- $t_3 = a_3 - t_1 = a_3 - a_1$,
- $t_4 = a_4 - t_2 = a_4 - (a_2 - a_0) = a_4 - a_2 + a_0$.

In general,

$$t_n = a_n - a_{n-2} + a_{n-4} - a_{n-6} + \cdots$$

(3d) So it seems that we have proved that any polynomial A is a multiple $A = (X^2 + 1)T$ of $X^2 + 1$. However, it is clear that for $A = 1$ there is no polynomial T such that $1 = (X^2 + 1) \cdot T^{(2)}$ Where is the mistake?

The point is that T that we found need not be a polynomial – it may have infinitely many non-zero coefficients and then it will be a formal power series but not a polynomial.

(3e) In order for T to be a polynomial we need

$$t_n = a_n - a_{n-2} + a_{n-4} - a_{n-6} + \cdots$$

to be zero for all sufficiently large n .

What this means is that the alternating sums of even coefficients of A has to vanish and the alternating sums of odd coefficients of A has to vanish. However, these are precisely the conditions for A to be in $\text{Ker}(\phi)$.

So, $(X^2 + 1)\mathbb{R}[X]$ equals $\text{Ker}(\phi)$.

Remark. Part (3) of the proof was long because we understood the situation directly, without any smart tricks. We will redo part (3) when we learn how to divide polynomials. Then the proof will be very simple.

4.1. Division of polynomials.

4.1.1. *Degree of a polynomial.* The degree $\deg(A)$ of the polynomial $A = \sum a_i X^i \in R[X]$ is

$$\deg(A) = \begin{cases} \text{the greatest } i \text{ such that } a_i \neq 0 & \text{if } A \text{ has some coefficient } \neq 0, \text{ i.e., if } A \neq 0, \\ -1 & \text{if } A = 0. \end{cases}$$

(One can say it in one go by using the idea of infimum:

$$\deg(A) = \inf\{i; a_i \neq 0\}.$$

Remark. Notice that $\deg(P) \geq 0$ iff $P \neq 0$.

²Because $\deg[(X^2 + 1)T] = \deg(X^2 + 1) + \deg(T) \geq 2 + 0$, while $\deg(A) = \deg(1) = 0$.

Lemma. (a) For any ring R and any $P, Q \in R[X]$

- (1) $\deg(P + Q) \leq \max[\deg(P), \deg(Q)]$.
- (2) $\deg(P \cdot Q) \leq \deg(P) + \deg(Q)$.

(b) If R has no zero divisors and $P, Q \neq 0$ then more is true:

$$\deg(P \cdot Q) = \deg(P) + \deg(Q).$$

(c) The only situation when

$$\deg(P + Q) \leq \max[\deg(P), \deg(Q)]$$

fails to be equality is when the highest powers cancel in addition, i.e., $\deg(P) = \deg(Q)$ and the leading coefficient of Q is (-1) times the leading coefficient of P .

Corollary. When A is an integral domain so is $A[X]$.

Proof. Integral domain means: commutative and no zero divisors. Since A is commutative so is $A[X]$. It remains to check that $A[X]$ has no zero divisors. So, let $P, Q \in A[X]$ and $P, Q \neq 0$. Then $\deg(P) \geq 0$ and $\deg(Q) \geq 0$, so

$$\deg(P \cdot Q) = \deg(P) + \deg(Q) \geq 0 + 0 = 0.$$

This guarantees that $PQ \neq 0$.

4.1.2. *Division of polynomials.* If a, b are elements of a field F and $b \neq 0$ then we can define the quotient q of a by b by $q \stackrel{\text{def}}{=} ab^{-1}$. (We often denote it $\frac{a}{b}$ as in real numbers.)

Recall that in the ring of integers we also have a process of division. However it is more complicated: from integers $a, b \in \mathbb{Z}$ it produces the quotient q but there is also the remainder r . The crucial properties of q and r are

- $a = bq + r$ and
- the remainder r is strictly lesser than b .

It turns out that a similar notion of division exists in rings of polynomials over a field:

4.1.3. *Theorem.* If F is a field then for any polynomials $A, B \in F[X]$, with $B \neq 0$, there exist unique polynomials Q, R such that

- $A = BQ + R$ and
- $\deg(R) < \deg(B)$.

Actually, more is true. For polynomials over a ring R one can again needed divide A by B provided that B has the property that its leading coefficient is invertible in R :

4.1.4. *Theorem.* If \mathcal{A} is an integral domain field then for any polynomials $A, B \in F[X]$, with $B \neq 0$, there exist unique polynomials Q, R such that

- $A = BQ + R$ and
- $\deg(R) < \deg(B)$.

