Algebraic structures

Application of discrete models

1 Ring

We have seen many different kinds of mathematical objects with similar operations than numbers (such as addition, subtraction, multiplication): matrices, polynomials, residue classes etc. These operations have many similar properties (e.g. it is true for all of them that $a + b = b + a$), but they also have some differences (e.g. for numbers, $a \cdot b = b \cdot a$, but this is not true for matrices). In this part, we will do *abstraction*: for all these different mathematical objects, we will study only their common properties, and "forget" their individual characteristics. In this way, any conclusion we will have in general will automatically apply to all particular cases.

Definition 1.1 (ring)**.** Let R be a set with two binary operations: "$+$" and "$\cdot$". The tuple $(R, +, \cdot)$ is called a *ring* if the following properties (*ring axioms*) are satisfied:

A. *additive operation* $(+)$:
   1. closed: $\forall a, b \in R : \ a + b \in R$;
   2. associative: $\forall a, b, c \in R : \ (a + b) + c = a + (b + c)$;
   3. commutative: $\forall a, b \in R : \ a + b = b + a$;
   4. *zero element*: $\exists z \in R : \forall a \in R : \ a + z = a$;
   5. *additive inverse*: $\forall a \in R : \exists x \in R : \ a + x = z$ (z is from above) (x is denoted by $-a$);

M. *multiplicative operation* $(\cdot)$:
   1. closed: $\forall a, b \in R : \ a \cdot b \in R$;
   2. associative: $\forall a, b, c \in R : \ (a \cdot b) \cdot c = a \cdot (b \cdot c)$;

D. "$\cdot$" is distributive over "$+$":
   $\forall a, b, c \in R : \ a \cdot (b + c) = a \cdot b + a \cdot c \ \wedge \ (b + c) \cdot a = b \cdot a + c \cdot a$.

The most obvious example is the integers, i.e. $(\mathbb{Z}, +, \cdot)$, which satisfy all required properties above, so they form a ring. The *zero element* from A/4 is the zero number ($z = 0$), since $a + 0 = a$, and the *additive inverse* from A/5 is the negative of the number ($x = -a$), since $a + (-a) = 0$.

Notice that rings have much fewer requirements on the multiplicative operation than on the additive operation. This allows more kinds of objects to fit into the definition of rings (e.g. matrices, where multiplication is not commutative). Later we will see other algebraic structures with more requirements on the multiplicative operation.

(Note: although the two ring operations were denoted by "$+$" and "$\cdot$", they can really be any operations, not just addition and multiplication, as long as they satisfy all the requirements above. But since we mostly deal with these two operations anyway, we will often – slightly inaccurately – omit them from the notation, e.g. instead of $(\mathbb{Z}, +, \cdot)$, we simply write that "$\mathbb{Z}$ is a ring". Furthermore, if the operation is multiplication, its sign may be omitted, e.g. we may write ab instead of $a \cdot b$.)

Examples of rings:

1. Most of the basic number sets are rings:
   (a) $\mathbb{Z}$ (the set of integers),
   (b) $\mathbb{Q}$ (the set of rational numbers),
   (c) $\mathbb{R}$ (the set of real numbers),
   (d) $\mathbb{C}$ (the set of complex numbers)

