Math 330-2: Abstract Algebra

Winter 2022, Northwestern University

Shuyi Weng

Disclaimer: These lecture notes are written for class-planning purposes only. They are not meant to be a substitution for the textbook. It is likely that the notes contain typos and mistakes. You are encouraged to let me know when you see any of those.

Last modified: Thursday, March 10, 2022, 8:13pm

Contents

1 Rings

Class 1
01/03/2022

In Math 330-1, we studied the algebraic structure *groups* in detail. We will study a new type of algebraic structure—*rings* in this quarter's Math 330-2.

1.1 Definitions and examples

1.1 Definition (*Rings*)
A **ring** is a set R together with two binary operations $+$ and $\cdot$, which we call **addition** and **multiplication**, respectively, that satisfies

- $(R, +)$ forms an abelian group, *i.e.*,
  - $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$;
  - there exists a $0 \in R$ such that $0 + a = a + 0 = a$ for all $a \in R$;
  - for all $a \in R$, there exists $b \in R$ such that $a + b = 0$;
  - $a + b = b + a$ for all $a, b \in R$.
- $(R, \cdot)$ forms a **monoid**, *i.e.*,
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$;
  - there exists a $1 \in R$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in R$.
- Multiplication is **distributive** with respective to addition, *i.e.*,
  - $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in R$;
  - $(a + b) \cdot c = a \cdot c + b \cdot c$ for all $a, b, c \in R$.

1.2 Remark
Fraleigh's textbook does not require the existence of $1 \in R$ as one of the ring axioms. There is no consensus in the mathematical community as to whether the existence of a multiplicative identity must be one of the ring axioms. When we talk about rings in Math 330-1, we will always require the existence of a multiplicative identity, so that we do not have to repetitively say "rings with identity" when we simply want to refer to, well, rings. Mathematicians, with their undying humor, often use *rngs* to refer to rings without "i"dentity.

1.3 Proposition (*Elementary properties of rings*)
Let $(R, +, \cdot)$ be a ring. Some elementary properties follow quickly from the structural properties of binary operations.

(1) There is a unique additive identity $0 \in \mathbb{R}$.

(2) For all $a \in R$, its additive inverse is unique, which we denote by $-a$.

(3) There is a unique multiplicative identity $1 \in \mathbb{R}$.

(4) For all $a \in R$, we have $0 \cdot a = a \cdot 0 = 0$ and $(-1) \cdot a = a \cdot (-1) = -a$.

Proof
Properties (1)–(3) follow from Theorems 2.11 and 3.8 of Math 330-1 notes.

To prove (4), note that

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

Because additive identity is unique, $0 \cdot a = 0$. Likewise, $a \cdot 0 = 0$. On the other hand,

$$0 = 0 \cdot a = [1 + (-1)] \cdot a = 1 \cdot a + (-1) \cdot a = a + (-1) \cdot a.$$

Because additive inverse is unique, $(-1) \cdot a = -a$. Likewise, $a \cdot (-1) = -a$. □

1.4 Notation

Similar with groups, we will usually omit the multiplicative symbol $\cdot$ for ring operation, so that $a \cdot b$ can be written as ab. In addition, when the context is clear, we omit the additive and multiplicative symbols of the binary operations, and refer to the ring $(R, +, \cdot)$ as R.

Under the context of abelian groups, we frequently use na, where $n \in \mathbb{N}$ and $a \in R$, to denote the sum $a + a + \cdots + a$ with n copies of a as summands. When n is a negative integer, na denotes the sum $(-a) + (-a) + \cdots + (-a)$ with $|n|$ copies of $-a$. We will continue with this notation for the additive structure of a ring. We should *not* interpret this notation as the product of n and a, as n may not be in the ring R to begin with.

We will also use the power notation a^n for $n \in \mathbb{N}$ to denote the product $a \cdot a \cdot \cdots \cdot a$ with n copies of a. We should be careful that a^n does not necessarily exist for negative n, as we will see in upcoming discussions.

A few algebraic structures are related to rings with relaxations or restrictions on the defining axioms of the multiplicative monoid structure.

1.5 Definition *(Rngs and commutative rings)*

A **rng** is a triple $(R, +, \cdot)$ such that

- $(R, +)$ forms an abelian group.
- $(R, \cdot)$ is associative, *i.e.*, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$.
- Multiplication is distributive with respective to addition.

In short, a rng is a ring (possibly) without a multiplicative identity.

A **commutative ring** is a ring $(R, +, \cdot)$ whose multiplicative monoid $(R, \cdot)$ is commutative, *i.e.*, $ab = ba$ for all $a, b \in R$.

1.6 Remark

The relation between these three algebraic structures can be summarized as

$$\text{rngs} \supseteq \text{rings} \supseteq \text{commutative rings}$$

We will add more classes of algebraic structures to this chain of inclusion relations through this quarter's discussion on ring theory.

We are now ready to see a large collection of examples for rings, rngs, and commutative rings.

Class 2
01/05/2022

1.7 Example

The most canonical example of a ring is $\mathbb{Z}$ with its usual addition and multiplication. The properties of $\mathbb{Z}$ enumerated in Definition 1.1 are familiar to us since middle school. In addition, the multiplication of integers is commutative, thus making $\mathbb{Z}$ a commutative ring.

Likewise, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all commutative rings, so is $(\mathbb{Z}_n, +_n, \cdot_n)$.

1.8 Example

The set of all even integers $2\mathbb{Z}$ does not have a multiplicative identity. However, it satisfies all the other ring axioms. It is therefore a rng.

Likewise, the set $x\mathbb{Z}$ of all integer multiples of x is a rng for all $x \in \mathbb{R}$. When $x = \pm 1$, we get the ring $\mathbb{Z}$.

1.9 Example (*Trivial ring*)

It is in fact true that $x\mathbb{Z}$ is a ring if $x = 0$. In this case, we have $0\mathbb{Z} = \{0\}$. This is called the **trivial ring**, which is a ring that contains a single element. An equivalent classification of the trivial ring is that the additive identity and the multiplicative identity coincide, as demonstrated in the following lemma.

1.10 Lemma

A ring R is trivial if and only if $0_R = 1_R$.

Proof

If a ring R contains a single element, then its only element is both the additive identity 0_R and the multiplicative identity 1_R.

Suppose $0_R = 1_R$ in a ring R. Let $a \in R$. Then $a = 1_R \cdot a = 0_R \cdot a = 0_R$. Thus every element in R is equal to the additive identity 0_R. Therefore, $R = \{0_R\}$. □

1.11 Example (*Matrix rings*)

Consider $\mathsf{Mat}_n(\mathbb{R})$ the set of $n \times n$ matrices with real entries. $\mathsf{Mat}_n(\mathbb{R})$ forms a ring under matrix addition and matrix multiplication, with the zero matrix $0 \in \mathsf{Mat}_n(\mathbb{R})$ as the additive identity, and the $n \times n$ identity matrix $I_n \in \mathsf{Mat}_n(\mathbb{R})$ as the multiplicative identity. Note that $\mathsf{Mat}_n(\mathbb{R})$ is a non-commutative ring if $n \geq 2$, as matrix multiplication is not commutative.

Matrix rings could be generalized to $\mathsf{Mat}_n(R)$, where we consider the set of $n \times n$ matrices with entries in a ring R. Addition and multiplication of matrices in $\mathsf{Mat}_n(R)$ are defined analogously, substituting our usual addition and multiplication with the corresponding operations in the ring R. The additive identity is the zero matrix, which has 0_R in every entry; the multiplicative identity is the $n \times n$ identity matrix, which has 1_R on the main diagonal, and 0_R elsewhere.

1.12 Example (*Polynomial ring over* $\mathbb{R}$)

A **polynomial** over $\mathbb{R}$ is an expression in x (or any other indeterminant) of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where $n \in \mathbb{N}$ and $a_0, \ldots, a_n \in \mathbb{R}$. The smallest natural number n such that

$a_n \neq 0$ is called the **degree** of $p(x)$. The collection of all polynomials over $\mathbb{R}$, denoted by $\mathbb{R}[x]$, is a ring under addition and multiplication of polynomials. The constant zero polynomial serves as the additive identity; the constant 1 polynomial serves as the multiplicative identity.

1.13 Example (*Polynomial ring over R*)
Let R be a ring. The **polynomial ring over** R with indeterminate x, denoted by $R[x]$, is the collection of all polynomials of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where $n \in \mathbb{N}$ and $a_0, \ldots, a_n \in R$. This generalized notion of polynomials also defines a ring under addition and multiplication, with the polynomial $0(x) = 0_R$ as the additive identity and the polynomial $1(x) = 1_R$ as the multiplicative identity. We will consider the polynomial rings $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ quite often in Math 330-2 and Math 330-3.

1.14 Example (*Rings of functions*)
Let S be a set, and let R be a ring. Consider the set R^S the set of all functions $f\colon S \to R$, and define addition and multiplication on R^S pointwise, that is, for all $f, g \in R^S$, define

$$(f+g)(x) = f(x) + g(x), \text{ and } (fg)(x) = f(x)g(x)$$

for all $x \in S$. We leave it as an exercise here to verify the ring axioms.

1.15 Definition (*Direct product of rings*)
Let $R_1, R_2, \ldots, R_n$ be rings. The **direct product** of $R_1, R_2, \ldots, R_n$ is the ring

Class 3
01/07/2022

$$\prod_{k=1}^{n} R_k = R_1 \times R_2 \times \cdots \times R_n = \{(a_1, a_2, \ldots, a_n) \mid a_k \in R_k\}$$

with addition and multiplication defined pointwise, that is,

$$(a_k) + (b_k) = (a_k + b_k), \text{ and } (a_k)(b_k) = (a_k b_k).$$

1.2 Units and zero divisors

In a ring R, the additive structure is richer than the multiplicative structure, distinctively with the existence of an additive inverse for every element. Yet, we can still attempt to find the multiplicative inverses, a.k.a. to do divisions, and we can *sometimes* succeed.

1.16 Definition (*Units*)
Let R be a non-trivial ring. An element $u \in R$ is a **unit** if there exists $v \in R$ such that $uv = vu = 1$.

1.17 Example
In the ring of integers $\mathbb{Z}$, the only units are ± 1.

1.18 Example
In each of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, all non-zero elements are units, as $x \cdot (1/x) = 1$ for every $x \neq 0$, and $1/x$ is in the ring that we started with, which is not true for $\mathbb{Z}$.

1.19 Example
In $\mathbb{Z}_n$, the units are all integers $1 \leq k \leq n-1$ that satisfies $\gcd(n,k) = 1$. See Theorem 1.11 in Math 330-1 notes.

1.20 Example
In the space $\mathsf{Mat}_n(\mathbb{R})$ of $n \times n$ real matrices, the units are precisely the invertible matrices, collectively forming the general linear group $\mathsf{GL}_n(\mathbb{R})$.

1.21 Example (*Gaussian integers*)
Consider the ring $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$. Elements of this ring are called **Gaussian integers**. A number $a + bi \in \mathbb{Z}[i]$ is a unit if and only if there is some $c + di \in \mathbb{Z}[i]$ such that $(a + bi)(c + di) = 1$. Because $\mathbb{Z}[i] \subseteq \mathbb{C}$, it inherits the absolute value of complex numbers. Thus $|(a+bi)(c+di)| = |a+bi||c+di|$. But the absolute value of a non-zero Gaussian integer is at least 1. Thus only ± 1 and $\pm i$ are candidates for units in $\mathbb{Z}[i]$. It turns out that they are the only units in $\mathbb{Z}[i]$.

1.22 Notation
If $u \in R$ is a unit, then associativity and unitality of the multiplicative structure guarantees that there is a unique $v \in R$ such that $uv = vu = 1$ (see Theorem 3.8 of Math 330-1 notes). We denote this v by u^{-1} to indicate that it is the multiplicative inverse of u.

All examples above give evidence suggesting that the collection of all units in a ring will form a multiplicative group. We prove this lemma here.

1.23 Lemma
Let R be a non-trivial ring. The subset

$$R^* := \{a \in R \mid a \text{ is a unit}\}$$

forms a group under the multiplicative operation of the ring.

PROOF
Associativity of elements in R^* inherits from the monoid structure of R.

Because R is non-trivial, $1 \in R$, and $1 \cdot 1 = 1$, thus $1 \in R^*$. Further, $1 \cdot a = a \cdot 1 = a$ for all $a \in R^*$. So $1 \in R^*$, is the identity element in R^*.

Let $a \in R^*$. Then there exists $b \in R$ such that $ab = ba = 1$. Further, $b \in R^*$ by definition. Thus b is the inverse of a in R^*.

Finally, let $a, b \in R^*$. Then they have multiplicative inverses a^{-1} and b^{-1}, respectively. Thus $(ab)(b^{-1}a^{-1}) = 1$, and $ab \in R^*$, showing R^* is closed. □

1.24 Definition (*Fields*)
Let R be a non-trivial ring. The subset R^* as defined in Lemma 1.23 is the **group of units** of R. A nontrivial commutative ring R is a **field** if $R^* = R \backslash \{0\}$.

1.25 Example
From Examples 1.17 and 1.18, we see that $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all fields, while $\mathbb{Z}$ is not a field.

1.26 Example
From Example 1.19, we see that $\mathbb{Z}_n$ is a field when $\gcd(k, n) = 1$ for all integers $1 \leq k \leq n-1$. This happens precisely when n is a prime integer. Thus $\mathbb{Z}_p$ is a field for every prime integer p. We call $\mathbb{Z}_p$ the **finite field of order** p.

1.27 Definition (*Zero divisors and integral domains*)
Let R be a ring. An non-zero element $a \in R$ is a **zero divisor** if there exists a non-zero $b \in R$ such that either $ab = 0$ or $ba = 0$. A nontrivial commutative ring R is an **integral domain** if it contains no zero divisors.

1.28 Example
For a and b in a number ring, if $a \neq 0$ and $b \neq 0$, then $ab \neq 0$. This shows number rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, as well as $\mathbb{Z}[i]$, have no zero divisors, and all of them are integral domains.

1.29 Example
Consider the ring $\mathbb{Z}_n$. We show that $a \in \mathbb{Z}_n$ is a zero divisor if $\gcd(a, n) > 1$. Let $d = \gcd(a, n) > 1$. Then $a = kd$ and $n = \ell d$ for some $k, \ell \in \mathbb{Z}$. By assumption, $d > 1$, thus $0 < \ell < n$. Thus $a\ell = kd\ell = kn \equiv 0 \pmod{n}$. Together with Example 1.19, this shows that every non-zero element in $\mathbb{Z}_n$ is either a unit or a zero divisor.

1.30 Theorem (*Polynomial rings over integral domains*)
If R is an integral domain, then $R[x]$ is also an integral domain.

Class 4
01/10/2022

PROOF
Let $f(x), g(x) \in R[x]$, and assume that $f(x) \neq 0$, $g(x) \neq 0$. Write

$$f(x) = \sum_{i=0}^{m} a_i x^i, \text{ and } g(x) = \sum_{j=0}^{n} b_j x^j,$$

where m and n are the degrees of $f(x)$ and $g(x)$, respectively, forcing $a_m \neq 0$ and $b_n \neq 0$. Thus

$$\begin{aligned} f(x)g(x) &= \sum_{i=0}^{m}\sum_{j=0}^{n} a_i b_j x^{i+j} = \sum_{k=0}^{m+n}\sum_{i+j=k} a_i b_j x^{i+j} \\ &= (\text{lower order terms}) + a_m b_n x^{m+n}. \end{aligned}$$

Because R is an integral domain, $a_m b_n \neq 0$, so that $f(x)g(x) \neq 0$. Thus $R[x]$ is an integral domain. □

1.31 Remark
This proof also shows that if $f(x), g(x) \in R[x]$ are non-zero polynomials, then

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

1.32 Theorem (*Cancellation property in an integral domain*)
Let R be an integral domain, and let $a, b, c \in R$. If $ab = ac$ and $a \neq 0$, then $b = c$.

PROOF
Suppose $ab = ac$ with $a \neq 0$. Then $a(b - c) = ab - ac = 0$. But $a \neq 0$ is not a zero divisor, so we must have $b - c = 0$. Thus $b = c$. □

1.33 Corollary
Let R be an integral domain, and let $a, b \in R$. If $ab = 0$, then either $a = 0$ or $b = 0$.

PROOF
The corollary follows by letting $c = 0$ in Theorem 1.32. □

We will now investigate the relation between integral domains and fields.