Example. A polynomial is said to be *monic* if its leading coefficient is 1. So one can always divide by monic polynomials.

4.1.5. *The proof of the theorems.* We will actually state and prove an even more general statement about existence and uniqueness of quotient and remainder:

Theorem. Let \mathcal{R} be a ring and let $A = a_0 + a_1X + \cdots + a_nX^n$ and $B = b_0 + b_1X + \cdots + b_mX^m$ be two polynomials in $\mathcal{R}[X]$.

(a) If $\deg(A) = n \geq \deg(B) = m$ and $c \in R$ is such that $b_m \cdot c = a_n$ then

$$\deg[A - B(cX^{n-m})] < \deg(A).$$

(b) If the leading coefficient of B is invertible in \mathcal{R} then there exist polynomials Q, R in $\mathcal{R}[X]$ such that

- $A = BQ + R$ and
- $\deg(R) < \deg(B)$.

(c) If R is also an integral domain then the polynomials Q and R in part (b) are unique.

Proof. (a) Since $n \geq m$, $q = cX^{n-m}$ is a polynomial. One has

$$\begin{aligned} A - BQ &= (a_0 + a_1X + \cdots + a_nX^n) - (b_0 + b_1X + \cdots + b_mX^m) \cdot cX^{n-m} \\ &= (a_0 + a_1X + \cdots + a_nX^n) - (b_0cX^{n-m} + b_1cX^{n-m+1}X + \cdots + b_m cX^n X). \end{aligned}$$

We have chosen c so that $b_m c = a_n$ and this has the effect of canceling the two appearances of X^n . We are left with powers of X that are $< n$, so $\deg[A - B(cX^{n-m})] < \deg(A)$.

(b) If $\deg(A) < \deg(B)$ then $Q = 0$ and $R = A$ satisfy $BQ + R = B \cdot 0 + A = A$ and $\deg(R) = \deg(A) < \deg(B)$.

So, it remains to consider the case when $\deg(A) \geq \deg(B)$, i.e., $n \geq m$. Since b_m is invertible we can choose $c_0 \in R$ to be $b_m^{-1}a_n$. It satisfies $b_m \cdot c_0 = a_n$. Also, since $n \geq m$, $q_0 = c_0X^{n-m}$ is a polynomial. Let $A_1 = A - Bq_0$, then according to (a) we have $\deg(A_1) < \deg(A)$.

We can now repeat this procedure with A_1 instead of A . This gives q_1 (chosen as above but for A_1 and B instead of A and B) such that $A_2 = A_1 - Bq_1$ satisfies $\deg(A_2) < \deg(A_1)$. And so on:

$$A_1 = A - Bq_0, A_2 = A_1 - Bq_1, A_3 = A_2 - Bq_2, \dots \quad \text{with} \quad \deg(A) > \deg(A_1) > \deg(A_2) > \dots$$

The degrees of A_k 's keep decreasing and eventually we get to some $A_p = A_{p-1}B + q_{p-1}$ with $\deg(A_p) < \deg(B)$.

Now e can not go any further, but we are done since

$$A = Bq_0 + A_1 = Bq_0 + Bq_1 + A_2 = \dots = Bq_0 + Bq_1 + \dots + Bq_{p-1} + A_p = B(q_0 + q_1 + \dots + q_{p-1}) + A_p = BQ + A_p$$

if we take $Q = q_0 + q_1 + \dots + q_{p-1}$ and $R = A_p$.

The important thing is that after applying the procedure (a) several times we have managed to decrease the degree of A_p beneath the degree of B . So, $\deg(R) = \deg(A_p) < \deg(B)$.

(c) If we have two solutions Q_i, R_i for $i = 1, 2$, i.e., $A = BQ_i + R_i$ and $\deg(R_i) < \deg(B)$ for $i = 1, 2$. Now we have $BQ_1 + R_1 = BQ_2 + R_2$, i.e., $B(Q_1 - Q_2) = R_2 - R_1$.

We know that $B \neq 0$, so if also $Q_1 - Q_2 \neq 0$, then since R has no zero divisors we would have

$$\deg[B(Q_1 - Q_2)] = \deg(B) + \deg(Q_1 - Q_2) \geq \deg(B) + 0 = \deg(B).$$

However, this is impossible since

$$\deg[B(Q_1 - Q_2)] = \deg(R_2 - R_1) \leq \max[\deg(R_2), \deg(R_1)] < \deg(B).$$

This implies that $Q_1 - Q_2 = 0$, i.e., $Q_2 = Q_1$ and then we also get $R_2 = A - BQ_2 = A - BQ_1 = R_1$.