   with the usual addition and multiplication. The zero element of each ring is the number 0, and the additive inverse of a number is its negative $(-a)$. But the set of natural numbers ($\mathbb{N}$) is *not* a ring, as it has no additive inverse: there is no such natural number x that e.g. $2 + x = 0$.
2. The set of $n \times n$ matrices, e.g. $\mathbb{R}^{n\times n}$ (for a fixed n) is a ring, with matrix addition and matrix multiplication. The zero element is the zero matrix, i.e. the matrix with all zeros, e.g. $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. The additive inverse of an $A \in \mathbb{R}^{n\times n}$ is the matrix $-A \in \mathbb{R}^{n\times n}$ with all the elements negated (e.g. if $A = \left(\begin{smallmatrix} 2 & 1 \\ 0 & -3 \end{smallmatrix}\right)$, then $-A = \left(\begin{smallmatrix} -2 & -1 \\ 0 & 3 \end{smallmatrix}\right)$), because then $A + (-A)$ is the zero matrix.
3. The set of polynomials, e.g. $\mathbb{R}[x]$ is also a ring. The zero element is the zero polynomial ($0 \in \mathbb{R}[x]$), and the additive inverse of a polynomial p is $-p = -1 \cdot p$ (e.g. $-(2x^2-3) = -2x^2+3$).
4. The residue classes for a given $m \in \mathbb{N}^+$ also form a ring: $\mathbb{Z}_m = \{\overline{0}, \overline{1}, \ldots, \overline{m-1}\}$ (see the part about Modular arithmetic). Reminder: the operations are performed modularly, i.e. the result of each operation is the remainder of the numerical result by m, e.g. in $\mathbb{Z}_5$, $\overline{3} + \overline{4} = \overline{2}$ and $\overline{3} \cdot \overline{3} = \overline{4}$. The zero element is the residue class $\overline{0}$. The additive inverse of an $\overline{a} \in \mathbb{Z}_m$ is $\overline{-a} = \overline{m-a}$, e.g. in $\mathbb{Z}_5$, the additive inverses of $\overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}$ are $\overline{0}, \overline{4}, \overline{3}, \overline{2}, \overline{1}$ respectively (since $\overline{0} + \overline{0} = \overline{0}$, $\overline{1} + \overline{4} = \overline{0}$ etc.).
5. The even numbers, $2\mathbb{Z} := \{2k \mid k \in \mathbb{Z}\}$ also form a ring. Since $2\mathbb{Z} \subseteq \mathbb{Z}$, and we already know that $\mathbb{Z}$ is a ring, many ring axioms are inherited automatically (associativity, commutativity and distributivity). The others must be checked: $2\mathbb{Z}$ is closed under addition (A/1) and multiplication (M/1), because the sum and product of even numbers are even; the zero element is inherited, because 0 is even; and so is the additive inverse, because $-2k$ is also even.
   But the set of odd numbers do *not* form a ring, because they are not closed under addition (A/1): e.g. $1 + 3 = 4$, i.e. the sum of two odd numbers are not odd.
6. The following example demonstrates that the two ring operations need not necessarily be addition and multiplication, they can be any operations. Let H be an arbitrary set, and consider all possible subsets of H – in other words, take the power set of H, i.e. 2^H (e.g. if $H = \{a, b\}$, then $2^H = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$). Then $(2^H, \triangle, \cap)$ is a ring, i.e. the subsets of H with the symmetric difference ($\triangle$) and set intersection ($\cap$) operations (see: Discrete mathematics I) form a ring. All ring axioms are satisfied, but here the additive operation is $\triangle$, and the multiplicative operation is $\cap$. For example, the (first half of the) distributivity is: $\forall A, B, C \in 2^H : A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$. The zero element of the ring is the empty set, $\emptyset$ (since $A \triangle \emptyset = A$), and the additive inverse of each set is itself (since $A \triangle A = \emptyset$).
7. The simplest possible ring is a set with a single element: $\{a\}$, where the operations are defined in the only possible way: $a + a = a$ and $a \cdot a = a$. Then $(\{a\}, +, \cdot)$ is a ring, because all ring axioms are trivially satisfied (e.g. commutativity is $a + a = a + a$ etc.). The zero element of the ring is its only element (a), and the additive inverse of a is a itself. This ring is called the *zero ring* (because its only element is the zero element).

Theorem 1.2 (properties of rings)**.**

1. *The zero element is unique (i.e. if z_1 and z_2 are both zero elements, then $z_1 = z_2$).*
2. *The additive inverse is unique (i.e. if x_1 and x_2 are both additive inverses of a, then $x_1 = x_2$).*
3. *The additive inverse of the zero element is itself: $-z = z$.*
4. *The zero element absorbs any element by multiplication: $\forall a \in R : za = z \wedge az = z$.*

(These are simple consequences of the ring axioms. For example, for the first one, since z_1 is a zero element, $z_1 + z_2 = z_2$, but since z_2 is also a zero element, $z_1 + z_2 = z_1$, so $z_1 = z_2$.)

The definition of the ring requires only two operations (addition and multiplication), but for numbers, there are four fundamental operations. What about the others? For example, subtraction can be defined using addition and additive inverse:

Definition 1.3 (subtraction)**.** For a ring R and $a, b \in R$: $a - b := a + (-b)$.

It follows from the definition that $(a - b) + b = a$ and $(a + b) - b = a$, as expected.

(Note: if the additive operation of the ring is not addition, then this operation is not called subtraction either, but the *inverse* of that operation. E.g. for the ring $(2^H, \triangle, \cap)$ above, the inverse of $\triangle$ is $\triangle$ itself, because the additive inverse of each element is itself. Later we will define division as the inverse of the multiplication in a similar way, but we need some additional definitions before that.)

2 Integral domain

The definition of the ring is sufficiently general, but sometimes we need more requirements, especially on the multiplicative operation. So we introduce some further notions.

The additive operation is commutative in every ring by definition, but it can be useful if the multiplicative operation is commutative too:

Definition 2.1 (commutative ring)**.** A ring R is *commutative* if the multiplicative operation $(\cdot)$ is commutative, i.e. $\forall a, b \in R : a \cdot b = b \cdot a$.

Almost all of the above examples are commutative rings. The exception is the matrices, e.g. if $A = \left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right)$, then $AB = \left(\begin{smallmatrix} 4 & 4 \\ -4 & -4 \end{smallmatrix}\right)$, but $BA = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

We can create the analogue of the zero element for the multiplicative operation:

Definition 2.2 (identity)**.** $e \in R$ is the *identity element* of the ring R if $\forall a \in R : ae = a \wedge ea = a$.

A ring that contains an identity element is called a *ring with identity*.

Theorem 2.3. *The identity element is unique, i.e. if e_1 and e_2 are both identities, then $e_1 = e_2$.*

Most of the above example rings have an identity element:

1. The identity of the number rings ($\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) is 1.
2. The identity of a matrix ring is the *identity matrix*, e.g. $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ in $\mathbb{R}^{2\times 2}$.
3. The identity of a polynomial ring is the constant polynomial 1.
4. The identity of $\mathbb{Z}_m$ is the residue class $\overline{1}$.
5. The ring of even numbers ($2\mathbb{Z}$) has no identity element.
6. The identity element of the ring $(2^H, \triangle, \cap)$ is H itself, since for all $A \in 2^H$ (i.e. $A \subseteq H$) sets, $A \cap H = A$ and $H \cap A = A$.
7. The identity element of the zero ring is its only element. (Note: this is the only ring where the identity and the zero element are the same, because in larger rings, the absorption property of the zero element ($za = z$) prevents it from being also the identity element.)