1.34 Theorem
Every field is an integral domain.

PROOF
Let F be a field, and $a, b \in F$ with $a \neq 0$ and $ab = 0$. Then a has a multiplicative inverse a^{-1}. Thus

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

This shows a is not a zero divisor. But $a \neq 0$ is arbitrarily chosen in F. Thus F has no zero divisors. □

1.35 Remark
Theorem 1.34 allows us to extend the chain of inclusion relations in Remark 1.6:

$$\text{rngs} \supseteq \text{rings} \supseteq \text{commutative rings} \supseteq \text{integral domains} \supseteq \text{fields}$$

Most of our discussion will happen close to the right end of this chain—the majority of rings that we will study in Math 330-2 are integral domains.

1.36 Theorem
Every finite integral domain is a field.

PROOF
Let R be a finite integral domain, and let $a \in R \backslash \{0\}$. Consider the subset $\{a^k \mid k \in \mathbb{N}\}$ of R. Because R is finite, there must exist $m, n \in \mathbb{N}$ such that $a^m = a^n$. Without loss of generality, assume that $m > n$. By the cancellation property in an integral domain, $a^{m-n} = 1$, with $m - n \geq 1$. Thus a is a unit, hence R is a field. □

2 The structure of rings

2.1 Subrings and ideals

Class 5
01/12/2021

Recall from Math 330-1 that a *subgroup* of a group G is a subset H of G that is also a group under the same binary operation of G. We define *subrings* analogously.

2.1 Definition (*Subrings*)
Let R be a ring. A **subring** of R is a subset $S \subseteq R$ that is a ring under the (inherited, restricted) ring operations of R. We write $S \leq R$ to denote that S is a subring of R.

2.2 Example
Here is a chain of subrings $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$. Note that we can also insert the Gaussian integers in between $\mathbb{Z}$ and $\mathbb{C}$ to form another chain $\mathbb{Z} \leq \mathbb{Z}[i] \leq \mathbb{C}$.

2.3 Example
It is worth emphasizing (probably again) that $\mathbb{Z}_n$ is not a subring of $\mathbb{Z}$. It is not even a subset of $\mathbb{Z}$. Nominally, it may seem to be the case that elements in $\mathbb{Z}_n$ are also integers. But this is far from reality—the elements of $\mathbb{Z}_n$ are *integer classes modulo* n, not integers. We may write $k \in \mathbb{Z}_n$. However, this k does not represent the integer k, but rather the *equivalence class* of $\mathbb{Z}$ modulo n that contains k, that is, the subset $\{\dots, k-2n, k-n, k, k+n, k+2n, \dots\}$.

2.4 Example
If R is a ring, then $R \leq R[x]$ as constant polynomials.

2.5 Unlike subgroups, subrings do not capture the interesting substructures within rings in general. For example, the only subrings of $\mathbb{Z}$ are the trivial subring $\{0\}$ and $\mathbb{Z}$ itself: if we want a non-trivial subring of $\mathbb{Z}$, it must contain 1, which generates the entire ring $\mathbb{Z}$ (as an abelian group).

We appeal to a different type of substructure within rings. Recall that $n\mathbb{Z}$ are normal subgroups of $\mathbb{Z}$. Most of them do not form subrings, because they do not contain 1 if $n \neq \pm 1$. But they are rngs—they are closed under addition, subtraction, and multiplication. This leads to the definition of an *ideal* in a ring, which plays a similar role in ring theory as normal subgroups in group theory.

2.6 Definition (*Ideals*)
Let R be a ring. A **left ideal** of R is a an additive subgroup $I \leq R$ that is closed under left multiplication by elements in R, that is, $ax \in I$ for all $a \in R$ and $x \in I$. A **right ideal** of R is an additive subgroup $I \leq R$ that is closed under right multiplication by elements in R, that is, $xa \in I$ for all $a \in R$ and $x \in I$. A **two-sided ideal**, or simply an **ideal**, of R is an additive subgroup $I \leq R$ that is both a left ideal and a right ideal of R.

2.7 Remark
For commutative rings, the notions of left, right, and two-sided ideals coincide,

and we simply use the term *ideal* alone.

2.8 Example
$n\mathbb{Z}$ is an ideal of $\mathbb{Z}$. Integer multiples of n are closed under addition, subtraction, and multiplication by integers.

2.9 Example (*Trivial ideal*)
If R is a ring, then the trivial subring $\{0\}$ is also an ideal of R, called the **trivial ideal** of R. Indeed, $0a = a0 = 0$ for all $a \in R$, so it is closed under left and right multiplication by elements in R.

2.10 Example
Consider the polynomial ring $R[x]$, where R is a nontrivial ring. Let

$$I = \{f \in R[x] \mid f(0) = 0\}.$$

Then I is an ideal of $R[x]$. Let $f, g \in I$ and let $h \in R[x]$. Then

$$f(0) - g(0) = 0 - 0 = 0.$$

Thus $f - g \in I$, and $I \leq R[x]$ is a additive subgroup. Further,

$$f(0)h(0) = 0h(0) = 0, \text{ and } h(0)f(0) = h(0)0 = 0.$$

Thus $fh, hf \in I$, and I is an ideal of $R[x]$.

2.11 Example
A left ideal is not necessarily a right ideal in a non-commutative ring. Consider the matrix ring $\mathsf{Mat}_2(\mathbb{R})$, and consider the subset

$$L = \left\{ \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} \,\middle|\, x, y \in \mathbb{R} \right\}.$$

L is clearly an additive subgroup of $\mathsf{Mat}_2(\mathbb{R})$ with entry-wise addition. L is also closed under left multiplication by any 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} = \begin{bmatrix} ax + by & 0 \\ cx + dy & 0 \end{bmatrix} \in L,$$

so L is a left ideal of $\mathsf{Mat}_2(\mathbb{R})$. But L is not closed under right multiplication by 2×2 matrices. For example,

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \notin L,$$

so L is not a right ideal of $\mathsf{Mat}_2(\mathbb{R})$.

2.12 Unlike the relation between subgroups and normal subgroups, subrings and ideals are two different classes of substructures within a ring. In fact, the only ideals that are also subrings are the trivial ideal and the ring itself.

2.13 Lemma *(Ideals containing a unit)*
Let R be a ring, and let I be an ideal of R. If $u \in I$ is a unit in R, then $I = R$.

PROOF
We have $I \subseteq R$ by definition. Suppose that $u \in I$ is a unit in R. Then $vu = 1$ for some $v \in R$. Thus $1 \in I$ because I is an ideal. Let $a \in R$. Then $a1 = 1a = a \in I$ because I is an ideal. Thus $R \subseteq I$. We can conclude that $I = R$. □

2.14 Corollary *(Ideals of a field)*
If F is a field, then the only ideals of F are $\{0\}$ and F itself.

PROOF
Let $I \subseteq F$ be an ideal of F. Suppose that $a \in I$ and $a \neq 0$. Then a is a unit, and thus $I = F$ by Lemma 2.13. □

2.15 Definition *(Ideals generated by subsets)*
Let R be a ring, and let $A \subseteq R$ be a subset of R. The **left ideal generated by** A is the collection of all finite left R-linear combinations of elements in A, that is,

$$\{r_1a_1 + \cdots + r_na_n \mid n \in \mathbb{N},\ r_1, \ldots, r_n \in R,\ a_1, \ldots a_n \in A\}.$$

Likewise, the **right ideal generated by** A is the collection of all finite right R-linear combinations of elements in A, that is,

$$\{a_1s_1 + \cdots + a_ns_n \mid n \in \mathbb{N},\ s_1, \ldots, s_n \in R,\ a_1, \ldots a_n \in A\}.$$

The **(two-sided) ideal generated by** A, denoted by (A), is the collection of all finite two-sided R-linear combinations of elements in A, that is,

$$\{r_1a_1s_1 + \cdots + r_na_ns_n \mid n \in \mathbb{N},\ r_1, \ldots, r_n, s_1, \ldots, s_n \in R,\ a_1, \ldots a_n \in A\}.$$

Note that the elements $a_1, \ldots a_n \in A$ in these definitions need not to be distinct.

For a commutative ring R, the notions of the left, right, and two-sided ideals generated by a subset A coincide. In this case, the convention is to take the definition of the left ideal generate by A as the definition.

A lemma is due here to show that (A) is indeed an ideal of R.

2.16 Lemma
Let R be a ring, and let $A \subseteq R$. Then (A) is an ideal of R.

PROOF
Let $x, y \in (A)$. We can write

$$x = r_1a_1s_1 + \cdots + r_ma_ms_m, \text{ and } y = p_1b_1q_1 + \cdots + p_nb_nq_n,$$

where $r_i, s_i, p_j, q_j \in R$, and $a_i, b_j \in A$. Then

$$x - y = r_1a_1s_1 + \cdots + r_ma_ms_m - p_1b_1q_1 - \cdots - p_nb_nq_n \in (A).$$

Thus $(A) \leq R$ is an additive subgroup by the one-step subgroup test.

Now let $z \in R$, then

$$zx = z(r_1a_1s_1 + \cdots + r_ma_ms_m) = (zr_1)a_1s_1 + \cdots + (zr_m)a_ms_m \in (A)$$
$$xz = (r_1a_1s_1 + \cdots + r_ma_ms_m)z = r_1a_1(s_1z) + \cdots + r_ma_m(s_mz) \in (A)$$

Thus (A) is an ideal of R. □

2.17 Example
In any ring R, we have $(0) = \{0\}$, and $(1) = R$.

2.18 Example
Consider the number ring $\mathbb{Z}$. The ideal generated by an integer (n) coincides with the subgroup generated by the same integer $\langle n \rangle = n\mathbb{Z}$.

2.19 Example
Given a ring R and an element $a \in R$, it is not always true that the ideal generated by a coincides with the subgroup generated by a, considering only the additive group structure of R in the latter case. For example, consider the ring $\mathbb{R}$. The subgroup generated by $2 \in \mathbb{R}$ contains all integer multiples of 2. Thus $\langle 2 \rangle = 2\mathbb{Z}$. However, $\mathbb{R}$ is a field, so (2) is either trivial or $\mathbb{R}$. But $2 \in (2)$ is a unit, so it must be the case that $(2) = \mathbb{R}$.

2.20 Example
Consider the number ring $\mathbb{Z}$ again. Let $a, b \in \mathbb{Z}$. By Bézout's identity (see Theorem 1.10 of Math 330-1 notes), $d = \gcd(a, b) = ma + nb$. Thus $d \in (a, b)$. Because $\gcd(a, b)$ is the smallest positive integer that could be written as an integer linear combination of a and b, we have $(a, b) = (d)$. In particular, if a and b are relatively prime, $(a, b) = \mathbb{Z}$.

Class 6
01/14/2022

2.21 Definition *(Finitely generated and principal ideals)*
An ideal I of R is **finitely generated** if $I = (a_1, \ldots, a_n)$ for a finite list of $a_1, \ldots, a_n \in R$. An ideal I of R is a **principal ideal** if $I = (a)$ for some $a \in R$.

2.22 Definition *(Principal ideal domains)*
A **principal ideal domain** (abbreviated as **PID**) is an integral domain whose ideals are all principal.

2.23 Example
Corollary 2.14 and Example 2.17 together tells us that every field is a PID.

2.24 Example
In the number ring $\mathbb{Z}$, Example 2.20 tells us that any ideal of the form (a, b) is principal: $(a, b) = (\gcd(a, b))$. In fact, $\mathbb{Z}$ is a principal ideal domain: Let I be an ideal of $\mathbb{Z}$. Then $I \leq \mathbb{Z}$ as an additive subgroup. But $\mathbb{Z}$ is cyclic, so I must be cyclic as well (see Theorem 5.11 of Math 330-1 notes). Thus $I = \langle d \rangle$ as a cyclic subgroup of $\mathbb{Z}$, where $d \in \mathbb{Z}$. But $\langle d \rangle = d\mathbb{Z} = (d)$. So I is a principal ideal of $\mathbb{Z}$.

2.25 Example
The ideal $(2, x)$ in $\mathbb{Z}[x]$ is not a principal ideal. First note that a polynomial in

$(2, x)$ is of the form $2p(x) + xq(x)$ for some $p(x), q(x) \in \mathbb{Z}$. In particular, it must have an even constant term. Assume that $(2, x) = (a(x))$ for some $a(x) \in \mathbb{Z}[x]$. Then $2 \in (a(x))$. Thus $2 = p(x)a(x)$ for some $p(x) \in \mathbb{Z}[x]$. By Remark 1.31, $\deg(p(x)) + \deg(a(x)) = \deg(2) = 0$. Thus both $p(x)$ and $a(x)$ must be constant polynomials in $\mathbb{Z}[x]$, *i.e.*, integers. But 2 is a prime integer, then one of $p(x)$ and $a(x)$ must be ± 1, and the other must be ± 2. If $a(x) = \pm 1$, then $(a(x)) = \mathbb{Z}[x]$ by Lemma 2.13. But $(2, x)$ does not contain all polynomials in $\mathbb{Z}[x]$, so it must be the other case, where $a(x) = \pm 2$. But now $x \in (\pm 2)$, so that $x = 2q(x)$ for some $q(x) \in \mathbb{Z}[x]$, which is not possible. The contradiction shows that $(2, x)$ is not a principal ideal of $\mathbb{Z}[x]$.

This tells us that although $\mathbb{Z}$ is a principal ideal domain, $\mathbb{Z}[x]$ is not a principal ideal domain (cf. Theorem 1.30).

2.26 Remark

Principal ideal domains can be inserted into the chain of inclusion relations in Remark 1.35 in between integral domains and fields.

$$\text{rngs} \supseteq \text{rings} \supseteq \text{commutative rings} \supseteq \text{integral domains} \supseteq \text{PIDs} \supseteq \text{fields}$$

2.2 Ring homomorphisms

Just like normal subgroups correspond to kernels of group homomorphisms, ideals of rings correspond to kernels of ring homomorphisms. This section is analogous to Chapter 4 of Math 330-1 notes on group homomorphisms.

2.27 Definition (*Ring homomorphisms*)

Let R and S be rings. A **(ring) homomorphism** is a function $f\colon R \to S$ that satisfies

(1) $f(a +_R b) = f(a) +_S f(b)$ for all $a, b \in R$,

(2) $f(a \cdot_R b) = f(a) \cdot_S f(b)$ for all $a, b \in R$, and

(3) $f(1_R) = 1_S$.

A **(ring) isomorphism** is a bijective ring homomorphism. Two rings R and S are said to be **isomorphic**, denoted by $R \cong S$ if there exists an isomorphism between R and S.

2.28 Remark

Condition (1) in Definition 2.27 is the same condition for f to be a group homomorphism between $(R, +_R)$ and $(S, +_S)$. Conditions (2) and (3) together give the conditions for f to be a **monoid homomorphism** between $(R, \cdot_R)$ and $(S, \cdot_S)$. Note that Fraleigh's textbook does not require rings to have 1, thus it omits condition (3).

2.29 Example (*Identity homomorphisms*)

If R is a ring, then the function $\iota\colon R \to R$ defined by $\iota(a) = a$ for all $a \in R$ is a ring homomorphism called the **identity homomorphism**.

2.30 Example (*Projection homomorphism*)
Consider the direct product of rings $R_1, R_2, \ldots, R_n$. For each k, the map

$$\pi_k \colon R_1 \times R_2 \times \cdots \times R_n \to R_k$$

defined by $\pi_k(a_1, a_2, \ldots, a_n) = a_k$ is a ring homomorphism called the **projection homomorphism** onto the k-th component.

2.31 Example (*Reduction homomorphism*)
Let $n \geq 2$ be an integer. Define $f \colon \mathbb{Z} \to \mathbb{Z}_n$ by setting $f(a)$ to be the remainder of a modulo n for all $a \in \mathbb{Z}$. Addition and multiplication are preserved by congruence. Further, the remainder of 1 modulo n is 1 itself.