4.1.6. *The algorithm for dividing polynomials.* For given polynomials A and B the question is how do you divide A by B , i.e., how do you find quotient Q and remainder R ?

This is actually explained in the proof of the preceding theorem. Q is found as a sum of several terms $Q = q_0 + q_1 + \dots + q_{p-1}$.

- One finds q_0 by dividing the highest term $a_n X^n$ of A by the highest term $b_m X^m$ of B . This gives $q_0 = b_m^{-1} a_n X^{n-m}$. Then one replaces A by $A_1 = A - Bq_0$ (which has lower degree than A).
- Now one repeats the same for A_1 and B , we get q_1 and $A_2 = A_1 - Bq_1$,
- And so on until we get some $A_p = A_{p-1} - Bq_{p-1}$ such that $\deg(A_p) < \deg(B)$.
- Then $R = A_p$ and $Q = q_0 + q_1 + \dots + q_{p-1}$.

Remark. Now we know how to divide polynomials over \mathbb{R} and \mathbb{C} but also over more unusual fields such as $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5, \mathbb{Z}_7, \dots$. We can even divide polynomial $A \in \mathbb{Z}[X]$ by a polynomial $B \in \mathbb{Z}[X]$ over the ring of integers, provided that B is monic.

4.2. Some consequences of division of polynomials.

4.2.1. $F[X]$ is PID.

Lemma. For any field F , the ring of polynomials $F[X]$ is a principal ideal domain, i.e., any ideal is principal.

Proof. Principal ideal domain means that (i) there are no zero divisors and (ii) any ideal is principal. We already know that $F[X]$ satisfies (i) (remember that we proved this using the degrees).

So, let I be an ideal in $F[X]$. If I is the zero ideal $\{0\}$ the claim is true since $\{0\} = 0 \cdot F[X]$ is principal.

If $I \neq \{0\}$ then it contains some polynomial $S \neq 0$. and then $\deg(S) \geq 0$. Now choose $P \in I$ so that $P \neq 0$ and that $\deg(P)$ is the least possible among the degrees of all nonzero elements of I .

We will see that $I = (P)$. Since $P \neq 0$ we can divide any $A \in I$ with P , we get $Q, R \in F[X]$ such that $A = PQ + R$ and $\deg(R) < \deg(P)$.

Then,

4.2.2. Division with polynomials $X - a$.

Lemma. Consider polynomials over a field F and let $\alpha \in F$. When we divide a polynomial $A \in F[X]$ with the remainder is $A(\alpha) \in F$.

Proof. If $A = (X - \alpha)Q + R$ and $\deg(R) < \deg(X - \alpha) = 1$, then $\deg(R)$ can be 0 or -1 , so at any rate R is a constant polynomial, i.e., $R \in F$.

Now we evaluate $A = (X - \alpha)Q + R$ at α and get

$$A(\alpha) = (\alpha - \alpha)Q(\alpha) + R(\alpha) = R(\alpha).$$

Then, since R is a constant we have $R = R(\alpha) = A(\alpha)$.

Corollary. $X - \alpha$ divides polynomial A iff $A(\alpha) = 0$.

Proof. Dividing A by $X - \alpha$ we get $A = (X - \alpha)Q + A(\alpha)$. So, if $A(\alpha) = 0$ then $A = (X - \alpha)Q$ is divisible by $X - \alpha$.

On the other hand, if A is divisible by $X - \alpha$, i.e., $A = (X - \alpha)Q$; then $A(\alpha) = (\alpha - \alpha)Q(\alpha) = 0$.

4.2.3. Example: Division in $\mathbb{R}[X]$ by $X^2 + 1$. We will reprove the following fact:

Lemma. Consider the evaluation homomorphism $ev_i : \mathbb{R}[X] \rightarrow \mathbb{C}$. Then

$$\text{Ker}(ev_i) = (X^2 + 1).$$

Proof. Certainly, $ev_i(X^2 + 1) = i^2 + 1 = 0$ so $X^2 + 1 \in \text{Ker}(ev_i)$. Then all multiples of $X^2 + 1$ lie in the ideal $\text{Ker}(ev_i)$, i.e., $(X^2 + 1) \subseteq \text{Ker}(ev_i)$.