The following notion is not found in number rings, and it can be annoying, so we try to avoid them:
Definition 2.4 (zero divisor)**.** $a \in R \setminus \{z\}$ is a *zero divisor* in the ring R if there exists $b \in R \setminus \{z\}$ such that $ab = z$ or $ba = z$ (where z is the zero element).
So if R has no zero divisors, then $ab = z$ implies that either $a = z$ or $b = z$. The number rings have no zero divisors, because if $ab = 0$, then $a = 0$ or $b = 0$. This property was fundamental for solving equations, because this ensured that e.g. if $(x-1)(x-2) = 0$, then $x = 1$ or $x = 2$, and there is no other solution. With zero divisors, this would be much more complicated.
Examples of zero divisors:

1. For matrices, e.g. $\left(\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, i.e. $\left(\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right)$ are zero divisors in $\mathbb{R}^{2\times 2}$.
2. In $\mathbb{Z}_6$, e.g. $\overline{2} \cdot \overline{3} = \overline{0}$ (because $2 \cdot 3 = 6$), i.e. $\overline{2}$ and $\overline{3}$ are zero divisors in $\mathbb{Z}_6$.

Theorem 2.5 (zero divisors in $\mathbb{Z}_m$)**.**

1. *$\overline{a} \in \mathbb{Z}_m$ is a zero divisor if and only if $\overline{a} \neq \overline{0}$ and $\gcd(a, m) \neq 1$.*
2. *$\mathbb{Z}_m$ has zero divisors if and only if m is a composite number.*

So for prime numbers, $\mathbb{Z}_p$ has no zero divisors.
Rings that have all the good properties from above will be very useful later:
Definition 2.6 (integral domain)**.** The ring R is called an *integral domain* if

0. it has at least two elements,
1. it is commutative,
2. it has an identity element and
3. it has no zero divisors.

Examples of integral domains include the four basic number rings ($\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$), polynomial rings over numbers ($\mathbb{Z}[x]$, $\mathbb{R}[x]$ etc.), and $\mathbb{Z}_p$ for prime numbers.
(The reason for the "at least two elements" criterion is to exclude the zero ring, therefore ensure that the identity element is not the same as the zero element, which would lead to complications later.)

3 Field

We need one more important property: the analogue of the additive inverse $(-a)$ for multiplication:
Definition 3.1 (multiplicative inverse)**.** Let R be a ring with identity. The *multiplicate inverse* of $a \in R$ is x if $ax = e$ and $xa = e$ (where e is the identity element). Notation: a^{-1}. (It is sometimes simply called the *inverse* of a.) An element is called *invertible* if it has a multiplicative inverse.
Examples of the multiplicative inverse:

1. The multiplicative inverse of $2 \in \mathbb{Q}$ is $1/2 \in \mathbb{Q}$, because $2 \cdot 1/2 = 1$.
2. In $\mathbb{Z}$, only 1 and -1 are invertible (the inverse of each is itself).
3. The multiplicative inverse of $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 5 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$ is $A^{-1} = \left(\begin{smallmatrix} -5 & 2 \\ 3 & -1 \end{smallmatrix}\right)$, because $AA^{-1} = I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $A^{-1}A = I$. On the other hand, $B = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 6 \end{smallmatrix}\right)$ is not invertibe, because $\det B = 0$.
4. In $\mathbb{Z}_m$, the invertible elements are those $\overline{a} \in \mathbb{Z}_m$ for which $\gcd(a, m) = 1$.

Theorem 3.2 (properties of the multiplicative inverse)**.**

1. *The multiplicative inverse is unique.*
2. *The multiplicative inverse of the identity element is itself: $e^{-1} = e$.*
3. *The zero element is not invertible.*
4. *A zero divisor is not invertible.*

Now we can define one of the most important algebraic structures (besides the ring):

Definition 3.3 (field)**.** The ring R is called a *field* if

0. it has at least two elements,
1. it is commutative,
2. it has an identity element and
3. all nonzero elements are invertible.

Notice that any field is also an integral domain, because their definitions differ only in point 3, and an invertible element cannot be a zero divisor (see the above theorem).

(Note: point 3 had to exclude the zero element, which cannot be invertible in any ring, as multiplying any element by it yields the zero element. This corresponds to the fact that numbers cannot be divided by 0.)

From the list of example rings above, the following ones are fields:

1. $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ – they are called *number field*s.
2. $\mathbb{Z}_p$ for any prime number p, because all of its non-$\overline{0}$ elements are invertible.

But note that the ring of integers, $\mathbb{Z}$ is *not* a field, because only 1 and -1 are invertible there.