2.32 Example (*Evaluation homomorphism*)
Let R and S be rings that satisfies $R \leq S$. Let $a \in S$. Then the **evaluation homomorphism** of the polynomial ring $R[x]$ at a is the function $\varphi_a \colon R[x] \to S$ defined by $\varphi_a(f(x)) = f(a)$ for all $f(x) \in R[x]$. We verify that φ_a is indeed a ring homomorphism. Let $f(x), g(x) \in R[x]$. Then

Class 7
01/19/2022

- $\varphi_a(f(x) + g(x)) = f(a) + g(a) = \varphi_a(f(x)) + \varphi_a(g(x))$,
- $\varphi_a(f(x)g(x)) = f(a)g(a) = \varphi_a(f(x))\varphi_a(g(x))$, and
- $\varphi_a(1(x)) = 1(a) = 1$.

2.33 Lemma
If $f \colon R \to S$ is a ring homomorphism, then

(1) $f(0_R) = 0_S$,
(2) $f(-a) = -f(a)$ for all $a \in R$,
(3) $f(1_R) = 1_S$,
(4) if $a \in R$ is a unit, then so is $f(a) \in S$, and $f(a)^{-1} = f(a^{-1})$, and

PROOF
(1) and (2) follows from the fact that f is an additive group homomorphism between R and S. (3) is an axiom of ring homomorphisms. Thus the only statements we need to prove here is (4).

Suppose that $a \in R$ is a unit. Then it has a multiplicative inverse $a^{-1} \in R$ such that $aa^{-1} = a^{-1}a = 1_R$. Thus $f(a)f(a^{-1}) = f(aa^{-1}) = f(1_R) = 1_S$. Likewise, $f(a^{-1})f(a) = 1_S$. □

2.34 Remark
Zero divisors are not necessarily preserved under ring homomorphisms. Consider the projection homomorphism $\pi_1 \colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$. Every element of the form $(a, 0)$ with $a \neq 0$ is a zero divisor in $\mathbb{Z} \times \mathbb{Z}$, but $\pi_1(a, 0) = a$ is not a zero divisor in $\mathbb{Z}$. For the other direction, consider the reduction homomorphism $f \colon \mathbb{Z} \to \mathbb{Z}_4$. We have $2 \in \mathbb{Z}$ not a zero divisor, but $f(2) = 2 \in \mathbb{Z}_4$ a zero divisor.

2.35 Definition (*Images and kernels*)
Let $f \colon R \to S$ be a ring homomorphism. The **image** of f, denoted by $\mathrm{im}(f)$, is

the subset

$$\operatorname{im}(f) = \{f(a) \mid a \in R\} \subseteq S.$$

The **kernel** of f, denoted by $\ker(f)$, is the subset

$$\ker(f) = \{a \in R \mid f(a) = 0\} \subseteq R.$$

2.36 Example

The images and kernels of the homomorphisms in Examples 2.29 to 2.31 are

Identity homomorphism: $\operatorname{im}(f) = R$, $\ker(f) = \{0\}$.

Projection homomorphism: $\operatorname{im}(\pi_k) = R_k$, $\ker(\pi_k) = \{(a_i)_{i=1}^n \mid a_k = 0\}$.

Reduction homomorphism: $\operatorname{im}(f) = \mathbb{Z}_n$, $\ker(f) = n\mathbb{Z}$.

2.37 Lemma

If $f\colon R \to S$ is a ring homomorphism, then $\operatorname{im}(f)$ is a subring of S, and $\ker(f)$ is an ideal of R.

Proof

First, $\operatorname{im}(f) \leq (S, +)$ and $\ker(f) \leq (R, +)$ as additive subgroups because f is a homomorphism between abelian groups. Additionally,

- $1_S = f(1_R)$, and if $x, y \in \operatorname{im}(f)$, then $x = f(a)$ and $y = f(b)$ for some $a, b \in R$. Thus $xy = f(a)f(b) = f(ab) \in \operatorname{im}(f)$.
- If $a \in R$ and $x \in \ker(f)$, then $f(ax) = f(a)f(x) = f(a)0 = 0$. Likewise, $f(xa) = f(x)f(a) = 0f(a) = 0$. Thus $ax, xa \in \ker(f)$.

Hence, $\operatorname{im}(f)$ is a subring of S, and $\ker(f)$ is an ideal of R. □

2.38 Example

The image and kernel of the evaluation homomorphism in Example 2.32 is more complicated, and thus yields a lot of interesting results.

- The image is

$$\begin{aligned}\operatorname{im}(\varphi_a) &= \{f(a) \mid f(x) \in R[x]\} \\ &= \{c_0 + c_1 a + c_2 a^2 + \cdots + c_n a^n \in S \mid n \in \mathbb{N}, c_0, \ldots c_n \in R\}.\end{aligned}$$

  This is the set of all finite R-linear combinations of the powers of $a \in S$, oftentimes denoted by $R[a]$. Note that this is not a set of polynomials anymore: its elements are in the larger ring S.

  An important class of examples is given by $\mathbb{Z}[\sqrt{n}]$, where $n \in \mathbb{Z}$ is a *squarefree* integer—that is, all primes in its prime factorization appear only once. Some of these rings do not satisfy *unique factorization*. For example, $6 \in \mathbb{Z}[\sqrt{-5}]$ can be expressed as a product of *prime* elements in $\mathbb{Z}[\sqrt{-5}]$ in two different ways:

$$6 = 2 \cdot 3 \text{ and } 6 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

  We will study *unique factorization domains* in the coming weeks.

- The kernel is
$$\ker(\varphi_a) = \{f(x) \in R[x] \mid f(a) = 0\}.$$
This is precisely the set of all polynomials in $R[x]$ with a as a root. In the case that R is a field, we will show at a later point that $x - a$ is a factor of $f(x)$ in $S[x]$, allowing us to further conclude that $\ker(\varphi_a) = (x-a) \cap R[x]$.

2.3 Quotient rings

We are now ready to define the analogue of quotient groups in ring theory. This section compares to Chapter 7 of Math 330-1 notes.

2.39 Definition (*Cosets*)
Let R be a ring, and $I \subseteq R$ be an ideal. A **coset** of I is a subset of R of the form
$$a + I = \{a + x \mid x \in I\} \subseteq R.$$
The element a here is called a **representative** of the coset $a + I$.

2.40 Remark
Cosets of $I \subseteq R$ are precisely the cosets of the normal subgroup $(I, +) \trianglelefteq (R, +)$ as abelian groups. In fact, the notation $a + I$ is the same notation for cosets of $I \trianglelefteq R$ if the additive notation is adopted for abelian groups.

2.41 Definition (*Quotient rings*)
Let R be a ring, and $I \subseteq R$ be an ideal. The **quotient** of R by I, denoted by R/I, is the ring
$$R/I = \{a + I \mid a \in R\},$$
with ring operations defined by
$$(a + I) + (b + I) = (a + b) + I, \text{ and } (a + I)(b + I) = (ab) + I$$
Such a ring R/I is called a **quotient ring**, or a **factor ring**.

The following lemma verifies that the operations in a quotient ring are *well-defined*.

2.42 Lemma
Let R be a ring, and $I \subseteq R$ be an ideal. If $c \in a + I$ and $d \in b + I$, then $c + d \in (a + b) + I$, and $cd \in (ab) + I$.

Proof
Addition is well-defined because $I \trianglelefteq R$ is a normal subgroup. We only verify well-defined-ness of multiplication here.

Class 8
01/21/2022

Suppose that $c \in a + I$ and $d \in b + I$. Then $c = a + x$ and $d = b + y$ for some $x, y \in I$. Thus
$$cd = (a + x)(b + y) = ab + ay + xb + xy.$$

Because I is an ideal, $ay, xb, xy \in I$, so is their sum. Thus

$$cd = ab + (ay + xb + xy) \in ab + I. \qquad \square$$

2.43 Example

We have seen in group theory that $\mathbb{Z}/n\mathbb{Z} = \{r + n\mathbb{Z} \mid 0 \leq r \leq n-1,\ r \in \mathbb{Z}\}$ is isomorphic to $\mathbb{Z}_n$ as abelian groups. In fact, they are isomorphic as rings as well, which can be demonstrated by the following theorem.

2.44 Theorem (*First isomorphism theorem for rings*)

If $f\colon R \to S$ is a ring homomorphism, then $R/\ker(f) \cong \operatorname{im}(f)$.

Proof

This is left as an exercise in your homework assignment. $\square$

2.45 Example

Now we return to Example 2.43, and consider the reduction homomorphism $f\colon \mathbb{Z} \to \mathbb{Z}_n$ defined in Example 2.31. It is clearly a surjective homomorphism, with $\ker(f) = n\mathbb{Z}$. Apply the first isomorphism theorem, $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

2.46 Example (*Canonical homomorphism*)

Let R be a ring and let I be an ideal of R. The **canonical homomorphism** is the map $f\colon R \to R/I$ defined by $f(a) = a + I$. This is a natural definition that arises from the structure of the ring and its ideal, hence the name. The canonical homomorphism is clearly surjectivity, and its kernel is precisely the ideal I. The first isomorphism theorem only tells us the trivial result $R/I \cong R/I$.

2.47 Example (*Field extension*)

Consider the evaluation homomorphism $\varphi_i\colon \mathbb{R}[x] \to \mathbb{C}$, given by $\varphi_i(f(x)) = f(i)$ for all $f(x) \in \mathbb{R}[x]$. Then $\ker(\varphi_i)$ consists of all polynomials $f(x) \in \mathbb{R}[x]$ that has i as a root. But because $f(x)$ is real-valued, it must also have $-i$ as a root. Thus both $x + i$ and $x - i$ are factors of $f(x)$, and we can write

$$f(x) = (x + i)(x - i)g(x) = (x^2 + 1)g(x)$$

for some $g(x) \in \mathbb{R}[x]$. It then follows that

$$\ker(\varphi_i) = \{(x^2 + 1)g(x) \mid g(x) \in \mathbb{R}[x]\} = (x^2 + 1).$$

Moreover, φ_i is surjective because $a + bi = \varphi_i(a + bx)$. By the first isomorphism theorem, we have $\mathbb{C} \cong \mathbb{R}[x]/(x^2 + 1)$.

The idea behind this example is more important than the example itself. We know that the polynomial $p(x) = x^2 + 1$ does not have a root in $\mathbb{R}$, so it does not *factor* further. However, if we claim that $i \notin \mathbb{R}$ (or any other symbol) is a root of $p(x)$, then we can "throw i into $\mathbb{R}$" to force $p(x)$ to have a root. In order to maintain ring structure, we also need to have $-i$, bi, and $a + bi$, thus we get all complex numbers $\mathbb{C}$. The more technical term for the action of "throwing i into $\mathbb{R}$" is called *adjoining* i to $\mathbb{R}$, denoted by $\mathbb{R}[i]$ (cf. Example 1.21). It turns out that $p(x)$ is what we call a *minimal polynomial* of $i \in \mathbb{C}$ over $\mathbb{R}$, and adjoining i to $\mathbb{R}$ is the same as taking the quotient $\mathbb{R}[x]/(p(x))$. You will see more of this in Math 330-3, but we must continue to build foundations for now.

2.4 Maximal and prime ideals

Class 9
01/24/2022

So far we have seen many different ideals of rings. For every non-trivial ring R, there are at least the trivial ideal $\{0\}$ and the improper ideal R, and there may be some more interesting ideals of R out in the wild in between these two. We study some of these ideals in this subsection.

2.48 Definition (*Maximal and prime ideals*)
Let R be a commutative ring.

- A **maximal ideal** of R is an ideal $M \neq R$ such that if I is an ideal satisfying $M \subseteq I \subseteq R$, then either $I = M$ or $I = R$.
- A **prime ideal** of R is an ideal $P \neq R$ such that for all $a, b \in R$, if $ab \in P$, then either $a \in P$ or $b \in P$.

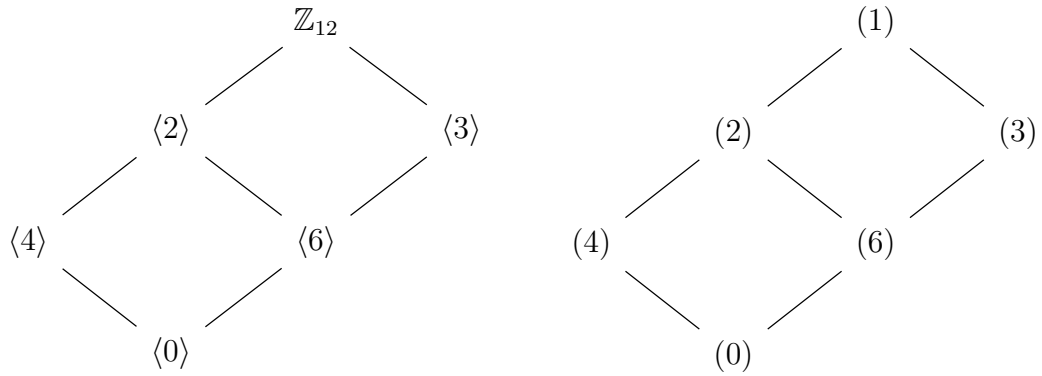
2.49 Example (*Ideals of* $\mathbb{Z}$)
We know that ideals of $\mathbb{Z}$ are of the form $k\mathbb{Z}$ where k is an integer. We also know that $n\mathbb{Z} \subseteq m\mathbb{Z}$ if and only if m divides n.

- To have a maximal ideal $n\mathbb{Z}$, we would need $n \neq \pm 1$ and if m divides n, then either $m = \pm 1$ or $m = \pm n$. This is to say that n is a prime integer. Thus all maximal ideals of $\mathbb{Z}$ are $p\mathbb{Z}$ where p is a prime integer.
- To have a prime ideal $n\mathbb{Z}$, we would need $ab \in n\mathbb{Z}$ to imply either a or b is an integer multiple of n. This is to say that n is either a prime integer or 0. Thus all prime ideals of $\mathbb{Z}$ are $\{0\}$ and $p\mathbb{Z}$ where p is a prime integer.

We see that prime ideals and maximal ideals of $\mathbb{Z}$ (mostly) coincide.

2.50 Example
The trivial ideal $\{0\}$ is a prime ideal in any integral domain R. (Why is it not prime if R is not an integral domain?)

2.51 Example
Consider $\mathbb{Z}_{12}$. The subgroups of $\mathbb{Z}_{12}$ can be organized in the following *subgroup diagram* (cf. Example 5.15 in Math 330-1 notes).



It turns out that all subgroups of $\mathbb{Z}_{12}$ are ideals of $\mathbb{Z}_{12}$ as well (Warning: this is not true in general for any ring R), and the lattice indicates containment

relation: if I_1 is below I_2 with an edge connecting two ideals, then $I_1 \subseteq I_2$. For example, $(6) \subseteq (2)$, and $(4) \nsubseteq (3)$. We can see from this diagram that $\mathbb{Z}_{12}$ has two maximal ideals: (2) and (3).

2.52 Example
The ideal $(x) \subseteq \mathbb{Z}[x]$ is prime, but not maximal. If $p(x), q(x) \in \mathbb{Z}[x]$ such that their product $p(x)q(x)$ is a polynomial with zero constant term, then either $p(x)$ or $q(x)$ must have zero constant term. However, it is not maximal, because $(x) \subsetneq (2, x) \subsetneq \mathbb{Z}[x]$ (see Example 2.25).

2.53 Example
The ideal $(x^2 + 1) \subseteq \mathbb{R}[x]$ is maximal. Let's suppose that there is an ideal I of $\mathbb{R}[x]$ such that $(x^2 + 1) \subsetneq I \subseteq R[x]$. Then there exists $f(x) \in I \backslash (x^2 + 1)$. By polynomial long division (which we will see in more detail in the next chapter), we can find $q(x) \in \mathbb{R}[x]$ and $r(x) = ax + b \neq 0$ such that

$$f(x) = (x^2 + 1)q(x) + r(x).$$

Thus $ax + b = (x^2 + 1)q(x) - f(x) \in I$, and

$$\begin{aligned}(ax + b)(ax - b) &= a^2x^2 - b^2 \in I,\\ a^2(x^2 + 1) &= a^2x^2 + a^2 \in I.\end{aligned}$$

Thus $a^2 + b^2 \in I$. But $a^2 + b^2$ is a unit in $\mathbb{R}[x]$. Hence, $I = \mathbb{R}[x]$ by Lemma 2.13.