In the opposite direction we will see that any $A \in \text{Ker}(ev_i)$ is a multiple of $X^2 + 1$. For this we divide any A in $\text{Ker}(ev_i)$ by $X^2 + 1$. We get

$$A = (X^2 + 1)Q + R \quad \text{and} \quad \deg(R) < \deg(X^2 + 1) = 2.$$

So, $R = a + bX$ for some $a, b \in \mathbb{R}$. But then

$$R = A - (X^2 + 1)Q \quad \text{hence} \quad R(i) = A(i) - (i^2 + 1)Q(i) = 0.$$

This means that $0 = R(i) = a + bi$. The $a = 0 = b$ and $R = 0$. So, A is a multiple of $X^2 + 1$! a

4.2.4. Example of a non principal ideal. Let F be a field and consider the ring $A = F[X, Y]$ of polynomials in two variables. Let I be the ideal which is a sum of two principal ideals $I \stackrel{\text{def}}{=} (X) + (Y)$. So, I consists of all sums $A + B$ where $A \in (X)$ is a multiple of X and $B \in (Y)$ is a multiple of Y , i.e., I consists of all $X \cdot P + Y \cdot Q$ with $P, Q \in F[X]$.

- (1) A polynomial in two variables

$$R = r_{0,0} + r_{1,0}X + r_{0,1}Y + r_{2,0}X^2 + r_{1,1}XY + r_{0,2}Y^2 + \dots$$

lies in I iff its constant term $r_{0,0}$ vanishes.

Proof. We see that all monomials $X^i Y^j$ lie in I if $i > 0$ or $j > 0$.

- (2) Ideal I in $A = F[X, Y]$ is not a principal ideal. In particular, $A = F[X, Y]$ is not a principal ideal domain.

If I were principal then there would exist some polynomial $R \in I$ such that $I = (R)$, i.e., elements of I are multiples of R . This would mean that since X and Y are in I that they are multiples of R , i.e., $X = RU$ and $Y = RV$ for some polynomials $U, V \in F[X, Y]$.

However, $X = RU$ implies that R and U contain no Y 's. For that we view $F[X, Y]$ as $(F[X])[Y]$ and accordingly we use the degree \deg_Y which means the highest power in Y . Since $X = RU$ we have $R \neq 0$ and $U \neq 0$. Then $0 = \deg_Y(X) = \deg_Y(RU) = \deg_Y(R) + \deg_Y(U)$ implies that $\deg_Y(R) = 0$ and $\deg_Y(U) = 0$. This is just what we claimed: no Y 's appear in R or U .

Similarly, since $Y = RV$, no X 's appear in R or V .

But then R has only the constant term, i.e., $R \in F$. But this is impossible since then $R \neq 0$ would imply that R is invertible and therefore $I = (R) = R \cdot F[X, Y]$ would be the whole ring $F[X, Y]$ and this is not true for our ideal I .

5. Fraction fields

5.1. The problem. The ring of integers \mathbb{Z} is not a field but it lies in a field of rational numbers, i.e., ring \mathbb{Z} is a subring of a field \mathbb{Q} . This allows us to do some computations with integers that we could not do inside integers themselves (i.e., divisions such as $\frac{2}{3}$).

For which rings R is it true that R is a subring of a field? The following are certainly necessary conditions:

- R should be commutative (because fields are commutative),
- R should have no zero divisors, i.e., it should be an integral domain (because fields have no zero divisors).

The two conditions are stated together in the phrase “ R is an integral domain”.

So, now, the question has reduced to the following one: for which integral domains A is it true that A is a subring of a field?

It will turn out that this is true for all integral domains.

Theorem. Any integral domain A is a subring of some field F . Moreover, there is the smallest such field F that contains A . It is called the *fraction field* of A .

The following two sub-subsections deal with minor clarifications of the theorem, they can be skipped in the first reading.

5.1.1. *A stupid clarification on the level of set theory I.* If one has an injective map of sets $i : X \rightarrow Y$ this is in some sense as good as a situation where A is a subset of X . The reason is that if i is injective then X contains a subset which is the image of the map $A' \stackrel{\text{def}}{=} \{i(a); a \in A\} \subseteq X$ (usually denoted $i(A)$), and we know a bijection between A and A' – to $a \in A$ there corresponds an element $i(a)$ of A' . So,

*A' is a subset of X and “everything in A' works the same as in A ”
(because we know a bijection).*

Similarly, if we know an injective homomorphism of rings $i : A \rightarrow F$ with F a field, this is for all practical purposes as good as knowing that A is a subfield of some field.⁽³⁾

5.1.2. *A clarification on the level of set theory II.* Here we are interested just in what it means that there exists the *smallest* field F that contains integral domain A . This is not literally true, the precise meaning is that *smallest* field F that contains A is *unique up to (canonical) isomorphism*. This means the following:

- For a field F that contain A we will say that it is minimal, if any subfield F' of F that contains A has to be all of F .