To summarize fields, let's give their full definition without the use of rings, in a similar format than the definition of rings above:

Definition 3.4 (field)**.** $(F, +, \cdot)$ is a *field* if the following properties (*field axioms*) are satisfied:

A. *additive operation* $(+)$:
   1. closed: $\forall a, b \in F : \ a + b \in F$;
   2. associative: $\forall a, b, c \in F : \ (a + b) + c = a + (b + c)$;
   3. commutative: $\forall a, b \in F : \ a + b = b + a$;
   4. *zero element*: $\exists z \in F : \forall a \in F : \ a + z = a$;
   5. *additive inverse*: $\forall a \in F : \exists x \in F : \ a + x = z$ (where z is the zero element);

M. *multiplicative operation* $(\cdot)$:
   1. closed: $\forall a, b \in F : \ a \cdot b \in F$;
   2. associative: $\forall a, b, c \in F : \ (a \cdot b) \cdot c = a \cdot (b \cdot c)$;
   3. commutative: $\forall a, b \in F : \ a \cdot b = b \cdot a$;
   4. *identity element*: $\exists e \in F \setminus \{z\} : \forall a \in F : \ a \cdot e = a$;
   5. *multiplicative inverse*: $\forall a \in F \setminus \{z\} : \exists x \in F : \ a \cdot x = e$ (where e is the identity element);

D. "$\cdot$" is distributive over "$+$":
   $\forall a, b, c \in F : \ a \cdot (b + c) = a \cdot b + a \cdot c$.

We can see that unlike rings, fields have exactly as many requirements on the multiplicative operation as on the additive one (except for the multiplicative inverse of the zero element, which is impossible).

(Note: because the multiplication is now known to be commutative, certain properties were simplified, e.g. now the distributivity can be stated in only one direction.)

In fields, we can finally define the fourth fundamental operation (similarly to subtraction, see above):

Definition 3.5 (division)**.** For a field F and $a, b \in F$, $b \neq z$: $a/b := ab^{-1}$.

For example, in $\mathbb{Q}$, it means that $5/2 = 5 \cdot 2^{-1} = 5 \cdot (1/2)$.

It follows from the definition that $(a/b)\, b = a$ and $(ab)/b = a$, as expected.

4 Rings in Sage

In Sage, the most common rings can be created as follows:

```
sage: ZZ
Integer Ring
sage: QQ
Rational Field
sage: RR     # avoid RR and CC, because they can only be represented numerically
Real Field with 53 bits of precision
sage: CC
Complex Field with 53 bits of precision
sage: Mat(QQ, 5)    # matrices
Full MatrixSpace of 5 by 5 dense matrices over Rational Field
sage: ZZ["x"]
Univariate Polynomial Ring in x over Integer Ring
sage: Zmod(6)
Ring of integers modulo 6
```

The elements of these rings can be accessed in various ways:

```
sage: type(12)      # integers are already in ZZ
<class 'sage.rings.integer.Integer'>
sage: type(2/3)     # and fractions are in QQ
<class 'sage.rings.rational.Rational'>
sage: type(QQ(12))  # conversion to another ring: ZZ -> QQ
<class 'sage.rings.rational.Rational'>
sage: ZZ(2/3)       # QQ -> ZZ doesn't always work
[...]
TypeError: no conversion of this rational to integer
sage: RR(2/3)       # RR and CC can only be represented numerically
0.666666666666667
sage: Zmod(5)(13)   # ZZ -> Zmod(5)
3
sage: Zmod(10).list()  # finite rings can be enumerated
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
sage: [a^2 for a in Zmod(10)]
[0, 1, 4, 9, 6, 5, 6, 9, 4, 1]
sage: ZZ.zero()     # zero element
0
sage: ZZ.one()      # identity element
1
sage: Mat(ZZ, 3).one()
[1 0 0]
[0 1 0]
[0 0 1]
```

5 Polynomials over rings

In the previous part (Polynomials), we discussed polynomials, but only with numbers as coefficients (e.g. $\mathbb{Z}[x]$, $\mathbb{R}[x]$). Now we generalize them to rings.

In this section, let K be an arbitrary commutative ring, and consider $K[x]$, the set of *polynomials over* K, i.e. the polynomials whose coefficients are from the ring K.

For example, $\mathbb{Z}_4[x]$ is the set of polynomials over the residue classes modulo 4 ($\mathbb{Z}_4 = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$):

```
sage: P4.<x> = Zmod(4)[]
sage: P4
Univariate Polynomial Ring in x over Ring of integers modulo 4
sage: p = 2*x + 3; q = 3*x + 1
sage: p + q
x
sage: p * q
2*x^2 + 3*x + 3
```

Why is $p + q = x$? Because in $\mathbb{Z}_4$, $\overline{2} + \overline{3} = \overline{1}$ and $\overline{3} + \overline{1} = \overline{0}$. In general, operations in $\mathbb{Z}_m[x]$ can be performed by first calculating the result in $\mathbb{Z}[x]$, and then taking the coefficients modulo m. For example, in $\mathbb{Z}[x]$, the above $p \cdot q = 6x^2 + 11x + 3$, so in $\mathbb{Z}_4[x]$, $p \cdot q = \overline{2}x^2 + \overline{3}x + \overline{3}$.