The argument above is quite specific to the example. The next two theorems give some nice characterizations of maximal and prime ideals, which in turn provides easily generalized arguments for whether a given ideal is maximal or prime.

2.54 Theorem
Let R be a commutative ring and $I \subseteq R$ be an ideal. Then I is maximal if and only if R/I is a field.

Proof
($\Rightarrow$) Suppose that I is a maximal ideal of R. Let $a + I \in R/I$, with $a \notin I$. Define

$$J = I + (a) = \{x + ra \mid x \in I,\ r \in R\}.$$

Then $I \subsetneq J$. Further, if $x + ra, y + sa \in J$, then

$$(x + ra) - (y + sa) = (x - y) + (r - s)a \in I + (a) = J,$$

and if $t \in R$, then

$$t(x + ra) = tx + (tr)a \in I + (a) = J.$$

Thus J is an ideal of R. Because I is maximal, $J = R$, so that $1 \in J$. This means $1 = x + ra$ for some $x \in I$ and $r \in R$. Thus

$$(r + I)(a + I) = ra + I = (1 - x) + I = 1 + I,$$

which is the multiplicative identity in R/I. This shows that $a+I$ is a unit in R/I, and thus R/I is a field.

($\Leftarrow$) Now suppose that R/I is a field. Let J be an ideal of R that satisfies $I \subsetneq J \subseteq R$. Then there exists $a \in J \backslash I$. Thus $a \notin I$, and $a+I$ is not the additive identity element $I \in R/I$. Thus $a+I \in R/I$ is a unit, and there exists some $b \in R$ such that $ab+I = (a+I)(b+I) = 1+I$. This implies $1-ab \in I \subseteq J$. But $a \in J$, so $ab \in J$ as well, which means $1 = (1-ab)+ab \in J$. Therefore, $J = R$. This shows that I is a maximal ideal of R. □

2.55 Theorem

Let R be a commutative ring and $I \subseteq R$ be an ideal. Then I is prime if and only if R/I is an integral domain.

Class 10
01/26/2022

Proof

Note that the additive identity in R/I is $0+I = I$. Thus we need to show I is prime if and only if $(a+I)(b+I) = I$ implies either $a+I = I$ or $b+I = I$.

($\Rightarrow$) Suppose that I is a prime ideal of R. Let $a, b \in R$ with $(a+I)(b+I) = I$. Then $ab+I = I$, and $ab \in I$. Thus either $a \in I$ or $b \in I$, hence either $a+I = I$ or $b+I = I$. This shows R/I is an integral domain.

($\Leftarrow$) Now suppose that R/I is an integral domain. Then $(a+I)(b+I) = I$ implies either $a+I = I$ or $b+I = I$. Let $a, b \in R$ with $ab \in I$. Then $I = ab+I = (a+I)(b+I)$. Thus either $a+I = I$ or $b+I = I$, hence either $a \in I$ or $b \in I$. This shows I is a prime ideal of R. □

2.56 Corollary

A maximal ideal in a commutative ring is prime.

Proof

Suppose that R is a commutative ring with a maximal ideal I. Then R/I is a field, which is also an integral domain. Thus I is a prime ideal. □

2.57 Example

We know from Example 2.47 that $\mathbb{C} \cong \mathbb{R}[x]/(x^2+1)$. This indicates that (x^2+1) is a maximal ideal of $\mathbb{R}[x]$.

2.58 Example

If p is a prime integer, $\mathbb{Z}/p\mathbb{Z} \cong \mathbb{Z}_p$ is a field (cf. Example 1.26).

3 Factorization in an integral domain

The factorization problem has occurred a handful of times in our previous discussion of integral domains. We will study factorization in more detail in this chapter, specifically addressing the question of whether a given integral domain has the *unique factorization property*—is every element of the integral domain a unique product of "prime" elements? Before we begin, let's recall a few definitions from Homework 2 that are relevant to factorization of elements.

3.1 Definition

Let R be an integral domain, and let $a, b, d \in R$. We say that d is

- a **divisor** of a, denoted by $d \mid a$, if there exists $q \in R$ such that $a = qd$;
- a **common divisor** of a and b if $d \mid a$ and $d \mid b$;
- a **greatest common divisor** of a and b if d is a common divisor of a and b, and every common divisor of a and b is a divisor of d.

3.1 Polynomial rings over a field

We will first study the ring $F[x]$ of polynomials over a field F. We adopt the convention that the degree of the zero polynomial is $-\infty$. To factor a polynomial, we would first need to define the analog of "prime numbers" in the ring of polynomials.

3.2 Definition *(Irreducible polynomials)*

A non-constant polynomial $f(x) \in F[x]$ is **reducible** over F if there exist non-constant $g(x), h(x) \in F[x]$ such that $f(x) = g(x)h(x)$. A non-constant polynomial $f(x) \in F[x]$ is **irreducible** over F if it is not reducible over F, that is, $f(x) \in F[x]$ is irreducible if and only if $f(x) = g(x)h(x)$ for $g(x), h(x) \in F[x]$ implies that either $g(x)$ or $h(x)$ is constant.

3.3 Remark

It is important to note that irreducibility of a polynomial is defined over a particular field. A polynomial may be irreducible over a field F, but may become reducible if we view it over a larger field $E \supseteq F$, as we will demonstrate in the following examples.

3.4 Example

The polynomial $f(x) = x^2 + 1$ is irreducible over $\mathbb{R}$. It cannot be written as a product of two linear factors. However, $f(x)$ is reducible over $\mathbb{C}$, because $f(x) = (x + i)(x - i)$ is the product of two non-constant polynomials in $\mathbb{C}$.

3.5 Example

In the polynomial ring $\mathbb{C}[x]$, the irreducible polynomials are precisely the linear polynomials $f(x) = ax + b$ with $a \neq 0$. This follows from the fundamental theorem of algebra, which tells that every non-constant polynomial in $\mathbb{C}[x]$ has a root $\alpha \in \mathbb{C}$, therefore a linear factor $x - \alpha$.

3.6 Example

"Not having a root" is not a sufficient condition for irreducibility in $\mathbb{R}[x]$. For example, $f(x) = x^4 + 2x^2 + 1 > 0$ for all $x \in \mathbb{R}$, so it does not have a root in $\mathbb{R}$. However, $f(x) = (x^2 + 1)^2$ is reducible.

The following theorem is the analog of the division algorithm in $\mathbb{Z}$ (cf. Theorem 1.1 of Math 330-1 notes). It also justifies the *polynomial long division*, which we will demonstrate in the examples succeeding the theorem.

3.7 Theorem (*Division algorithm for polynomials*)

Let $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Then there exist unique polynomials $q(x), r(x) \in F[x]$, called the **quotient** and the **remainder** of $f(x)$ when divided by $g(x)$, such that $f(x) = g(x)q(x) + r(x)$ and $\deg(r(x)) < \deg(g(x))$.

PROOF

If $f(x) = 0$, then $q(x) = r(x) = 0$ satisfies the conditions. Suppose that $f(x) \neq 0$. Write

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m, \ g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n,$$

with $a_m \neq 0$ and $b_n \neq 0$. Without loss of generality, we can take $a_m = b_n = 1$ because

$$f(x) = g(x)q(x) + r(x) \ \Leftrightarrow \ \frac{f(x)}{a_m} = \frac{g(x)}{b_n}\frac{b_nq(x)}{a_m} + \frac{r(x)}{a_m}$$

and the polynomials $f(x)/a_m$ and $g(x)/b_n$ are monic.

To prove existence, consider the subset

$$S = \{f(x) - g(x)q(x) \mid q(x) \in F[x]\}$$

of $F[x]$. If $0 \in S$, we take $r(x) = 0$ and $q(x)$ that satisfies $f(x) - g(x)q(x) = 0$, so $f(x) = g(x)q(x) + 0$. If $0 \notin S$, we take $r(x)$ to be an element of minimal degree in S, and take $q(x)$ that satisfies $f(x) - g(x)q(x) = r(x)$. Write

$$r(x) = c_0 + c_1x + c_2x^2 + \cdots + c_kx^k$$

with $c_k \neq 0$. We need to show that $k < n$, and we proceed by means of contradiction. Suppose that $k \geq n$. Then

$$f(x) - g(x)q(x) - c_kx^{k-n}g(x) = r(x) - c_kx^{k-n}g(x)$$

The left hand side is $f(x) - g(x)[q(x) - c_kx^{k-n}] \in S$, while the right hand side is

$$r(x) - c_kx^{k-n}g(x) = c_0 + c_1x + \cdots + c_kx^k - (\text{lower degree terms} + c_kx^k),$$

which has degree $< k$, contradicting minimality of $k = \deg(r(x))$ in S.

Now to prove uniqueness, suppose that $q_1(x), r_1(x), q_2(x), r_2(x) \in F[x]$ satisfies that $f(x) = g(x)q_i(x) + r_i(x)$ and $\deg(r_i(x)) < \deg(g(x))$. Then

$$g(x)[q_1(x) - q_2(x)] = r_2(x) - r_1(x)$$

Because $\deg(r_2(x) - r_1(x)) < g(x)$, this can only happen if $q_1(x) - q_2(x) = 0$, and thus $r_1(x) = r_2(x) = 0$. It follows that $q_1(x) = q_2(x)$ and $r_1(x) = r_2(x)$. □

3.8 Example

We find $q(x), r(x)$ in the case $f(x) = x^5 - x^3 + 4x^2 - 5x - 2$ and $g(x) = x^2 - 4x + 2$ in $\mathbb{Z}_7[x]$. To do this, we can use the usual technique of polynomial long division, but being sure to do all of the arithmetic in $\mathbb{Z}_7$. What we obtain is the following:

$$\begin{array}{r|l}
x^2 - 4x + 2 & \begin{array}{rrrrrr}
 & & x^3 & +4x^2 & -x & -1 \\
\hline
x^5 & & -x^3 & +4x^2 & -5x & -2 \\
x^5 & -4x^4 & +2x^3 & & & \\
\hline
 & 4x^4 & -3x^3 & +4x^2 & & \\
 & 4x^4 & -2x^3 & +x^2 & & \\
\hline
 & & -x^3 & +3x^2 & -5x & \\
 & & -x^3 & +4x^2 & -2x & \\
\hline
 & & & -x^2 & -3x & -2 \\
 & & & -x^2 & +4x & -2 \\
\hline
 & & & & & 0
\end{array}
\end{array}$$

so we see that $q(x) = x^3 + 4x^2 - x - 1$ and $r(x) = 0$. In another word, the polynomial $g(x) = x^2 - 4x + 2$ divides $f(x) = x^5 - x^3 + 4x^2 - 5x - 2$ in $\mathbb{Z}_7[x]$. To verify this, note that in $\mathbb{Z}[x]$ we have

$$(x^2 - 4x + 2)(x^3 + 4x^2 - x - 1) = x^5 - 15x^3 + 11x^2 + 2x - 2.$$

Reduce modulo 7, this is indeed equal to $x^5 - x^3 + 4x^2 - 5x - 2$ in $\mathbb{Z}_7[x]$.

The following factor theorem has been taken for granted since high school for polynomials in $\mathbb{R}[x]$. This is in fact, a consequence of the division algorithm for polynomials.

3.9 Corollary *(Factor theorem)*

Let F be a field and let $f(x) \in F[x]$. An element $a \in F$ is a root of $f(x)$ if and only if $x - a$ divides $f(x)$.

Proof

($\Rightarrow$) Suppose that $f(a) = 0$. By the division algorithm, $f(x) = (x-a)q(x)+r(x)$ with $r(x) = c \in F$ being a constant polynomial. But $c = r(a) = f(a) = 0$, so $f(x) = (x - a)q(x)$, and thus $x - a$ divides $f(x)$.

Class 11
01/28/2022

($\Leftarrow$) Suppose that $x - a$ divides $f(x)$. Then $f(x) = (x - a)q(x)$ for some $q(x) \in F[x]$, and so $f(a) = (a - a)q(a) = 0q(a) = 0$, as desired. □

Recall from Theorem 1.30 that $R[x]$ is an integral domain if R is an integral domain. But if $R = F$ is a field, we can say a bit more about the polynomial ring $F[x]$.

3.10 Theorem (*Polynomial rings over fields are PIDs*)
If F is a field, then $F[x]$ is a principal ideal domain.

PROOF
Let I be a nontrivial ideal of $F[x]$. Let $f(x) \in I$ be arbitrary, and let $g(x) \in I$ be a polynomial with minimal degree $n \geq 0$. By the division algorithm, there exist unique polynomials $q(x), r(x) \in F[x]$ with $\deg(r(x)) < \deg(g(x))$ such that $f(x) = g(x)q(x) + r(x)$. However, $r(x) = f(x) - g(x)q(x) \in I$, so $r(x) = 0$ by minimality of $\deg(g(x)) = n$ in I. Thus $f(x) = g(x)q(x)$. Because every element of I is a multiple of $g(x)$, it follows that $I = (g(x))$ is principal. □

3.11 Example
Theorem 3.10 provides us with a lot more examples of principal ideals beyond the number ring $\mathbb{Z}$: all of $\mathbb{Q}[x]$, $\mathbb{R}[x]$, and $\mathbb{C}[x]$ are principal ideal domains, as well as $\mathbb{Z}_p[x]$ for prime p. Note that $\mathbb{Z}[x]$ is not a principal ideal domain. As we have seen in Example 2.25, the ideal $(2, x)$ in $\mathbb{Z}[x]$ is not principal.

Another consequence of the division algorithm for polynomials is the existence of greatest common divisors.

3.12 Theorem
Let F be a field and let $f(x), g(x) \in F[x]$. Then $f(x)$ and $g(x)$ have a greatest common divisor in $F[x]$.

PROOF
Let $I = (f(x), g(x))$. Because $F[x]$ is a principal ideal domain, $I = (d(x))$ for some $d(x) \in F[x]$. We show $d(x)$ is a greatest common divisor of $f(x)$ and $g(x)$.

Because $(d(x)) = (f(x), g(x))$,

- we have $f(x) = p(x)d(x)$ and $g(x) = q(x)d(x)$ for some $p(x), q(x) \in F[x]$. Thus $d(x)$ is a common divisor of $f(x)$ and $g(x)$.
- we can write $d(x) = a(x)f(x) + b(x)g(x)$ for some $a(x), b(x) \in F[x]$. Assume $d'(x)$ is a common divisor of $f(x)$ and $g(x)$. Then $f(x) = p'(x)d'(x)$ and $g(x) = q'(x)d'(x)$ for some $p'(x), q'(x) \in F[x]$. Thus

$$d(x) = a(x)p'(x)d'(x) + b(x)q'(x)d'(x) = d'(x)[a(x)p'(x) + b(x)q'(x)],$$

  which shows $d'(x)$ divides $d(x)$.

Therefore, $d(x)$ is a greatest common divisor of $f(x)$ and $g(x)$. □

3.13 Remark
It is also worth noting that the greatest common divisor of two polynomials is unique up to multiplication by a constant. In particular, there is a unique monic greatest common divisor.

3.2 Unique factorization domains

We studied the factorization problem in the polynomial ring $F[x]$ extensively in the previous section. Now we make attempts to generalize it to arbitrary integral domains. Note that the entire factorization problem started with the goal to generalize the fundamental theorem of arithmetic that concerns factorization of integers into primes. We shall first review this proof.

3.14 Theorem (*Fundamental theorem of arithmetic*)
An integer $n \geq 2$ can be expressed as a product of primes, and the prime factors are unique up to reordering.

Proof
Let $n \geq 2$. We prove existence and uniqueness separately.

Existence. It is clear that $n = 2$ is a product of primes. This will serve as the base case of the following argument of strong induction. If $n > 2$, and if n is prime, we are done. If $n > 2$, and if n is not prime, then $n = ab$ for some $2 \leq a, b < n$. By the strong induction hypothesis, a and b are products of primes, and so is $n = ab$.