³If we really want to we can now cook up a field \mathcal{F} that contains A . As a set $\mathcal{F} \stackrel{\text{def}}{=} (F - i(A)) \cup A$, i.e., we take out of the field F all points of the form $i(a)$ with $a \in A$, and we replace them by the corresponding points a in A . Then we have a natural bijection $\iota : \mathcal{F} \rightarrow F$. The point is that $F = (F - i(A)) \cup i(A)$ and $\mathcal{F} \stackrel{\text{def}}{=} (F - i(A)) \cup A$, so we construct ι so that it is identity on the part on $F - i(A)$ that the two sets have in common and we define ι so that relates the remaining parts $A \subseteq \mathcal{F}$ and $i(A) \subseteq F$, so that on $A \subseteq \mathcal{F}$, ι is the same as i . Now we can make \mathcal{F} into a field by transporting via ι the structure of a field that we have on F .

- Now the claim is that any two minimal fields F and G that contain A have to be “the same for all practical purposes” (but they really need not be the same). The precise meaning of the phrase is that there is a canonical way to identify F and G :

Then there is a unique isomorphism $\phi : F \rightarrow G$ which is identity on A .

5.2. The quotient construction of new sets. The idea is that sometimes when looking at some set S it may turn out that from our point of view some elements “perform the same task” so in a sense there are unwanted duplications which we would like to eliminate. The most natural solution is to

- Partition the set S into groups $S = S_1 \cup S_2 \cup \dots$ so that two elements are in the same group iff they “perform the same task”.
- Squeeze each group into a point.

The mathematical meaning of the second step is that

- (1) We replace set S by a new set \mathcal{S} whose elements are not any more the elements of S but the above *groups* of elements of S ! So, $\mathcal{S} = \{S_1, S_2, \dots\}$.
- (2) The relation of the original set S and the new set \mathcal{S} is encoded into a natural map $q : S \rightarrow \mathcal{S}$ which assigns to each $a \in S$ the group that contains a .

The map q sends all elements of a group S_i into a single element of \mathcal{S} , so the map q can be thought of as “squeezing” each group S_i (a subset of S) into a single element of \mathcal{S} . The funny thing is that this element of \mathcal{S} is the group S_i itself.

Remark. An alternative would be to eliminate duplications by choosing from each group S_i one element a_i . Then all a_i ’s form a new set T which is obtained from S by eliminating duplications. This is less abstract however it also less natural because we had to make many choices (possibly infinitely many). This procedure also creates a new burden, whatever we do in the future we will have to remember these (arbitrary) choices and this will get cumbersome in calculations.

It remains to make (mathematical) sense of the above idea that “from a certain point of view elements a and b of S perform the same task”. This is elegantly accomplished by the idea of

5.2.1. Equivalence relations. A relation \sim on a set S is said to be an *equivalence relation* if it satisfies

- (Reflexivity) $(\forall a \in S) a \sim a$;
- (Symmetry) $(\forall a, b \in S) a \sim b \Rightarrow b \sim a$;
- (Transitivity), $(\forall a, b, c \in S) a \sim b \text{ and } b \sim c \Rightarrow a \sim c$.

We will see (first in examples and then in general) that the meaning of any equivalence relation $a \sim b$ is always that “ a and b are the same in a certain sense”, i.e., “in a certain sense a and b perform the same task”.

5.2.2. Examples.

- (1) Any subgroup H of a group G defines an equivalence relation by $a \sim b$ if $ba^{-1} \in H$.
- (2) On the set \mathcal{L} of lines in a plane the relation $L \parallel M$, i.e., L is parallel to M , is an equivalence relation.
- (3) Any function $f : S \rightarrow T$ from S to some set T defines an equivalence relation by:

$$a \sim b \text{ if } f(a) = f(b).$$

- (4) A *partition* of a set S is a family \mathcal{F} of subsets of S such that
 - (a) \mathcal{F} covers S :

$$\cup_{F \in \mathcal{F}} F = S.$$

- (b) This union is disjoint in the sense that different sets F do not have anything in common: if $\overline{F, G} \in \mathcal{F}$ and $F \neq G$ then $F \cap G = \emptyset$.

Lemma. Any partition \mathcal{F} defines an equivalence relation by:

$$a \sim b \text{ if there is some } F \in \mathcal{F} \text{ which contains both } a \text{ and } b.$$

5.2.3. *Equivalence classes.* An equivalence relation \sim on a set S attaches to each element $a \in S$ a subset of S called *the equivalence class of a* , and denoted by $[a]$. It consists of all $x \in S$ such that $a \sim x$.