Back to the general case, it is easy to prove that $K[x]$ is also a ring, and it inherits most of the additional properties of the coefficient ring K:

Theorem 5.1 (properties of $K[x]$)**.**

1. *If K is a commutative ring, then $K[x]$ is also a commutative ring.*
2. *If K is a ring with identity, then $K[x]$ is also a ring with identity.*
3. *If K has no zero divisors, then $K[x]$ has no zero divisors either.*
4. *If K is an integral domain, then $K[x]$ is also an integral domain.*

So for example $K[x]$ inherits the identity element of K, and it doesn't create any new zero divisors if K has none. But there is one property which cannot be fully inherited:

Theorem 5.2. *The polynomial $x \in K[x]$ has no multiplicative inverse.*

So the ring $K[x]$ can never be a field, regardless of whether K is a field or not.

So $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, $\mathbb{R}[x]$ and $\mathbb{C}[x]$ are integral domains, but none of them are fields.

Now let's evalulate the statements of the previous part (Polynomials), and see which of them work for any polynomial ring $K[x]$, and which of them need some further restrictions on K. In general, the convention was that if a theorem or definition was stated for $K[x]$ (as opposed to a more specific set, e.g. $\mathbb{R}[x]$), then it works for any commutative ring K with identity, but if it was stated for a more specific set, then it needs some restrictions on K (which might still be looser than what was given). More specifically, we need to pay attention to the following:

1. First of all, if K were not commutative, then even the multiplication of two polynomials were problematic, e.g. for $2x \cdot 2x = 4x^2$, we need that $x \cdot 2 = 2 \cdot x$. That's why we restricted K to be a commutative ring.
2. If K has no identity element, then we encounter problems with the root factors: the leading coefficient of $x - c$, though not indicated, is the identity element (e.g. $1x - c$); so if K has no identity, then $x - c \notin K[x]$. Therefore, K is better be a ring with identity.

3. The theorem about the degree of the product of two polynomials (Theorem 1.7. part 3.), i.e. $\deg(p \cdot q) = \deg p + \deg q$ is true only if K has no zero divisors. For example, in $\mathbb{Z}_6[x]$, we have $(\overline{2}x + \overline{1})(\overline{3}x + \overline{1}) = \overline{5}x + \overline{1}$, i.e. the product is not a quadratic polynomial, because the original leading coefficient has vanished due to zero divisors: $\overline{2}x \cdot \overline{3}x = \overline{0}x^2$.
4. The theorem that a polynomial of degree n has at most n roots (Theorem 4.5.) also breaks with zero divisors: e.g. in $\mathbb{Z}_6[x]$, the polynomial $p = x^2 + x$ has *four* roots: $\overline{0}, \overline{2}, \overline{3}, \overline{5} \in \mathbb{Z}_6$ (also, the root factor form is not unique: $p = (x - \overline{2})(x - \overline{3}) = x(x - \overline{5})$). That's why the theorem had the condition "K is a number set", which can now be generalized to "K has no zero divisors". (In the proof of the theorem, we used this fact when saying "if $p(c) = 0$, then one of the factors must be 0".) Because zero divisors cause so many inconveniences, we mostly deal with polynomials over integral domains.
5. Division with remainder (Theorem 5.2.) needs even more than integral domains. For example, the quotient of $f = x^2 + 3$ and $g = 2x$ is $q = 1/2 \cdot x$, and the remainder is $r = 3$, so in general, we can't perform this operation in $\mathbb{Z}[x]$, only in $\mathbb{Q}[x]$. This is because we need to divide by the leading coefficient, which in general works only in fields. That's why the theorem was stated for $\mathbb{R}[x]$, but it is true over any other field, e.g. for $\mathbb{Q}[x]$, $\mathbb{Z}_5[x]$ etc. Similarly, the Euclidean algorithm (and Bézout's identity) works for polynomials over fields.
6. The formal derivative of polynomials can be problematic even over fields. First, how can we apply the definition, which says that the derivative of $a_n x^n$ is $n a_n x^{n-1}$, if $a_n \in K$ but $n \in \mathbb{N}$? That is, how can we multiply an element of an arbitrary ring by n, which is an integer? (The exponents are always integers, no matter what ring the coefficients are from.) This can be solved by repeated addition: $n a_n := a_n + a_n + \ldots + a_n$ with exactly n terms. E.g. in $\mathbb{Z}_3[x]$, the formal derivative of $p = \overline{2}x^5$ is $p' = 5 \cdot \overline{2}x^4 = (\overline{2} + \overline{2} + \overline{2} + \overline{2} + \overline{2})x^4 = \overline{1}x^4$. (Note: the word "formal" is especially justified here, because calculus makes no sense in $\mathbb{Z}_3$.) But this can lead to unexpected results. For example, in $\mathbb{R}[x]$, we know that the derivative of a degree n polynomial has degree $n-1$, but this is not always true in $\mathbb{Z}_m[x]$: e.g. the cubic polynomial $p = x^3 + \overline{2}x^2 + \overline{2}x + \overline{1} \in \mathbb{Z}_3[x]$ has only a linear derivative: $p' = x + \overline{2}$, because the quadratic term of p' vanishes in $\mathbb{Z}_3$: $3 \cdot \overline{1}x^2 = (\overline{1} + \overline{1} + \overline{1})x^2 = \overline{0}x^2$. It can also happen that the whole polynomial vanishes, e.g. the derivative of $p = x^6 + x^3 + \overline{1} \in \mathbb{Z}_3[x]$ is $p' = \overline{0}$. This breaks the method $p \,/\, \gcd(p, p')$, which would filter out the multiple roots (see: Polynomials) – even though p has plenty of them: $p = x^6 + x^3 + \overline{1} = (x - \overline{1})^6$, i.e. it has a sextuple root, $c = \overline{1}$.