Uniqueness. Suppose that $n = p_1 p_2 \dots p_k$ and $n = q_1 q_2 \dots q_\ell$ are two ways to write n as a product of primes. Then $p_1 \mid (q_1 q_2 \dots q_\ell)$. Because p_1 is prime, we must have $p_1 \mid q_j$ for some j. Without loss of generality, assume this is q_1. But then $q_1 = p_1$ because q_1 is prime as well. Iterating this process, and we will get $k = \ell$ and $p_i = q_j$, possibly with some reordering. □

3.15 Remark
The existence and uniqueness parts in the preceding proof use different properties of prime integers: the existence part uses the fact that primes cannot be expressed as nontrivial products of other integers; the uniqueness parts uses the fact that if a prime divides a product of integers, then it must divide at least one of them. We have been taking it for granted that these are both defining properties of prime integers. In another word, these two properties are equivalent in $\mathbb{Z}$. But it is not obvious that they are equivalent in an arbitrary ring, so they deserve separate definitions in rings.

3.16 Definition (*Irreducible and prime elements*)
Let R be an integral domain, and let $p \in R$ be a non-zero, non-unit element. We say that p is

Class 12
01/31/2022

- **irreducible** if for all $a, b \in R$, if $p = ab$, then either a or b is a unit;
- **prime** if for all $a, b \in R$, if $p \mid ab$, then either $p \mid a$ or $p \mid b$.

3.17 Remark
The definition of prime elements in a ring is closely related to that of prime ideals—to say $p \mid x$ is the same as saying $x = rp$ for some $r \in R$. Thus a non-zero, non-unit $p \in R$ is prime if and only if (p) is a prime ideal.

3.18 Example

The prime elements, as well as the irreducible elements, of $\mathbb{Z}$ are precisely $\pm p$, where p is a positive prime number.

3.19 Example

Consider the polynomial ring $\mathbb{R}[x]$. All linear polynomials $ax + b \in \mathbb{R}[x]$, where $a \neq 0$, is both irreducible and prime.

To see this, let $f(x)g(x) = ax+b$. Then one of $f(x)$ and $g(x)$ must have degree 1, while the other must have degree 0, which is a unit in $\mathbb{R}[x]$.

Now assume that $ax + b$ divides $f(x)g(x)$. Then Corollary 3.9 tells us that $f(\alpha)g(\alpha) = 0$, where $\alpha = -b/a$. Thus either $f(\alpha) = 0$ or $g(\alpha) = 0$, which implies that either $ax + b$ divides $f(x)$ or $ax + b$ divides $g(x)$.

It would be convenient to have irreducibility and primality equivalent in an integral domain, like the case of $\mathbb{Z}$. One direction of this ideal scenario is guaranteed.

3.20 Theorem (*Prime elements are irreducible*)

Let R be an integral domain, and let $p \in R$. If p is prime, then it is irreducible.

PROOF

Suppose that p is prime. Let $p = ab$ for some $a, b \in R$. Then $p \mid ab$, so that either $p \mid a$ or $p \mid b$. Without loss of generality, suppose it is the case that $p \mid a$. Then there exists $r \in R$ such that $a = pr = abr$. Apply cancellation in an integral domain, we see $br = 1$. Thus b is a unit, and so p is irreducible. □

Unfortunately, irreducible elements are not necessarily prime in an arbitrary integral domain, as demonstrated in the following somewhat involved example.

3.21 Example

Consider the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$. We show that 2 is irreducible, but not prime in $\mathbb{Z}[\sqrt{-5}]$.

Suppose that $2 = xy$ for some $x, y \in \mathbb{Z}[\sqrt{-5}]$. Write $x = a + b\sqrt{-5}$ and $y = c + d\sqrt{-5}$ for $a, b, c, d \in \mathbb{Z}$. Consider complex absolute values:

$$\begin{aligned} 4 = |2|^2 = |xy|^2 = xy\overline{xy} &= (a + b\sqrt{-5})(a - b\sqrt{-5})(c + d\sqrt{-5})(c - d\sqrt{-5}) \\ &= (a^2 + 5b^2)(c^2 + 5d^2) = a^2c^2 + 5(b^2c^2 + a^2d^2 + 5b^2d^2). \end{aligned}$$

Thus $b^2c^2 + a^2d^2 + 5b^2d^2 = 0$ and $a^2c^2 = 4$. In particular, $a, c \neq 0$. Thus $b = d = 0$. It then follows that $x = a$, $y = c$, and so $2 = xy$ with $x, y \in \mathbb{Z}$. Thus either $x = \pm 1$ or $y = \pm 1$, showing that 2 is irreducible.

However, 2 is not prime. To see this, note that $2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. Suppose, for a contradiction, that $2 \mid (1 + \sqrt{-5})$. Then $1 + \sqrt{-5} = 2z$ for some $z \in \mathbb{Z}[\sqrt{-5}]$. Consider complex absolute values:

$$6 = |1 + \sqrt{-5}|^2 = |2z|^2 = 4|z|^2.$$

However, $|z|^2 \in \mathbb{Z}$ for $z \in \mathbb{Z}[\sqrt{5}]$. Thus $2 \nmid (1 + \sqrt{-5})$. Likewise, $2 \nmid (1 - \sqrt{-5})$. Thus 2 is not prime in $\mathbb{Z}[\sqrt{-5}]$.

3.22 Remark

?? shows that irreducibility and primality are not equivalent in $\mathbb{Z}[\sqrt{-5}]$. So we cannot produce an analogous fundamental theorem of arithmetic for the ring $\mathbb{Z}[\sqrt{-5}]$ following the argument for $\mathbb{Z}$. It can be shown that all of $2, 3, 1 \pm \sqrt{5} \in \mathbb{Z}[\sqrt{-5}]$ are irreducible, yet $2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$, so any such "fundamental theorem" for $\mathbb{Z}[\sqrt{-5}]$ will fail the uniqueness part.

With some additional conditions on the integral domain, irreducibility can imply primality.

3.23 Theorem (*Irreducible elements are prime in a PID*)

Let R be a principal ideal domain, and let $p \in R$. If p is irreducible, then it is prime.

PROOF

Suppose $p \in R$ is irreducible. We want to show p is prime, or equivalently, (p) is a prime ideal. Let I be an ideal containing (p). Then $I = (a)$ for some $a \in R$. Because $p \in (a)$, we have $p = ra$ for some $r \in R$. Irreducibility of p implies that either r or a is a unit. If a is a unit, then $I = (a) = R$; if r is a unit, then $I = (a) = (p)$. This shows (p) is a maximal ideal, thus a prime ideal. □

3.24 Remark

Notice that we proved a result that is quite a bit stronger than Theorem 3.23 asked—we showed that (p) is maximal for an irreducible $p \in R$, while only a prime ideal was asked. This hints that PID may not be a necessary condition for the equivalence between irreducibility and primality. We shall soon see examples of non-PIDs satisfying this equivalence.

3.25 Definition (*Unique factorization domains*)

Class 13
02/02/2022

An integral domain R is a **unique factorization domain** (abbreviated as **UFD**) if it satisfies

(1) Every non-zero non-unit element $a \in R$ can be expressed as a finite product of irreducible elements, and
(2) The expression is *essentially unique*, meaning that if $p_1 p_2 \dots p_m = q_1 q_2 \dots q_n$ with irreducible p_i and q_j, then $m = n$, and there exists a permutation $\sigma \in S_n$ and units $u_1, \dots, u_n \in R$ such that $q_i = u_i p_{\sigma(i)}$.

3.26 Example

We have made frequent reference to the fact that $\mathbb{Z}$ is a unique factorization domain through the fundamental theorem of arithmetic. For example, we can factor 18 in a few different ways:

$$18 = 2 \cdot 3 \cdot 3 = 3 \cdot 2 \cdot 3 = 3 \cdot (-2) \cdot (-3)$$

But all of them are essentially the same, as they only differ by reordering and multiplication by the units $\pm 1 \in \mathbb{Z}$.

The following lemma shows that the use of "irreducible" in the definition of unique factorization domains is not arbitrary—it is equivalent with "prime."

3.27 Lemma

Let R be a unique factorization domain. An element $p \in R$ is prime if and only if it is irreducible.

Proof

Theorem 3.20 has shown primality implies irreducibility.

Suppose that $p \in R$ is irreducible. Let $a, b \in R$ with $p \mid ab$. Then there exists $c \in R$ such that $ab = cp$. Factor a, b, c into irreducibles by

$$a = a_1 \dots a_m, \quad b = b_1 \dots b_n, \quad c = c_1 \dots c_k.$$

Then we have $a_1 \dots a_m b_1 \dots b_n = c_1 \dots c_k p$. By uniqueness of factorization, one of a_i or b_j must equal up for some unit $u \in R$. Thus either $p \mid a$ or $p \mid b$. □

3.3 Noetherian rings

One of the main results in the study of unique factorization domains is that every principal ideal domain is a unique factorization domain. The standard proof utilizes the technology of *ascending chain condition*, which gives rise to *Noetherian rings*, named after Emmy Noether.

3.28 Definition (*Noetherian rings*)

A ring R is **Noetherian** if it satisfies the **ascending chain condition** (abbreviated **ACC**) on its ideals, that is, any ascending chain of ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

of R eventually terminates, meaning that there exists $n \in \mathbb{N}$ such that $I_m = I_n$ for all $m \geq n$.

3.29 Example

Many of our familiar examples of integral domains are Noetherian, including our dearest friend in ring theory—$\mathbb{Z}$. Note that ideals of $\mathbb{Z}$ has a containment relation $(m) \subseteq (n)$ if and only if $n \mid m$. But an integer cannot be infinitely factored non-trivially, so any ascending chain of ideals must terminate.

3.30 Example

Let R be a ring, and consider the polynomial ring on countably many indeterminants $R[x_1, x_2, \dots]$. Then the ascending chain of ideals

$$(x_1) \subseteq (x_1, x_2) \subseteq (x_1, x_2, x_3) \subseteq \cdots$$

do not terminate. Thus this is an example of a non-Noetherian ring.

The ascending chain condition in integral domains guarantees existence of factorization into irreducible elements, as demonstrated in the following theorem.

3.31 Theorem

Let R be a Noetherian integral domain, and let $a \in R$ be a non-zero, non-unit

element. Then a can be expressed as a product of finitely many irreducible elements in R.

PROOF
We proceed by a contradiction argument. Suppose $a \in R$ is a non-zero, non-unit element that cannot be expressed as a product of finitely many irreducible elements. Then a itself is not irreducible, and we can write $a = bc$, where $b, c \in R$ are non-units, and at least one of b and c cannot be expressed as a product of finitely many irreducible elements. With this observation, we can then construct a sequence $a_1, a_2, a_3, \cdots \in R$ recursively:

- Let $a_1 = a$;
- Let a_{n+1} be a non-unit element such that $a_n = a_{n+1}b$, where b is non-unit and a_{n+1} cannot be expressed as a product of finitely many irreducible elements.

Consider the ascending chain of ideals

$$(a_1) \subseteq (a_2) \subseteq (a_3) \subseteq \cdots$$

The containment $(a_n) \subseteq (a_{n+1})$ is proper for all n. Suppose that $a_{n+1} = a_n c$ for some $c \in R$. Then $a_{n+1} = a_{n+1}bc$. By cancellation, $bc = 1$. But b was assumed to be a non-unit element. Thus $a_{n+1} \notin (a_n)$. Thus we have a non-terminating ascending chain

$$(a_1) \subsetneq (a_2) \subsetneq (a_3) \subsetneq \cdots$$

which contradicts the assumption that R is a Noetherian ring. □

Class 14
02/04/2022

To prove that every PID is a UFD, we would need to take the intermediate step to show every PID is Noetherian. The following equivalent classification of Noetherian rings provides this conclusion readily.

3.32 Theorem
A ring R is Noetherian if and only if every ideal I of R is finitely generated.

PROOF
($\Rightarrow$) Suppose that there is an ideal I of R that is not finitely generated. Then we can construct a sequence $a_1, a_2, a_3, \cdots \in I$ recursively:

- Let $a_1 \in I$;
- Let $a_{n+1} \in I \backslash (a_1, \ldots, a_n)$. Such an a_{n+1} is guaranteed to exist because I is not finitely generated and $(a_1, \ldots, a_n) \subsetneq I$.

Then we have a non-terminating ascending chain of ideals

$$(a_1) \subseteq (a_1, a_2) \subseteq (a_1, a_2, a_3) \subseteq \cdots$$

so R is not Noetherian.

($\Leftarrow$) Suppose that every ideal of R is finitely generated. Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ be an ascending chain of ideals. Then their union $I = \bigcup_{k=1}^{\infty} I_k$ is an ideal of R. To see this, let $x \in I$ and $a \in R$. Then $x \in I_k$ for some $k \geq 0$. Thus $ax \in I_k$.

It follows that $ax \in \bigcup_{k=1}^{\infty} I_k = I$. Likewise, $xa \in I$ if R is not commutative. Additionally, let $x, y \in I$, then there exists k such that $x, y \in I_k$. But I_k is a ideal to begin with, thus $x - y \in I_k \subseteq I$. So I is an ideal of R. In particular, I is finitely generated by assumption. Say $I = (a_1, \ldots, a_n)$. But $\{a_1, \ldots, a_n\} \subseteq I_N$ for some $N \geq 0$. Thus $I_N = I$, and $I_m = I_N = I$ for all $m \geq N$. So R is Noetherian. □

Because all ideals of a PID are principal—they are generated by single elements—we can immediately conclude that all PIDs are Noetherian.

3.33 Corollary (*PIDs are Noetherian*)
If R is a principal ideal domain, then R is Noetherian.

3.34 Theorem (*PIDs are UFDs*)
If R is a principal ideal domain, then R is a unique factorization domain.

Proof
By Corollary 3.33 and Theorem 3.31, every $a \in R$ can be expressed as a product of irreducibles. It remains to show essential uniqueness.

Let $a = p_1 p_2 \ldots p_m = q_1 q_2 \ldots q_n$ be two expressions of a as products of irreducibles. By Theorem 3.23, all these irreducibles are prime. The remaining proof follows a similar argument as that in the fundamental theorem of calculus. By primality of p_1, we have $p_1 \mid q_j$ for some $1 \leq j \leq s$. Without loss of generality, assume $p_1 \mid q_1$. Now by irreducibility of q_1, we have $q_1 = p_1 u_1$ for some unit $u_1 \in R$. Canceling p_1 from both sides, we get

$$p_2 p_3 \ldots p_m = u_1 q_2 q_3 \ldots q_n$$

Iterating this process, we will get $m = n$ and $q_i = p_i u_i$ for units $u_i \in R$, possibly with some reordering of irreducible factors. □

3.35 Example
Let F be a field. By Theorem 3.10, $F[x]$ is a principal ideal domain. Thus $F[x]$ is a unique factorization domain as well.

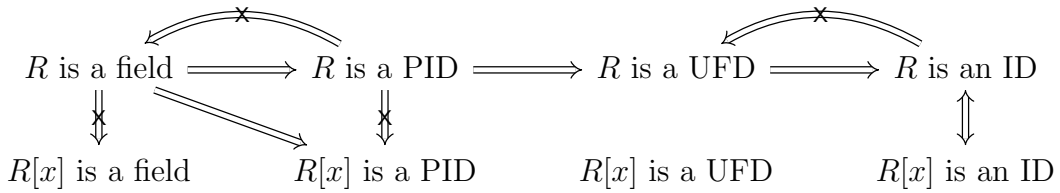
3.36 Example
Example 3.21 shows that $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain. Thus it cannot be a principal ideal domain, either.

3.37 Remark
We shall revisit the chain of inclusions in Remark 2.26, restricted within integral domains. We can now insert UFDs in between integral domains and PIDs.

$$\text{integral domains} \supseteq \text{UFDs} \supseteq \text{PIDs} \supseteq \text{fields}$$

We shall also note that Noetherian rings do not fit well in this chain. Indeed, we have shown that all PIDs are Noetherian. However, there are examples, within integral domains, of UFDs that are not Noetherian, as well as Noetherian rings that are not UFDs. More generally speaking, Noetherian rings are not required to be commutative, so they might not be integral domains at all!

We have also constructed part of the following relations between hierarchy of rings.



There are still some gaps within this diagram. Most notably, we do not yet have an example of a UFD that is not a PID, and "$R[x]$ is a UFD" is disconnected from the main diagram. We will fill in both gaps in the next section.