Lemma. Let \sim be an equivalence relation on a set S .

- (a) For any $a \in A$, the equivalence class $[a]$ contains a .
- (b) For any $a, b \in A$, the equivalence classes $[a]$ and $[b]$ are either the same or disjoint (in other words, if they meet they are the same).
- (c) Equivalence class $[a]$ and $[b]$ are the same iff $a \sim b$.

Examples.

- (1) If H is a subgroup of G and $a \sim b$ if $ba^{-1} \in H$, the equivalence class of $g \in G$ is the coset Hg .
- (2) For relation $L \parallel M$ the equivalence class of L consists of all lines parallel to L . So, the meaning of an equivalence class of planes is that it defines one direction in a plane.
- (3) If $f : S \rightarrow T$ and $a \sim b$ if $f(a) = f(b)$ the equivalence class of a consists of all elements of S with the same image in T .
- (4) For the equivalence relation \sim defined by a partition \mathcal{F} , the equivalence class of a is one of the sets F in \mathcal{F} , the unique one that contains a .

Corollary. The grouping of elements of S into equivalence classes is a *partition* of S , i.e.,

- (1) Equivalence classes cover S :

$$\cup_{a \in S} [a] = S.$$

- (2) This union is disjoint in the sense that different sets $[a]$ do not have anything in common.

5.2.4. *Quotients by equivalence relations.* We denote by S/\sim the set of all equivalence classes in S with respect to the equivalence relation \sim . It comes with a map $q : S \rightarrow S/\sim$ which attaches to each $a \in S$ its equivalence class :

$$q(a) = [a].$$

Examples.

- (1) If H is a subgroup of G and $a \sim b$ if $ba^{-1} \in H$, then G/\sim is the set $H \backslash G = \{Hg, g \in G\}$ of all right cosets of H in G . Recall that if H is a normal subgroup then $G/\sim = H \backslash G$ is a new group!
- (2) For the parallelism relation $L||M$ on lines in a plane, the meaning of the set \mathcal{L}/\sim of equivalence classes is that this is the set of all directions in the plane!
- (3) [**The first isomorphism theorem for sets.**] If we use a map $f : S \rightarrow T$ to define equivalence relation on S by $a \sim b$ if $f(a) = f(b)$, then the map f factors to a bijection

$$\bar{f} : S/\times \rightarrow \text{Im}(f).$$

So, any function f from S to T gives rise to a bijection between a quotient of the set S and a subset of a set T .

- (4) For the equivalence relation \sim defined by a partition \mathcal{F} , the set of equivalence classes S/\sim is the same as the original partition \mathcal{F} .

5.2.5. *Notions equivalent to an equivalence relation.*

Theorem. For a set S , the following notions are equivalent

- (1) Having an equivalence relation \sim on S ;
- (2) Having a partition \mathcal{F} of S ;
- (3) Having a surjective function f from S to some set T .

Proof. We already saw all the ingredients of this theorem:

- (EP) A partition \mathcal{F} of S defines an equivalence relation \sim on S by $a \sim b$ if $(\exists F \in \mathcal{F}) F \ni a, b$.
- (PE) An equivalence relation \sim on S ; defines a partition S/\sim of S .
- (SE) A surjection f from S to some set T defines an equivalence relation \sim on S by $a \sim b$ if $f(a) = f(b)$.
- (PE) An equivalence relation \sim on S ; defines a surjection $q : S \rightarrow S/\sim$.

One could say that this constitutes an equivalence of these notions, but only in a very weak sense: from each of the three kinds of objects we can construct objects of two other kinds.

However, a more useful (and standard) meaning of equivalence of these notions is that when we do this “passing from objects of one kind to another kind” several times in a row, that there are no contradictions.

In our case this means for instance that if we start from a partition \mathcal{F} and pass to an equivalence relation \sim and then in the opposite direction from the equivalence relation \sim to the corresponding partition S/\sim ; that what we get in the end is what we started with, i.e., $S/\sim = \mathcal{F}$. (Actually we already checked that.) A short way to say this is that the composition $EP \circ PE$ of procedures PE and EP is identity. So, what one needs to check (to be sure that there are no contradictions) are the following four statements:

$$EP \circ PE = Id \quad \text{and} \quad PE \circ EP = Id$$

as well as

$$ES \circ SE = Id \quad \text{and} \quad SE \circ ES = Id.$$

The rule of a thumb is that you should do the check if you have any doubts. If not you proceed, but still you may have to come back (and check) if these ideas become important at some later time, i.e., necessary in order to understand some more involved questions.