6 Divisibility in rings

Earlier, we discussed divisibility for integers and for polynomials. Now we generalize this to rings, more precisely to integral domains. Reminder: the main integral domains are the integers ($\mathbb{Z}$), the polynomial rings ($K[x]$, where K is an integral domain) and all fields.
So in this section, let R be an arbitrary integral domain.

Definition 6.1 (divisor)**.** $a \in R$ is a *divisor* of $b \in R$ if $\exists c \in R : b = ac$. Notation: $a \mid b$.

The definition is exactly the same as for integers (e.g. $2 \mid 6$) and for polynomials (e.g. $x+1 \mid x^2+x$). In fields, divisibility is trivial, because all elements are divisible by all nonzero elements (since $c := b/a$), e.g. in $\mathbb{Q}$, 5 is divisible by 2, because $5 = 2 \cdot (5/2)$.

To discuss divisibility in general, we need a few more definitions:

Definition 6.2 (unit)**.** $u \in R$ is a *unit* if it divides the identity: $u \mid e$, where e is the identity element.

For example:

1. $\mathbb{Z}$ has two units: 1 and -1 (the divisors of 1).
2. In $\mathbb{R}[x]$, all nonzero constant polynomials are units.
3. In general, the units of $K[x]$ are the units of K (as constant polynomials), e.g. in $\mathbb{Z}[x]$, the units are 1 and -1.
4. In fields, all nonzero elements are units.

Note: units are exactly the invertible elements.

Definition 6.3 (associates)**.** $a \in R$ and $b \in R$ are *associates* if $a \mid b$ and $b \mid a$.

For example:

1. In $\mathbb{Z}$, the associates are those numbers which have the same magnitude, e.g. 5 and -5;
2. In $\mathbb{R}[x]$, the associates are the constant multiples of each other, e.g. $x - 1$, $2x - 2$, $-x + 1$, $-x/2 + 1/2$ etc. are all associates.
3. In fields, any two nonzero elements are associates.

The associates are always unit multiples of each other, e.g. in $\mathbb{Z}$, $-5 = -1 \cdot 5$, in $\mathbb{R}[x]$, $2x-2 = 2(x-1)$.

The point of associates is that they behave in exactly the same way with respect to divisibility (e.g. $2 \mid 6 \implies 2 \mid -6$ etc.). We could say that the associates are "essentially" the same.

Definition 6.4 (trivial divisor)**.** The *trivial divisors* of $b \in R$ are the units and the associates of b.

This corresponds to the definition in $\mathbb{Z}$, e.g. the trivial divisors of 6 are 1, -1, 6 and -6.

The notion of prime numbers can be generalized as follows:

Definition 6.5 (irreducible element)**.** An element is *irreducible* if it is nonzero, non-unit, and it has only trivial divisors.

(The "nonzero, non-unit" corresponds to excluding 0 and 1 from the definition of prime numbers.)

For example:

1. In $\mathbb{Z}$, the irreducible elements are the prime numbers and their negatives: $2, 3, 5, 7, 11, \ldots$, $-2, -3, -5, -7, \ldots$. (For simplicity, prime numbers were defined to be positive, but in general, it is not so easy to make a convenient choice from among the associates.)
2. In $\mathbb{Z}[x]$, $x^2 - 2$ is irreducible, but in $\mathbb{R}[x]$, it is not, because $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.
3. Fields have no irreducible elements (because all nonzero elements are units).

Note: the associates of irreducible elements are also irreducible.

In polynomial rings, irreducible polynomials have the following properties:

1. Over fields, all linear polynomials are irreducible, because the divisors of $p = ax+b$ are constant and linear polynomials, the former of which are units, and the latters are p's associates.
2. Higher degree irreducible polynomials cannot have roots (in the coefficient ring K), otherwise we could extract a root factor, and thus the polynomial would not be irreducible.
3. In $\mathbb{C}[x]$, only linear polynomials are irreducible, because the fundamental theorem of algebra (see: Polynomials) states that all nonconstant polynomials have roots.
4. In $\mathbb{R}[x]$, only linear and quadratic polynomials can be irreducible (e.g. $2x + 3$ and $x^2 + 1$ are irreducible over $\mathbb{R}$).
5. But in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$, irreducible polynomials can have arbitrarily large degree, e.g. all $x^n - 2$ polynomials are irreducible.

For integers, the fundamental theorem of arithmetic (see: Divisors, prime numbers) guarantees that the prime factorization is unique. But if we try to generalize this to integral domains, we will find that it won't always work, so we need to distinguish those rings for which it works:

Definition 6.6 (UFD)**.** An integral domain R is called a *unique factorization domain* (UFD) if all nonzero non-unit elements can be written uniquely as a product of irreducible elements, up to associatedness and the order of the factors.

The phrase "up to associatedness" means that e.g. the two factorizations $6 = 2 \cdot 3 = (-2) \cdot (-3)$ are considered "essentially" the same, because the only difference is that 2 and 3 were replaced by their associates. (For the fundamental theorem of arithmetic, we could get around associatedness by working with only positive integers.)