3.4 Polynomial rings over a UFD

The polynomial ring $\mathbb{Q}[x]$ is known to be a PID, thus a UFD. However, polynomials in $\mathbb{Q}[x]$ are closely related to those in $\mathbb{Z}[x]$—every polynomial in $\mathbb{Q}[x]$ can be expressed as a polynomial in $\mathbb{Z}[x]$ divided by a single integer. It should come as no surprise that properties, especially irreducibility, of polynomials in $\mathbb{Q}[x]$ are related to those in $\mathbb{Z}[x]$.

3.38 Definition (*Primitive polynomials*)
Let R be a unique factorization domain. A non-constant polynomial

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

in $R[x]$ is **primitive** if 1 is a greatest common divisor of $a_0, \ldots, a_n \in R$.

3.39 Example
If $R = F$ is a field, then every non-constant polynomial is primitive. This is due to the fact that $a \mid b$ for all $a, b \in F$ and $a \neq 0$ in a field.

3.40 Example
In $\mathbb{Z}[x]$, the polynomial x^2+4x+4 is primitive, while $4x^2+4x+4$ is not primitive. In particular, monic polynomials in $\mathbb{Z}[x]$ are primitive.

3.41 Theorem (*Gauss's lemma on primitivity*)
Let R be a unique factorization domain, and let $f(x), g(x) \in R[x]$. Then $f(x)g(x)$ is primitive if and only if both $f(x), g(x) \in R[x]$ are primitive.

Class 15
02/07/2022

Proof
($\Rightarrow$) We prove the contrapositive statement. Suppose, without loss of generality, that $f(x)$ is not primitive. Then there exists a non-zero, non-unit $a \in R$ such that $f(x) = ah(x)$. Thus $f(x)g(x) = ag(x)h(x)$ is not primitive.

($\Leftarrow$) Suppose, by means of contradiction, that $f(x)g(x)$ is not primitive with some primitive factors $f(x), g(x) \in R[x]$. Then there exists an irreducible $p \in R$ that divides all coefficients of $f(x)g(x)$. By primitivity of $f(x)$ and $g(x)$, p cannot divide all coefficients of either $f(x)$ or $g(x)$. Let a_kx^k be the first term

(lowest degree term) in $f(x)$ with $p \nmid a_k$, and $b_\ell x^\ell$ be the first term in $g(x)$ with $p \nmid b_\ell$. Then the $k+\ell$ degree term in $f(x)g(x)$ has coefficient

$$\cdots + a_{k-2}b_{\ell+2} + a_{k-1}b_{\ell+1} + a_k b_\ell + a_{k+1}b_{\ell-1} + a_{k+2}b_{\ell-2} + \cdots$$

By primality (which is equivalent to irreducibility in a UFD), $p \nmid a_k b_\ell$. Among all products above, only $a_k b_\ell$ is not divisible by p. Thus p does not divide the coefficient of the $k+\ell$ degree term in $f(x)g(x)$. We have derived a contradiction. Thus $f(x)g(x)$ must be primitive if $f(x)$ and $g(x)$ are primitive. □

3.42 Now if we have a unique factorization domain R, and we want to factor polynomials in $R[x]$, we can pass this task to a larger ring. If $R \leq F$ as a subring of a field F, we can carry a factorization of $f(x) \in R[x] \leq F[x]$ in $F[x]$ back into one in $R[x]$. A canonical choice of this field F is the *field of fractions* of R.

3.43 Definition (*Field of fractions*)
Let R be an integral domain. The **field of fractions** (or **field of quotients**) of R is the field $\mathsf{Frac}(R)$ whose elements are formal quotients $\frac{a}{b}$, where $a, b \in R$ and $b \neq 0$. Further, two quotients $\frac{a}{b}$ and $\frac{c}{d}$ are identified when $ad = bc$. The operations are defined by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}, \quad \text{and} \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

3.44 Remark
We are sweeping a lot of technical details under the rug in Definition 3.43. If we were being rigorous, field of fractions should be defined by $\mathsf{Frac}(R) = (R \times R_{\neq 0})/\sim$, where $\sim$ is the equivalence relation defined by $(a,b) \sim (c,d)$ if and only if $ad = bc$. The elements in $\mathsf{Frac}(R)$ are therefore equivalence classes $[(a,b)]$ under $\sim$. But then the binary operations become class functions, and should be proved to be well-defined. There are also the field axioms that needed to be verified. We will not dwell too much on these technical details, but rather appeal to our understanding of the relation between $\mathbb{Z}$ and $\mathbb{Q}$. Section 21 of the textbook presents a rigorous treatment on field of fractions.

3.45 Remark
In particular, an integral domain R is isomorphic to the subring of $\mathsf{Frac}(R)$ given by $\left\{\frac{a}{1} \mid a \in R\right\}$ with an isomorphism define by $a \mapsto \frac{a}{1}$. As a result, we usually just write a instead of $\frac{a}{1} \in \mathsf{Frac}(R)$. This also means that the additive and multiplicative identities in $\mathsf{Frac}(R)$ are 0 and 1 from R, respectively.

3.46 Example
Evidently, $\mathsf{Frac}(\mathbb{Z}) = \mathbb{Q}$. In fact, from the set theoretic point of view, $\mathbb{Q}$ is *defined to be* the field of fraction of $\mathbb{Z}$.

3.47 Example (*Gaussian rational numbers*)
$\mathsf{Frac}(\mathbb{Z}[i])$ is (isomorphic to) the set $\mathbb{Q}[i]$ of Gaussian rational numbers. To see

this, let $a+bi, c+di \in \mathbb{Z}[i]$ with $c+di \neq 0$, then

$$\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i \in \mathbb{Q}[i].$$

3.48 Example (*Rational functions*)
Let R be an integral domain. A **rational function over** R is a function in the indeterminant x of the form $\dfrac{p(x)}{q(x)}$, where $p(x), q(x) \in R[x]$ and $q(x) \neq 0$. The **field of rational functions over** R, denoted by $R(x)$, is the collection of all rational functions over R. We see from the definition that $\mathsf{Frac}(R[x]) = R(x)$.

For a unique factorization domain R, the following theorem relates the non-constant irreducibles in $R[x]$ to those in $\mathsf{Frac}(R)[x]$.

3.49 Theorem (*Gauss's lemma on irreducibility*)
Let R be a unique factorization domain, and let $F = \mathsf{Frac}(R)$. A non-constant polynomial $p(x) \in R[x]$ is irreducible if and only if it is irreducible in $F[x]$ and primitive in $R[x]$.

PROOF
($\Leftarrow$) We prove the contrapositive.

Class 16
02/09/2022

Suppose $p(x) \in R[x]$ is reducible over R. Further suppose that $p(x) \in R$ is primitive—if not, the proof for this direction is complete. Then $p(x)$ does not have non-unit constant factors in $R[x]$. Write $p(x) = f(x)g(x)$ for some non-constant $f(x), g(x) \in R[x]$. But this is also a non-trivial factorization of $p(x)$ over F. Thus $p(x)$ is reducible in $F[x]$.

($\Rightarrow$) We prove the contrapositive for this direction as well.

Suppose $p(x) \in R[x]$ is not primitive. Then a we have a non-unit $a \in R$ as a greatest common divisor of coefficients of $p(x)$. Thus $p(x) = aq(x)$, with both $a, q(x) \in R[x]$ non-units.

Now suppose $p(x) \in R[x]$ is primitive and reducible in $F[x]$. Then we can write $p(x) = f(x)g(x)$ for some non-constant $f(x), g(x) \in F[x]$. Multiplying through by the denominators of coefficients in $f(x)$ and $g(x)$, we get

$$d \cdot p(x) = r(x)s(x),$$

where $d \in R$ and $r(x), s(x) \in R[x]$ are non-constant polynomials. If d is a unit in R, we are done with $f(x) = d^{-1}r(x)s(x)$. Suppose d is not a unit in R. Let $d = d_1 d_2 \dots d_k$ be a factorization into irreducibles. We want to show that each d_i divides either $r(x)$ or $s(x)$ in $R[x]$. Write

$$r(x) = \sum_{i=0}^{m} a_i x^i \text{ and } s(x) = \sum_{j=0}^{n} b_j x^j$$

Suppose, for a contradiction, that $d_1 \nmid a_i$ and $d_1 \nmid b_j$, where $0 \leq i \leq m$ and $0 \leq j \leq n$ are, respectively, the smallest indices for the condition to hold. Then

the coefficient of x^{i+j} in $r(x)s(x)$ is

$$c_{i+j} = a_0 b_{i+j} + a_1 b_{i+j-1} + \cdots + a_{i-1} b_{j+1} + a_i b_j + a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0.$$

Note that $d_1 \mid c_{i+j}$ because $r(x)s(x) = d \cdot p(x)$. Additionally, $d_1 \mid (a_k b_{i+j-k})$ for all $k < i$, and $d_1 \mid (a_{i+j-\ell} b_\ell)$ for all $\ell < j$. Thus $d_1 \mid a_i b_j$ as well. However, d_1 is irreducible, thus prime, so d_1 divides either a_i or b_j, which is a contradiction. Iterating this process, we have

$$d \cdot p(x) = (d_1 d_2 \dots d_k) r_0(x) s_0(x)$$

with non-constant polynomials $r_0(x), s_0(x) \in R[x]$. Canceling d from both sides, we get $p(x) = r_0(x)s_0(x)$. □

We are now ready to draw the missing connection in Remark 3.37.

3.50 Theorem

Let R be an integral domain. Then $R[x]$ is a unique factorization domain if and only if R is a unique factorization domain.

Proof

($\Rightarrow$) Suppose that $R[x]$ is a unique factorization domain. Let $a \in R$ be non-zero non-unit. Then $a \in R[x]$ is also non-zero and non-unit. By the UFD condition, we can write $a = p_1(x)p_2(x)\cdots p_n(x)$ with unique (up to permutation and multiplication by units) irreducible polynomials $p_i(x) \in R[x]$. But each $p_i(x)$ must be constant, thus in R. The irreducible elements in R are precisely the irreducible constant polynomials in $R[x]$. Thus unique factorization in R follows from unique factorization of constant polynomials in $R[x]$.

($\Leftarrow$) The converse direction requires a strong induction argument, which is analogous to the proof of the fundamental theorem of arithmetic (Theorem 3.14). Suppose that R is a unique factorization domain. We proceed by induction on the degree of polynomials in $R[x]$.

- **Base case.** Non-zero, non-unit degree 0 polynomials in $R[x]$ are precisely non-zero non-unit elements in R. Further, factoring constant polynomials in $R[x]$ will only result in constant polynomial factors, and will give the same outcome as factoring elements in R. Thus existence and uniqueness of factorization of constant polynomials into irreducibles are guaranteed from the fact that R is a unique factorization domain.
- **Inductive step.** Let $N \geq 1$ and suppose that every polynomial of degree $0 \leq n \leq N$ factors uniquely into irreducibles in $R[x]$. Let $f(x) \in R[x]$ be a polynomial of degree $N+1$. By factoring the coefficients of $f(x)$ and dividing by their gcd, we may assume $f(x)$ is primitive. Let $F = \mathsf{Frac}(R)$.
  - If $f(x)$ is irreducible in $F[x]$, then it is irreducible in $R[x]$, and we are done with existence.
  - If $f(x)$ is reducible in $F[x]$, then it is reducible in $R[x]$ by Gauss's lemma. Further, because $f(x)$ is primitive in $R[x]$, we can write

$$f(x) = g(x)h(x)$$

for some non-constant polynomials $g(x), h(x) \in R[x]$. In particular, $\deg(g(x)), \det(h(x)) \leq N$, so the induction hypothesis applies. Both $g(x)$ and $h(x)$ are products of irreducibles in $R[x]$, so is $f(x)$.

For uniqueness, Suppose that

$$f(x) = p_1(x)p_2(x)\cdots p_k(x) = q_1(x)q_2(x)\cdots q_\ell(x)$$

for irreducible polynomials $p_i(x), q_j(x) \in R[x]$. Because $f(x)$ is primitive, by Gauss's lemma on primitivity, all of $p_i(x), q_j(x)$ are primitive. By Gauss's lemma on irreducibility, they are irreducible in $F[x]$ as well. But $F[x]$ is a unique factorization domain, so we must have $k = \ell$, and (up to reordering) $p_i(x) = u_i q_i(x)$ for some $u_i \in F$. Write $u_i = \dfrac{a_i}{b_i}$ for $a_i, b_i \in R$. Then $b_i p_i(x) = a_i q_i(x)$. Because $p_i(x)$ and $q_i(x)$ are primitive, we must have b_i and a_i differ by a unit in R, which implies that $u_i \in R$. This completes the proof for uniqueness, and the inductive step. □

3.51 Example

Note that $\mathbb{Z}$ has been our canonical example of a unique factorization domain. Theorem 3.50 now tells us that $\mathbb{Z}[x]$ is a unique factorization domain as well. Recall from Example 2.25 that $\mathbb{Z}[x]$ is not a principal ideal domain. This is our first example of a UFD that is not a PID.

3.52 Corollary

If R is a unique factorization domain, then $R[x_1, x_2, \ldots, x_n]$ is a unique factorization domain.

Proof

Note that $R[x_1, x_2, \ldots, x_n] = R[x_1, x_2, \ldots, x_{n-1}][x_n]$. The proof then proceed by induction on the number of indeterminants. □

3.53 Corollary

If R is a unique factorization domain, then $R[x_1, x_2, \ldots]$ (polynomials over R on countably many indeterminants) is a unique factorization domain.

Proof

Let $f \in R[x_1, x_2, \ldots]$. Because f is a finite sum of monomial terms, there exists $n \in \mathbb{N}$ such that $f \in R[x_1, x_2, \ldots, x_n]$. Let $m \geq n$. A reducible $p \in R[x_1, x_2, \ldots, x_n]$ is still reducible in $R[x_1, x_2, \ldots, x_m]$ by taking the same factors $p = ab$. More importantly, an irreducible $p \in R[x_1, x_2, \ldots, x_n]$ is expected to stay irreducible in $R[x_1, x_2, \ldots, x_m]$. This can be seen by taking the partial evaluation $x_1 = x_2 = \cdots = x_n = 1$ on a factorization $p = ab$. The left hand side will become a constant in R, while the terms involving $x_{n+1}, \ldots, x_m$ will remain on the right hand side. Thus $a, b \in R[x_1, x_2, \ldots, x_n]$, and one of them must be a unit. Therefore, the unique factorization of $f \in R[x_1, x_2, \ldots, x_n]$ into irreducibles will remain as a unique factorization into irreducibles in $R[x_1, x_2, \ldots]$. □

3.54 Remark

Corollary 3.53 cannot follow from the same proof of Corollary 3.52. Note that

induction can only prove statements $P(n)$ indexed by arbitrarily large *finite* natural numbers $n \in \mathbb{N}$. It is not capable of proving the statement $P(\infty)$.

4 Euclidean domains

Class 17
02/11/2022

We have seen two different, but yet analogous, versions of the division algorithm: over $\mathbb{Z}$ and over $F[x]$. The division algorithm over $\mathbb{Z}$ allows us to classify all cyclic groups, as well as to show that subgroups of cyclic groups are cyclic (see Theorems 5.8 and 5.11 of Math 330-1 notes). These result transfers to ring theory to justify that $\mathbb{Z}$ is a principal ideal domain. The division algorithm over $F[x]$, as demonstrated in Theorem 3.10, shows that $F[x]$ is a principal ideal domain as well. In this chapter, we generalize the division algorithm in an integral domain.

4.1 The Euclidean algorithm

4.1 Definition (*Euclidean domains*)
A **Euclidean norm** on an integral domain R is a function $\nu\colon R_{\neq 0} \to \mathbb{Z}_{\geq 0}$ that satisfies

(1) For all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $\nu(r) < \nu(b)$.
(2) For all non-zero $a, b \in R$, we have $\nu(a) \leq \nu(ab)$.

An integral domain R is a **Euclidean domain** (abbreviated as **ED**) if there exists a Euclidean norm on R.

4.2 Example
Evidently, $\mathbb{Z}$ is a Euclidean domain with a Euclidean norm $\nu(n) = |n|$:

- The division algorithm over $\mathbb{Z}$ gives existence of q and r.
- Given $a, b \neq 0$, we have $|b| \geq 1$, thus $\nu(ab) = |ab| = |a||b| \geq |a| = \nu(a)$.