Remark. Now we see that for any equivalence relation \sim , the expression $a \sim b$ is equivalent to $q(a) = q(b)$, i.e., to $[a] = [b]$. So, the meaning of any equivalence relation is that elements are “the same in some aspect” (their q images are the same), or that they “perform the same task” (they represent the same equivalence classes).

5.3. Constructing Numbers: \mathbb{Z} from \mathbb{N} .

5.4. Construction of the fraction field of an integral domain. Let A be an integral domain. We assume that the definition of integral domain requires that $0 \neq 1$ – if $1 = 0$ then $A = \{0\}$ since for any $a \in A$ we would have $a = a \cdot 1 = a \cdot 0 = 0$.

5.4.1. *The smallest subfield that contains A .* If A lies in some field \mathcal{G} then \mathcal{G} contains for each pair $a \in A$ and $b \in A - \{0\}$ an element ab^{-1} . Denote by G the subset of \mathcal{G} consisting of all elements $x \in \mathcal{G}$ such that $x = ab^{-1}$ for some $a \in A$ and $b \in A - \{0\}$.

Lemma. (a) G is a subfield of \mathcal{G} .

(b) G is the smallest subfield of \mathcal{G} that contains A .

Proof. (a) G is a subfield because it is closed under products and sums

$$ab^{-1} \cdot a'b'^{-1} = (aa')(bb')^{-1} \quad \text{and} \quad ab^{-1} + a'b'^{-1} = (ab' + a'b)(bb')^{-1},$$

and also under inverses because if $ab^{-1} \neq 0$ then $a \neq 0$, and then $ba \in G$ and $ab^{-1} \cdot ba^{-1} = 1$.

(b) For $a \in A$ we can choose $b = 1 \in A - \{0\}$, then G contains $a \cdot 1^{-1} = a$, so G contains A .

If G' is any subfield of \mathcal{G} that contains A then for each pair $a \in A$ and $b \in A - \{0\}$ G' would contain a, b , and therefore also b^{-1} and ab^{-1} . So, G' contains A .

5.4.2. *The description of a minimal field that contains A in terms of $A \times (A - \{0\})$.* We have seen that in order to have a field G that contains A , we can choose G so that all elements of G are the form ab^{-1} for some pair $(a, b) \in A \times (A - \{0\})$. Then all elements of the field G come from $A \times (A - \{0\})$ by means of a function $f : A \times (A - \{0\}) \rightarrow G$, $f(a, b) = ab^{-1}$. So, G is obtained from $A \times (A - \{0\})$ by identifying any two pairs $(a, b), (a', b') \in A \times (A - \{0\})$ which give the same element of G , i.e., $f(a, b) = f(a', b')$.

To make this idea precise we define a relation \sim on $A \times (A - \{0\})$ so that $(a, b) \sim (a', b')$ if $f(a, b) = f(a', b')$. We know that \sim is an equivalence relations, because this is true for any relation defined by equality of values of some function. Now we know what does it mean to identify pairs (a, b) and (a', b') which give the same element of G . It means that we pass from $A \times (A - \{0\})$ to its quotient $A \times (A - \{0\}) / \sim$ by the equivalence relation \sim .

This gives a precise description of G in terms of $A \times (A - \{0\})$. To any equivalence class $[(a, b)] \in$ is that there is a natural bijection from $A \times (A - \{0\}) / \sim$ to G , it associates to equivalence class $[(a, b)]$ the element $f(a, b) = ab^{-1}$ of G .

However, this is not really a description in terms of $A \times (A - \{0\})$ alone because the relation $(a, b) \sim (a', b')$ if defined by an equality $ab^{-1} = a'b'^{-1}$ in G . So, we are using G in order to describe G .

Fortunately, this relation can be described without using G . By multiplying both sides of $ab^{-1} = a'b'^{-1}$ with bb' we see that this equality is equivalent to the equality $ab' = a'b$ in the field G . Moreover, since a, b, a', b' are in A , both ab' and $a'b$ are in A , therefore $ab' = a'b$ holds in G iff it holds in A .

So, the final conclusion is the following.

- If we know that A lies in some field then a minimal field that contains A can be described as $A \times (A - \{0\}) / \sim$ for the equivalence relation \sim on $A \times (A - \{0\})$, defined by $(a, b) \sim (a', b')$ if $ab' = a'b$ holds in A .

5.4.3. *The idea.* Recall that our problem is that we start with some integral domain A , and we *do not know* whether it lies in a field, and we wish to construct a field F that contains A .

So far we have noticed that if F exists then its smallest version can be described as $A \times (A - \{0\}) / \sim$ where $(a, b) \sim (a', b')$ if $ab' = a'b$ holds in A .