For example:

1. $\mathbb{Z}$ is a UFD by the fundamental theorem of arithmetic. For example, $-10 \in \mathbb{Z}$ can be factored in multiple ways: $-10 = (-2) \cdot 5 = 2 \cdot (-5) = 5 \cdot (-2) = (-5) \cdot 2$, but these factorizations differ only in order and associatedness (i.e. sign).
2. All fields are trivially UFD's (since they have no nonzero non-unit elements).
3. If K is a UFD, then $K[x]$ is also a UFD, i.e. $\mathbb{Z}[x]$, $\mathbb{R}[x]$, $\mathbb{Z}_5[x]$ etc. are all UFD's. E.g. in $\mathbb{R}[x]$, $x^2 - 1 = (x + 1)(x - 1) = (2x + 2)(x/2 - 1/2) = (-3x + 3)(-x/3 - 1/3)$ etc., but these factorizations differ only in order and associatedness (i.e. constant multiplier).

In Sage, the `factor()` function (see: Divisors, prime numbers) works for arbitrary UFD's, with the following caveat: it selects the irreducible elements (from its associates) so that they are as simple as possible (e.g. in $\mathbb{Z}$, it selects the positive ones), but if it doesn't work that way (e.g. if the factored number is negative), then it adds a unit (e.g. -1) to the beginning of the product. The result of `factor()` behaves like a sequence (e.g. it can be iterated through by a `for` loop), but the unit, if any, is not part of this sequence, it can be queried separately by `unit()`. For example:

```
sage: F = factor(-60)
sage: F
-1 * 2^2 * 3 * 5
sage: for (p, n) in F:
....:     print(p, n)
2 2
3 1
5 1
sage: F.unit()
-1
sage: P.<x> = QQ[]
sage: F = factor(2*x^2 - 1/2)
sage: F
(2) * (x - 1/2) * (x + 1/2)
sage: [p for (p, n) in F]
[x - 1/2, x + 1/2]
sage: F.unit()
2
```

Now let's generalize two more familiar notions for arbitrary integral domains:

Definition 6.7 (gcd)**.** A *greatest common divisor* of $a \in R$ and $b \in R$ is $c \in R$ if $c \mid a \ \wedge \ c \mid b \ \wedge \ \forall d \in R : d \mid a \ \wedge \ d \mid b \implies d \mid c$. Notation: $\gcd(p, q)$.

Definition 6.8 (lcm)**.** A *least common multiple* of $a \in R$ and $b \in R$ is $c \in R$ if $a \mid c \ \wedge \ b \mid c \ \wedge \ \forall d \in R : a \mid d \ \wedge \ b \mid d \implies c \mid d$. Notation: $\mathrm{lcm}(a, b)$.

Generally, in integral domains, there are two problems with these notions: first, they do not necessarily exist, and they are not unique. So we need to make some restrictions:

Theorem 6.9. *In a unique factorization domain (UFD), any two elements have a greatest common divisor and a least common multiple.*

So in $\mathbb{Z}$, $\mathbb{Z}[x]$, $\mathbb{R}[x]$ etc., gcd and lcm do exist. (The theorem is not surprising if we consider how gcd and lcm can be calculated from the prime factorization in $\mathbb{Z}$.)

Theorem 6.10. *The greatest common divisor and the least common multiple, if they exist, are unique up to associatedness.*

This means that two elements may have more than one gcd's, but they are all associates (and similarly for the lcm's). For polynomials, we have seen what this means: the gcd's are all constant multiples of each other. For integers, the only reason gcd and lcm were unique is that we restricted them to $\mathbb{N}$ (otherwise e.g. $\gcd(6, 4)$ could be either 2 or -2).

7 Field extensions

Definition 7.1. The field F is a *field extension* of the field G if $G \subseteq F$, and the corresponding operations ($+$ and $\cdot$) coincide in G.

For example, $\mathbb{R}$ is a field extension of $\mathbb{Q}$, and $\mathbb{C}$ is a field extension of both of them.

Definition 7.2. The field F is a *simple extension* of the field G if $\exists \alpha \in F \setminus G$ such that F is the smallest field extension of G that contains α. Notation: $F = G(\alpha)$. F is called the *field generated by* α *over* G, and α is the *primitive element* of the extension.

For example, $\mathbb{C}$ is a simple extension of $\mathbb{R}$: $\mathbb{C} = \mathbb{R}(i)$, because there is no other field extension $F \subset \mathbb{C}$ of $\mathbb{R}$ that contains i. This is because the field is closed under addition and multiplication, so if it contains i, it must also contain $2i$, $-5i$, $i + 1$, $2i - 3/2$ etc., i.e. all numbers of the form $a + bi$, where $a, b \in \mathbb{R}$, but then we already got the whole $\mathbb{C}$.

But $\mathbb{R}$ is *not* a simple extension of $\mathbb{Q}$, because there is no such α that $\mathbb{R} = \mathbb{Q}(\alpha)$. (The reason lies in the cardinality of the sets: $\mathbb{Q}$ is countably infinite, but $\mathbb{R}$ is continuum, i.e. it is "much bigger".)