4.3 Example
Similarly, if F is a field, then $F[x]$ is a Euclidean domain with a Euclidean norm $\nu(f(x)) = \deg(f(x))$:

- The division algorithm over $F[x]$ gives existence of $q(x)$ and $r(x)$.
- Given $f(x), g(x) \neq 0$, we have $\deg(g(x)) \geq 0$, thus

$$\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x)) \geq \deg(f(x)).$$

4.4 Example
Fields are trivial examples of Euclidean domains. A Euclidean norm on a field would be the constant 1 function, or any constant function.

4.5 Example
Another Euclidean domain that we have already seen is $\mathbb{Z}[i]$, with a Euclidean norm defined by $\nu(a+bi) = a^2+b^2$. This is the same with the square of complex absolute value.

- The existence of q and r has a neat geometric proof. Let $z, w \in \mathbb{Z}[i]$ with $w \neq 0$. Then $\dfrac{z}{w} \in \mathbb{C}$. Note that $\mathbb{Z}[i]$ forms a lattice in $\mathbb{C}$, and the

largest possible distance to the nearest lattice point is $1/\sqrt{2}$, attained at the center points of individual lattice grids $(0.5+0.5i)+\mathbb{Z}[i]$. This allows us to find $q \in \mathbb{Z}[i]$ such that

$$\left|\frac{z}{w} - q\right| \leq \frac{1}{\sqrt{2}} < 1.$$

Now let $r = z - wq$. Then evidently, $z = wq + r$. Further, if $r \neq 0$, then

$$\nu(r) = |r|^2 = |w|^2 \left|\frac{z}{w} - q\right|^2 < |w|^2 = \nu(w).$$

- Note that for a non-zero $a + bi \in \mathbb{Z}[i]$, $\nu(a+bi) = a^2 + b^2 \geq 1$ because either $|a| \geq 1$ or $|b| \geq 1$. Thus if $z, w \in \mathbb{Z}[i]$ are non-zero, then

$$\nu(zw) = |z|^2|w|^2 \geq |z|^2 = \nu(z).$$

The following theorem extends the chain of inclusions in Remark 3.37.

4.6 Theorem (*Euclidean domains are PIDs*)
If R is a Euclidean domain, then R is a principal ideal domain.

Proof
We will adapt the proof of Theorem 3.10 by replacing the degree function with an arbitrary Euclidean norm.

Suppose that R is a Euclidean domain with Euclidean norm $\nu\colon R_{\neq 0} \to \mathbb{Z}_{\geq 0}$. Let I be a non-trivial ideal of R, and let $x \in I$ be arbitrary. Let $y \in I$ be a non-zero element with minimal $\nu(y)$. Such an element is guaranteed to exist by the well-ordering principle (Theorem 5.9 of Math 330-1 notes). By the division algorithm, there exist $q, r \in R$ with either $r = 0$ or $\nu(r) < \nu(y)$ such that $x = yq + r$. However, $r = x - yq \in I$, so $r = 0$ by minimality of the Euclidean norm of y. Thus $x = yq$. Because x was arbitrarily chosen, every element in I is a multiple of y. Thus $I = (y)$ is principal. □

4.7 Remark
With Example 4.4 and Theorem 4.6, We now have

$$\text{integral domains} \supseteq \text{UFDs} \supseteq \text{PIDs} \supseteq \text{EDs} \supseteq \text{fields}$$

Additionally, every non-PID that we have seen is not a Euclidean domain. These include

- $\mathbb{Z}[\sqrt{-5}]$
- $\mathbb{Z}[x]$
- $F[x_1, \ldots, x_n]$

4.8 Example

Perhaps the most famous example of a PID that is not Euclidean is $\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$. The proof is not out of reach, but is quite long and intricate. You may refer to an exposition by Jack C. Wilson in *A Principal Ideal Ring That Is Not a Euclidean Ring*, Math. Mag., Vol. 46, No. 1 (Jan., 1973), pp. 34-38.

The Euclidean norm function in a Euclidean domain R gives a nice classification of units in R.

4.9 Theorem

Let R be a Euclidean domain with Euclidean norm $\nu\colon R_{\neq 0} \to \mathbb{Z}_{\geq 0}$. Define

Class 18
02/14/2022

$$m = \min\{\nu(a) \mid a \in R_{\neq 0}\}.$$

Then $\nu(a) = m$ if and only if $a \in R$ is a unit.

Proof

Let $a \in R_{\neq 0}$ be arbitrary. Then $\nu(1) \leq \nu(1a) = \nu(a)$. Thus $\nu(1) = m$.

($\Leftarrow$) Suppose that $a \in R$ is a unit. Then $\nu(a) \leq \nu(aa^{-1}) = \nu(1) = m$. But m is defined to be the minimal Euclidean norm. Thus $\nu(a) = m$.

($\Rightarrow$) Suppose that $a \in R$ satisfies $\nu(a) = m$. Then by the division algorithm, there exist $q, r \in R$ with either $r = 0$ or $\nu(r) < \nu(a)$ that satisfies $1 = aq + r$. But we assumed $\nu(a)$ to be the minimal Euclidean norm. We must have $r = 0$, and $1 = aq$. Thus a is a unit. □

4.10 Example

Theorem 4.9 provides a quick way to find all units in a Euclidean domain. For example,

- In $\mathbb{Z}$, the units are ± 1, the only integers n satisfying $|n| = 1$.
- In $\mathbb{Z}[i]$, the units are ± 1 and $\pm i$. These are the only Gaussian integers that satisfy $|a + bi|^2 = 1$.
- In $F[x]$, the units are precisely all non-zero constant polynomials, which are all degree-0 polynomials.

We can now formally state the *Euclidean algorithm* for computing greatest common divisors in a Euclidean domain.

4.11 Theorem (*Euclidean algorithm*)

Let R be a Euclidean domain with Euclidean norm ν. Let $a, b \in R$ with $a \neq 0$. A greatest common divisor of a and b can be computed by the following algorithm:

- If $b \neq 0$, find $q, r \in R$ that satisfy $a = qb + r$ with either $r = 0$ or $\nu(r) < \nu(b)$, then replace a by b and b by r, and repeat.
- If $b = 0$, then stop, and $\gcd(a, b) = a$.

Proof

Note that $\nu(b)$, from all steps of Euclidean algorithm, will form a strictly decreasing sequence of non-negative integers. Thus the algorithm will terminate—the remainder r will eventually become 0 after finitely many steps, at which point we can conclude that the current value of b is a greatest common divisor.

To see that the algorithm returns a greatest common divisor. If $b = 0$, then evidently a is a gcd of a and b. If $b \neq 0$, let $q, r \in R$ that satisfy $a = qb + r$. Then $r = a - qb$. Let d be a common divisor of a and b. Then $d \mid a$ and $d \mid b$.

Thus $d \mid r$. Conversely, if $d \mid r$ and $d \mid b$, then $d \mid a$. It follows that replacing a by b and b by r will not change the common divisors, including greatest common divisors. □

4.12 Example

Let's perform Euclidean algorithm on $\mathbb{Z}[i]$, which we have shown in Example 4.5 to be a Euclidean domain. We will use the idea from the proof to find q and r for Euclidean algorithm in $\mathbb{Z}[i]$. Consider $a = 23 - 11i$ and $b = 25i$ in $\mathbb{Z}[i]$.

(1) $\dfrac{23-11i}{25i} = -\dfrac{11}{25} - \dfrac{23}{25}i$, with $-i$ being the nearest Gaussian integer. Thus

$$23 - 11i = (-i) \cdot (25i) + (-2 - 11i),$$

and $\gcd(23 - 11i, 25i) = \gcd(25i, -2 - 11i)$.

(2) $\dfrac{25i}{-2-11i} = -\dfrac{11}{5} - \dfrac{2}{5}i$, with -2 being the nearest Gaussian integer. Thus

$$25i = (-2) \cdot (-2 - 11i) + (-4 + 3i),$$

and $\gcd(25i, -2 - 11i) = \gcd(-2 - 11i, -4 + 3i)$.

(3) $\dfrac{-2-11i}{-4+3i} = -1 + 2i \in \mathbb{Z}[i]$. Thus

$$-2 - 11i = (-1 + 2i) \cdot (-4 + 3i) + 0.$$

Stopping at this step will give $\gcd(23 - 11i, 25i) = \gcd(25i, -2 - 11i) = \gcd(-2 - 11i, -4 + 3i) = -4 + 3i$.

4.13 Remark

The division algorithm for Euclidean domains does not require uniqueness on q or r. In step (1) of Example 4.12, we could have chosen $q = -1 - i$, which is also within a distance of 1 from $\dfrac{23-11i}{25i}$. Similarly, in step (2), we could have chosen $q = -2 - i$ or $q = -3$.

Such phenomenon is not only limited to $\mathbb{Z}[i]$. It also occurs in $\mathbb{Z}$, but mostly going unnoticed. Consider a simple example of 7 divided by 3 with remainder. We have the familiar result $7 = 2 \cdot 3 + 1$, with 2 being the quotient and 1 being the remainder. However, writing $7 = 3 \cdot 3 + (-2)$ also satisfies the Euclidean norm condition: $\nu(-2) = 2 < 3 = \nu(3)$. So 3 could also be a quotient of 7 divided by 3, with -2 being the corresponding remainder.

4.2 Multiplicative norms

Class 19
02/16/2022

As demonstrated in the opening section, the Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain, therefore a unique factorization domain. However, irreducibility seems obscure in $\mathbb{Z}[i]$. While the notion of a Euclidean norm (or any norm) does not affect irreducibility of an element, we would certainly hope to have a norm that

pertains the multiplicative structure of the ring. The Euclidean norm for $\mathbb{Z}[i]$ has such a nice property: $\nu(z)\nu(w) = \nu(zw)$ for all $z, w \in \mathbb{Z}[i]$, which hints at an understanding towards irreducibility in $\mathbb{Z}[i]$. The notion of a *multiplicative norm* generalizes this effort in commutative rings.

4.14 Definition (*Multiplicative norms*)
Let R be a commutative ring. A **multiplicative norm** on R is a function $N\colon R \to \mathbb{Z}$ that satisfies

(1) $N(ab) = N(a)N(b)$ for all $a, b \in R$, and
(2) $N(a) = 0$ if and only if $a = 0$.

4.15 Remark
A Euclidean norm need not be a multiplicative norm; a multiplicative norm certainly need not be a Euclidean norm, either. We will demonstrate both in the following examples.

4.16 Example
In $\mathbb{Z}$, the canonical Euclidean norm $\nu_1(n) = |n|$ is multiplicative. Evidently, $|ab| = |a||b|$ and $|n| = 0$ if and only if $n = 0$. Note that the function $\nu_2(n) = 2|n|$ also satisfies the Euclidean norm conditions. However, ν_2 is not multiplicative:

$$\nu_2(ab) = 2|ab|, \text{ while } \nu_2(a)\nu_2(b) = 2|a| \cdot 2|b| = 4|ab|.$$

4.17 Example
In $F[x]$, the Euclidean norm $\nu(f(x)) = \deg(f(x))$ is not multiplicative:

$$\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x)) \neq \deg(f(x))\deg(g(x))$$

4.18 Example
For a square-free integer $n \geq 1$, we can define a multiplicative norm on $\mathbb{Z}[\sqrt{-n}]$ using complex absolute value. Let $N\colon \mathbb{Z}[\sqrt{-n}] \to \mathbb{Z}$ be given by

$$N(a + b\sqrt{-n}) = |a + b\sqrt{-n}|^2 = a^2 + nb^2$$

Because $\mathbb{Z}[\sqrt{-n}] \leq \mathbb{C}$, and $N(a + b\sqrt{-n}) = |a + b\sqrt{-n}|^2$, it is evidently a multiplicative norm. What perhaps is a little surprising is that this definition also works for negative integers n.

For a square-free integer n, that is, the prime factors of n do not repeat, define a norm on $\mathbb{Z}[\sqrt{n}]$ by

$$N(a + b\sqrt{n}) = a^2 - nb^2$$

We demonstrate that this is indeed a multiplicative norm:

- Let $a, b, c, d \in \mathbb{Z}$. Then

$$\begin{aligned} N((a+b\sqrt{n})(c+d\sqrt{n})) &= N(ac+nbd+(ad+bc)\sqrt{n}) \\ &= (ac+nbd)^2 - n(ad+bc)^2 \\ &= (a^2c^2+2nabcd+n^2b^2d^2) - (na^2d^2+2nabcd+nb^2c^2) \\ &= a^2c^2 - na^2d^2 - nb^2c^2 + n^2b^2d^2 \\ N(a+b\sqrt{n})N(c+d\sqrt{n}) &= (a^2-nb^2)(c^2-nd^2) \\ &= a^2c^2 - na^2d^2 - nb^2c^2 + n^2b^2d^2 \end{aligned}$$

- $N(a+b\sqrt{n}) = a^2 - nb^2 = 0$ if and only if $a^2 = nb^2$. Because n is square-free, this is possible if and only if $b = 0$, in which case we have $a^2 - nb^2 = a^2 = 0$, so $a = 0$, and so $a + b\sqrt{n} = 0$ as required.

Note that this collection of examples shows multiplicative norms do not need to be Euclidean. In particular, we know $\mathbb{Z}[\sqrt{-5}]$ is not Euclidean because it is not a UFD. However, the norm $N(a+b\sqrt{-5}) = a^2 + 5b^2$ is multiplicative.

4.19 Remark

The multiplicative norm presented in Example 4.18 is a special case of what we call a *field norm* for a field extension. You will learn more about field norms in Math 330-3.

We appeal to the notion of multiplicative norms because we want to further understand irreducibility of elements in a ring. The following two results relate the multiplicative norms to some fundamental building blocks in a ring.

4.20 Lemma

Let R be a commutative ring. If there exists a multiplicative norm N on R, then R is an integral domain.

Proof

Suppose that $N\colon R \to \mathbb{Z}$ is a multiplicative norm. Let $a, b \in R$ and suppose that $ab = 0$. Then $N(a)N(b) = N(ab) = 0$. But $N(a), N(b) \in \mathbb{Z}$, so either $N(a) = 0$ or $N(b) = 0$. Thus either $a = 0$ or $b = 0$. This shows R contains no zero divisor, thus an integral domain. □

4.21 Lemma

Let R be an integral domain with a multiplicative norm $N\colon R \to \mathbb{Z}$. If $u \in R$ is a unit, then $N(u) = \pm 1$.

Proof

First, for all $a \in R$, we have $N(a) = N(1a) = N(1)N(a)$. By cancellation in $\mathbb{Z}$, we have $N(1) = 1$. Now suppose that $u \in R$ is a unit. Then $1 = N(1) = N(uu^{-1}) = N(u)N(u^{-1})$. Thus $N(u)$ is a unit in $\mathbb{Z}$, so $N(u) = \pm 1$. □

4.22 Remark

Note that the converse of this lemma is not necessarily true—we could define a norm that is constant 1 on all non-zero elements of R. This is a multiplicative norm because R does not have zero divisors and $1 \cdot 1 = 1$. Most of the time, we

would like to work with multiplicative norms that respect units in the ring.

Class 20
02/18/2022

The next result provides a sufficient condition for irreducibility of an element in an integral domain equipped with a unit-preserving multiplicative norm.

4.23 Lemma
Let R be an integral domain with a multiplicative norm $N\colon R \to \mathbb{Z}$, and assume that $N(a) = \pm 1$ if and only if $a \in R$ is a unit. If $p \in R$ has a norm $N(p)$ that is prime in $\mathbb{Z}$, then p is irreducible in R.

PROOF
We prove the contrapositive. Suppose that $p \in R$ is reducible. Then $p = ab$ for some non-zero, non-unit $a, b \in \mathbb{Z}$. By assumption, $|N(a)|, |N(b)| \geq 2$. Thus $|N(p)| = |N(ab)| = |N(a)| \cdot |N(b)|$ is a product of two integers ≥ 2. Hence, $N(p)$ is not prime in $\mathbb{Z}$. □

4.24 Example
We revisit Example 3.21, where we presented two different factorizations into irreducibles for $6 \in \mathbb{Z}[\sqrt{-5}]$:

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 + \sqrt{-5}).$$

However, in that example, we showed irreducibility of 2 through a rather intricate computation, and we scrubbed irreducibility of the other three elements under the rug at the time. With the help of the multiplicative norm defined in Example 4.18, we have a slick way of showing irreducibility of 2, 3, and $1 \pm \sqrt{-5}$ all *at once*.