Therefore the only possible solution F is $A \times (A - \{0\}) / \sim$. The problem is to show that this really is a solution, i.e., that it has a natural structure of a field that contains A . This is what we will do:

5.4.4. *Construction.* Here we construct from an integral domain A a field F that contains A . This field F that we construct is called the *fraction field of A* .

Lemma. Let A be an integral domain. Then the following relation on $A \times (A - \{0\})$

$$(a, b) \sim (a', b') \text{ iff } ab' = a'b$$

is an equivalence relation, and the quotient

$$F \stackrel{\text{def}}{=} A \times (A - \{0\}) / \sim$$

is a minimal field that contains A .

More precisely, if we denote the equivalence class $[(a, b)]$ of the pair $(a, b) \in A \times (A - \{0\})$ by the symbol $\frac{a}{b}$, then the operations in F are given by

$$\frac{a}{b} + \frac{\alpha}{\beta} \stackrel{\text{def}}{=} \frac{a\beta + b\alpha}{b\beta} \quad \text{and} \quad \frac{a}{b} \cdot \frac{\alpha}{\beta} \stackrel{\text{def}}{=} \frac{a\alpha}{b\beta};$$

while the embedding of A into F is given by

$$\iota : A \rightarrow F, \quad \iota(a) \stackrel{\text{def}}{=} \frac{a}{1}, \quad a \in A.$$

Proof. **I.** \sim is an equivalence relation.

II. Operations $+$ and \cdot in F are well defined, i.e., they do not depend on the choice of representatives of equivalence classes.

III. $(F, +, \cdot)$ is a field.

IV. Map ι is an “embedding of rings”. Actually what this phrase means is that: (i) function ι is injective and (ii) ι is a morphism of rings.

V. Any subfield of F that contains A is equal to F .

6. Algebra and geometry

Algebra is fairly abstract and it is difficult to develop intuition about algebra. Fortunately, for algebraists (those who work with serious algebra), there is an intimate and very important relation between algebra and geometry. This relation

- Provides geometric intuition in algebra (now to solve an algebraic problem one can draw geometric pictures!).
- Among all kinds of geometries, the ones related algebra are the best understood ones.

- (1) There is a 1-1 correspondence between commutative rings and certain kind of geometric objects which are called *affine schemes*.
- (2) This correspondence is a complete translation, i.e., any question on commutative rings can be translated into a question on affine schemes, and vice versa.

The way that the correspondence works is that

- From a commutative ring A one can construct an affine scheme $\text{Spec}(A)$.
 - Points of $\text{Spec}(A)$ are the maximal ideals in A .
 - Geometric subobjects of X correspond to all ideals in A . (For instance if X is the three dimensional space geometric subobjects are points in X , curves in X and surfaces in X .)
 - Geometric subobjects which can not be decomposed into a union of two simpler geometric subobjects correspond to prime ideals in A .
- From any affine scheme X one constructs a commutative ring $\mathcal{O}(X)$, which is just the ring of (polynomial) functions on X .

Examples. Here are some examples of rings A and schemes X that correspond to each other, in other words $A = \mathcal{O}(X)$ is the ring of polynomial functions on X , and X is the spectrum $\text{Spec}(A)$ of the commutative ring A .

- (1) X is the real line \mathbb{R} and $A = \mathbb{R}[X]$ are the polynomials in one variable.
- (2) X is the real plane \mathbb{R}^2 and $A = \mathbb{R}[X_1, X_2]$ are the polynomials in two variables.
- (3) X is the real space \mathbb{R}^3 and $A = \mathbb{R}[X_1, X_2, X_3]$ are the polynomials in three variables.
- (4) X is the real n -dimensional space \mathbb{R}^n and $A = \mathbb{R}[X_1, X_2, \dots, X_n]$ are the polynomials in n variables.
- (5) X is the infinitesimal neighborhood of the point 0 on the real line and $A = \mathbb{R}[[X]]$ is the ring of formal power series with real coefficients.

Remarks. (0) Since we are expressing geometric idea through algebra we are creating many new kinds of geometries because we have many kinds of commutative rings. This makes many geometric ideas extend to totally new settings, for instance ring $\mathbb{Z}_p[X_1, X_2]$ corresponds to a geometric object which is a planes based on a finite field \mathbb{F}_p instead of the field \mathbb{R} of real numbers.

- (1) There is a larger class of geometric objects called *schemes* such that any scheme is obtained by gluing together several affine schemes. SO, in some sense schemes are obtained by gluing together commutative rings.

7. Modules

Groups act on sets. Rings act on modules.