We can create new fields using field extensions. For example, what is $\mathbb{Q}(\sqrt{2})$? (Remember: $\sqrt{2} \notin \mathbb{Q}$.) Besides all rational numbers, this field must contain $\sqrt{2}$, and all necessary numbers to make it closed under the operations. It can be shown that $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. For example, we can perform multiplication as follows: $(a+b\sqrt{2})(c+d\sqrt{2}) = (ac+2bd)+(ad+bc)\sqrt{2}$, and we can calculate the multiplicative inverse using rationalization: $(a + b\sqrt{2})^{-1} = \dfrac{a}{a^2 - 2b^2} + \dfrac{-b}{a^2 - 2b^2}\sqrt{2}$. (Notice how similar $\mathbb{Q}(\sqrt{2})$ and $\mathbb{C} = \mathbb{R}(i)$ are. This is not a coincidence: simple field extensions work similarly.)

Now consider fields with finitely many elements. Examples are all $\mathbb{Z}_p$ for prime p. They can be extended in the same way, e.g. we can create the field $\mathbb{Z}_3(\sqrt{2})$. But what does $\sqrt{2}$ mean here? It means a value whose square is $\overline{2} \in \mathbb{Z}_3$. There is no such element in $\mathbb{Z}_3$ (since $\overline{0}^2 = \overline{0}$, $\overline{1}^2 = \overline{1}$ and $\overline{2}^2 = \overline{1}$), so it is indeed an extension. The resulting field is similar to $\mathbb{Q}(\sqrt{2})$: $\mathbb{Z}_3(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}_3\}$,

e.g. $\overline{1}+\overline{2}\sqrt{2} \in \mathbb{Z}_3(\sqrt{2})$, and we can calculate with these elements similarly. Thus, we created a new finite field, which has $3^2 = 9$ elements (as both a and b have 3 possible values).
In fact, *all* finite fields can be constructed in this way:

Theorem 7.3 (properties of finite fields)**.**

1. *All finite fields have p^k elements, where p is a prime number and $k \geq 1$.*
2. *For all p^k, there exists a finite field with p^k elements, and it is unique up to isomorphism.*
3. *The finite field with p^k elements is a simple extension of $\mathbb{Z}_p$ (if $k \geq 2$).*

Definition 7.4. The finite field with q elements is denoted by $\mathbb{F}_q$.

So for each prime power, there exists a unique finite field with that many elements: $\mathbb{F}_2(=\mathbb{Z}_2)$, $\mathbb{F}_4$, $\mathbb{F}_8$, $\mathbb{F}_{16}$, ..., $\mathbb{F}_3(=\mathbb{Z}_3)$, $\mathbb{F}_9(=\mathbb{Z}_3(\sqrt{2}))$, $\mathbb{F}_{27}$, ..., $\mathbb{F}_5(=\mathbb{Z}_5)$, $\mathbb{F}_{25}$, ... But there is no $\mathbb{F}_6$ or $\mathbb{F}_{10}$, because they are not prime powers. Important: don't confuse $\mathbb{F}_m$ with $\mathbb{Z}_m$, because $\mathbb{F}_4 \neq \mathbb{Z}_4$, $\mathbb{F}_9 \neq \mathbb{Z}_9$ etc., and most of $\mathbb{Z}_m$ are not even fields. Only for prime numbers is it true that $\mathbb{F}_p = \mathbb{Z}_p$.

(The phrase "unique up to isomorphism" means that two finite fields with equal number of elements may look different, but they are always isomorphic, i.e. we can get one from the other by renaming the elements. For example, it can be shown that the set of logical values (T = "true", F = "false") with the "exclusive or" ($\oplus$) and "and" ($\wedge$) operations, i.e. $(\{\mathrm{T},\mathrm{F}\}, \oplus, \wedge)$ forms a field. This looks very different from the field $(\mathbb{Z}_2, +, \cdot)$, but in fact they are "essentially" the same (i.e. isomorphic), because we can get the latter from the former by renaming $\mathrm{F} \to \overline{0}$ and $\mathrm{T} \to \overline{1}$, e.g. $\mathrm{T} \oplus \mathrm{T} = \mathrm{F} \ \to\ \overline{1}+\overline{1}=\overline{0}$, $\mathrm{F} \wedge \mathrm{T} = \mathrm{F} \ \to\ \overline{0}\cdot\overline{1}=\overline{0}$ etc.)

Usually, the primitive element used for the field extension is not given in its explicit form (e.g. $\sqrt{2}$), but by a rule that it must satisfy (e.g. $\alpha^2 = 2$), more precisely a polynomial (e.g. $x^2 - 2$) for which it must be a root of. It can be proven that this process works for any non-linear irreducible polynomial. This polynomial is called the *minimal polynomial* of the primitive element.

In Sage, the finite field $\mathbb{F}_q$ can be created by `GF(q)` (GF = "Galois field"):

```
sage: GF(4)
Finite Field in z2 of size 2^2
sage: GF(4).list()
[0, z2, z2 + 1, 1]
sage: g = GF(4).gen()   # primitive element (the above z2)
sage: g.minpoly()       # minimal polynomial
x^2 + x + 1
sage: g^2 + g + 1       # is g really a root of that?
0
sage: GF(6)
[...]
ValueError: the order of a finite field must be a prime power
```

8 Author

These lecture notes were written by Uray M. János, partially using Nagy Ádám's notes in Hungarian:
https://compalg.elte.gitlab-pages.hu/dimoa-web/tematika/dimoa
Please report errors and suggestions here: uray.janos@inf.elte.hu.