First, we see that the field norm $N(a+b\sqrt{-5}) = a^2+5b^2$ satisfies the assumption for Lemma 4.23. Note that $N(a + b\sqrt{-5}) = 1$ if and only if $a = \pm 1$ and $b = 0$, which gives the only two units $\pm 1 \in \mathbb{Z}[\sqrt{-5}]$. Note that $N(2) = 4$, $N(3) = 9$, and $N(1 \pm \sqrt{-5}) = 6$. If any of these can be written as a product of two non-units $z, w \in \mathbb{Z}[\sqrt{-5}]$, we must have either ± 2 or ± 3 as the values of $N(z)$ and $N(w)$. But $a^2 + 5b^2 \neq 2$ and $a^2 + 5b^2 \neq 3$ for all $a, b \in \mathbb{Z}$. So none of 2, 3, and $1 \pm \sqrt{-5}$ is a product of two non-units in $\mathbb{Z}[\sqrt{-5}]$—they are all irreducible.

We conclude our discussion on norms with Fermat's $p = a^2 + b^2$ theorem, which will allow us to classify all prime (irreducible) elements in $\mathbb{Z}[i]$. We will make use of the next technical theorem from group theory.

4.25 Theorem
Let $p \geq 2$ be a prime integer. Then $\mathbb{Z}_p^* \cong \mathbb{Z}_{p-1}$ as groups.

4.26 Remark
This is a classic result in group theory. You can find several different proofs for this theorem, many of which are purely group-theoretic. We can use what we have learned in ring theory to prove it.

PROOF (not covered in class, non-examinable)
Because $|\mathbb{Z}_p^*| = p - 1$, it suffices to find an element in $\mathbb{Z}_p^*$ of order $p - 1$, which

will serve as a generator of the group. Let n be the maximum order among elements in $\mathbb{Z}_p^*$. We want to show that $n = p - 1$.

Let $a \in \mathbb{Z}_p^*$ be an element of order n, and let $b \in \mathbb{Z}_p^*$ be an element with order $m > 1$. Then $m \leq n$. Suppose, for a contradiction, that $m \nmid n$. Consider the prime factorizations of m and n, there exists a prime power q^k such that $q^k \mid m$ and $q^k \nmid n$. Let $\ell \geq 0$ be the largest integer such that $q^\ell \mid n$. Then $\ell < k$. Now consider the order of the element $a^{q^\ell} b^{m/q^k} \in \mathbb{Z}_p^*$. The order of a^{q^ℓ} is n/q^ℓ, which is free of q factors. The order of b^{m/q^k} is q^k. So $\gcd(|a^{q^\ell}|, |b^{m/q^k}|) = 1$. Then the intersection $\langle a^{q^\ell} \rangle \cap \langle b^{m/q^k} \rangle = \{1\}$. To see this, if $c \neq 1$ is in the intersection, then $|c| > 1$ divides both $|a^{q^\ell}|$ and $|b^{m/q^k}|$, but $\gcd(|a^{q^\ell}|, |b^{m/q^k}|) = 1$. Now suppose that $|a^{q^\ell} b^{m/q^k}| = N$. Then $a^{Nq^\ell} b^{Nm/q^k} = 1$, which implies that $a^{Nq^\ell} = b^{-Nm/q^k} \in \langle a^{q^\ell} \rangle \cap \langle b^{m/q^k} \rangle = \{1\}$. Thus $(n/q^\ell) \mid N$ and $q^k \mid N$. But they are relatively prime, so N is a multiple of $(n/q^\ell)q^k = nq^{k-\ell} > n$, contradicting the assumption that n is the maximal order. So we can conclude that every element in $\mathbb{Z}_p^*$ has an order that divides n.

Now we consider the polynomial $x^n - 1$ in $\mathbb{Z}_p[x]$. Every $a \in \mathbb{Z}_p^*$ is a root of this polynomial, because the order of a divides n, and thus $a^n = 1$. Because $\mathbb{Z}_p$ is a field, by the factor theorem, $x - a$ is a factor of $x^n - 1$ for all $a \in \mathbb{Z}_p^*$. Thus $\deg(x^n - 1) = n \geq p - 1$. But we had $n \leq p - 1$ by Lagrange's theorem (see Theorem 7.8 and Corollary 7.11 of Math 330-1 notes). Thus $n = p - 1$. □

4.27 Theorem (*Fermat's $p = a^2 + b^2$ theorem*)
Let $p \geq 3$ be an odd prime. Then p can be expressed as the sum of two perfect squares if and only if $p \equiv 1 \pmod 4$.

Proof
($\Rightarrow$) Suppose that $p = a^2 + b^2$. Then a and b cannot be both even or both odd—otherwise $p = a^2 + b^2$ would be even, contradicting the assumption that p is odd. Without loss of generality, suppose a is odd and b is even. Then $a = 2k + 1$ and $b = 2\ell$ for some $k, \ell \in \mathbb{Z}$, and so

$$p = (2k + 1)^2 + (2k)^2 = (4k^2 + 4k + 1) + (4k^2) = 4(2k^2 + k) + 1.$$

Thus $p \equiv 1 \pmod 4$.

($\Leftarrow$) Suppose that $p \equiv 1 \pmod 4$. Then $4 \mid (p - 1)$. Let $x \in \mathbb{Z}_p^*$ be a generator. Then the order of $y = x^{(p-1)/4}$ is 4, because $y^4 = x^{p-1} = 1$, and $y^k \neq 1$ for $k = 1, 2, 3$. Working in $\mathbb{Z}_p$ as a field, we have $0 = y^4 - 1 = (y^2 + 1)(y^2 - 1)$. Because $y^2 - 1 \neq 0$, we must have $y^2 + 1 = 0$. Thus $p \mid (y^2 + 1)$ in $\mathbb{Z}$, as well as in $\mathbb{Z}[i]$.

Now suppose, for a contradiction, that p is irreducible in $\mathbb{Z}[i]$. Then p is prime in $\mathbb{Z}[i]$. Because $y^2 + 1 = (y + i)(y - i)$, either $p \mid (y + i)$ or $p \mid (y - i)$. If $p \mid (y + i)$, then $p(a + bi) = y + i$ for $a, b \in \mathbb{Z}$ Thus $pb = 1$, which is impossible. Likewise, if $p \mid (y - i)$, then $p(a + bi) = y - i$ for $a, b \in \mathbb{Z}$ Thus $pb = -1$, which is also impossible. Therefore, p must be reducible in $\mathbb{Z}[i]$.

Write $p = (a + bi)(c + di)$. Take the multiplicative Euclidean norm on $\mathbb{Z}[i]$:

$$p^2 = (a^2 + b^2)(c^2 + d^2),$$

where $a^2 + b^2 \neq 1$ and $c^2 + d^2 \neq 1$. Thus $a^2 + b^2 = c^2 + d^2 = p$, as desired. □

4.28 Corollary (*Irreducibility in* $\mathbb{Z}[i]$)
The irreducible elements in $\mathbb{Z}[i]$ are precisely all products of a unit in $\mathbb{Z}[i]$ with an element $p \in \mathbb{Z}[i]$ that satisfies one of the following:

(1) A positive odd prime $p \in \mathbb{Z}$ that is congruent to 3 mod 4,
(2) $p = 1 + i$, or
(3) $p = a + bi$, where $a^2 + b^2$ is a prime integer congruent to 1 mod 4.

Proof
Suppose that $p = a + bi \in \mathbb{Z}[i]$ is irreducible, then so is $a - bi$. Thus $a^2 + b^2 = (a + bi)(a - bi)$ is a factorization into irreducibles in $\mathbb{Z}[i]$. It follows that $a^2 + b^2$ has at most two prime factors in $\mathbb{Z}$. If $a^2 + b^2$ is prime in $\mathbb{Z}$, then we are in case (2) or (3), depending on whether $a^2 + b^2 = 2$ or $a^2 + b^2 \geq 3$. If $a^2 + b^2$ is not prime in $\mathbb{Z}$, then its two prime factors must be either both $|a|$ or both $|b|$. So either $a = 0$ or $b = 0$. Thus $p \in \mathbb{Z}$ is a prime, and $N(p) = p^2$. Because $\mathbb{Z}[i]$ is a UFD, p must be an odd prime that is congruent to 3 mod 4 by Fermat's sum of squares theorem.

Now let's show that all elements listed are irreducible in $\mathbb{Z}[i]$. Note that elements in cases (2) and (3) all have prime norms, so they are irreducible. Now let $p \in \mathbb{Z}$ be an odd prime congruent to 3 mod 4. Suppose, for a contradiction that p is reducible in $\mathbb{Z}[i]$. Then $p = \alpha\beta$ for some $\alpha, \beta \in \mathbb{Z}[i]$ with $N(\alpha), N(\beta) > 1$. Because $N(\alpha)N(\beta) = N(p) = p^2$, we must have $N(\alpha) = N(\beta) = p$. But p is not a sum of two perfect squares, so no such $\alpha, \beta \in \mathbb{Z}[i]$ exist. Thus an odd prime $p \not\equiv 1 \pmod 4$ is irreducible in $\mathbb{Z}[i]$. □

5 Modules

Class 21
02/21/2022

Our canonical understanding of *vector spaces* revolves around $\mathbb{R}^n$ and its subspaces. There are many non-Euclidean vector spaces beyond what we saw in linear algebra; there are also vector spaces defined over $\mathbb{C}$, that is, allowing complex numbers to serve as scalars. We would like to generalize the notion of vector spaces through one of the most powerful idea in this course—abstraction.

5.1 Let's first consider the properties we expect a vector space V to satisfy. The "subspace test" in $\mathbb{R}^n$ usually involves three axioms to check:

- **Closure under addition:** if $v, w \in V$, then $v + w \in V$.
- **Closure under scalar multiplication:** if $v \in V$ and $c \in \mathbb{R}$, then $cv \in V$.
- **Non-emptiness:** $V \neq \varnothing$, also often manifested as $0 \in V$.

In linear algebra, addition and scalar multiplication are taken for grated as some nice operations we can perform on vectors. But if we put them under abstract algebra lens, what we have for these operations are

- An abelian group V under vector addition, and
- A (left) group action of $\mathbb{R}^*$ on V by scalar multiplication.

We then have the following axioms for addition, as a function $+\colon V \times V \to V$:

(1) For all $u, v, w \in V$, we have $(u + v) + w = u + (v + w)$.
(2) There exists $0 \in V$ such that $v + 0 = 0 + v = v$ for all $v \in V$.
(3) For all $v \in V$, there exists $w \in V$ such that $v + w = w + v = 0$.
(4) For all $v, w \in V$, we have $v + w = w + v$.

Scalar multiplication is thus a function $\alpha\colon \mathbb{R} \times V \to V$, written as $\alpha(c, v) = cv$, that satisfy

(5) For all $v \in V$, we have $1v = v$.
(6) For all $c, d \in \mathbb{R}^*$ and all $v \in V$, we have $c(dv) = (cd)v$.

Note that Axiom (6) can be extended to all of $\mathbb{R}$ to allow for 0 as a scalar:

(6*) For all $c, d \in \mathbb{R}$ and all $v \in V$, we have $c(dv) = (cd)v$.

Further, if we consider the additive structures of $\mathbb{R}$ and V together with scalar multiplication, we have two axioms on distribution:

(7) For all $c \in \mathbb{R}$ and all $v, w \in V$, we have $c(v + w) = cv + cw$.
(8) For all $c, d \in \mathbb{R}$ and all $v \in V$, we have $(c + d)v = cv + dv$.

These eight axioms together form the axiomatic approach to define **real vector spaces**. Evidently, $\mathbb{R}^n$ and its subspaces are real vector spaces.

5.2 Example (*A non-Euclidean vector space*)
The set $C^0(\mathbb{R})$ of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ is a real vector space, with vector addition defined by pointwise addition, the zero function $f(x) = 0$ as the zero vector, and $(-f)(x) = -(f(x))$. Details of verifying the vector space axioms are left as an exercise.

5.1 Definitions and examples

The field $\mathbb{R}$ of real numbers is a weirdly specific element in the definition of real vector spaces. In fact, what we have defined is also referred to as "vector spaces *over* $\mathbb{R}$." It is then not a surprise that we can define vector spaces over $\mathbb{C}$, or more generally, vector spaces over a field F, by simply replacing $\mathbb{R}$ with any field we like in the definition. It turns out that we can further relax the restriction from a field to a ring, and we obtain a *module*.

5.3 Definition *(Modules)*
Let R be a ring. A **left module over R**, or a **left R-module**, is an abelian group $(M, +_M)$ together with an action function $\alpha\colon R \times M \to M$, written as $\alpha(r, m) = rm$, that satisfies:

(1) For all $m \in M$, we have $1_R\, m = m$.
(2) For all $r, s \in R$ and all $m \in M$, we have $r(sm) = (rs)m$.
(3) For all $r \in R$ and all $m, n \in M$, we have $r(m +_M n) = rm +_M rn$.
(4) For all $r, s \in R$ and all $m \in M$, we have $(r +_R s)m = rm +_M sm$.

Right R-modules are defined analogously, and they are the same (isomorphic) if R is a commutative ring. If $R = F$ is a field, then we have defined a **vector space over F**.

5.4 Remark
We are dealing with increasing level of complexity in terms of operations as we start to discuss modules—there are two additions and two multiplications flying around at the same time! Both the ring R and the module M has their respective addition binary operation, and they are related by axiom (4) above. The operation rm is referred to as the "action" of R on M, or as "scalar multiplication." This is different from the multiplication in R, and they are related by axiom (2) above.

Like many algebraic structures, we would like to define *submodules* as well.

5.5 Definition *(Submodules)*
Let R be a ring, and let M be an R-module. An **R-submodule** of M is a subgroup $N \leq M$ which is closed under the R-action on M, *i.e.*, $rn \in N$ for all $n \in N$.

5.6 Example
All real vector spaces are $\mathbb{R}$-modules. These include $\mathbb{R}^n$ and their subspaces, as well as $C^0(\mathbb{R})$. Subspaces of them are $\mathbb{R}$-submodules.

5.7 Example
Just like $\mathbb{R}^n$ is an $\mathbb{R}$-module, the direct product R^n is an R-module for any ring R. The addition is defined component-wise:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

while scalar multiplication is defined just like scalar multiplication of vectors

$$r(x_1, \ldots, x_n) = (rx_1, \ldots, rx_n).$$

5.8 Example
Because a (left/right) ideal I of R is an abelian groups closed under (left/right) multiplication by elements from the entire ring, every (left/right) ideal I of R is a (left/right) R-module, and also a (left/right) submodule over R.

The integer ring $\mathbb{Z}$ is very special in the theory of ideals, as demonstrated in the following example.

5.9 Example (*Abelian groups are the same things as $\mathbb{Z}$-modules*)
Let $(G, +)$ be any abelian group. Then G can be made into a $\mathbb{Z}$-module with the following $\mathbb{Z}$-action:

$$ng = \begin{cases} g + g + \cdots + g \quad (n \text{ times}) & \text{if } n > 0, \\ e_G & \text{if } n = 0, \\ -g - g - \cdots - g \quad (-n \text{ times}) & \text{if } n < 0, \end{cases}$$

Conversely, every $\mathbb{Z}$-module has an underlying abelian group in its definition, and it respects the $\mathbb{Z}$-action defined above as well:

$$ng = (1 + 1 + \cdots + 1)g = 1g + 1g + \cdots + 1g = g + g + \cdots + g.$$

This shows that abelian groups and $\mathbb{Z}$-modules are the same things. This also indicates that $\mathbb{Z}$-submodules are the same things as subgroups of abelian groups.

5.10 Theorem (*The one-step submodule test*)
Let M be an R-module. A non-empty subset $N \subseteq M$ is an R-submodule if and only if $x + ry \in N$ for all $x, y \in N$ and all $r \in R$.

Class 22
02/23/2022

Proof
The forward direction follows from the definition of submodules.

Suppose that $x + ry \in N$ for all $x, y \in N$ and all $r \in R$. Setting $r = -1_R$ shows $N \leq M$ as a subgroup. Setting $x = 0_M$ shows that N is closed under the R-action. □

5.2 Module homomorphisms

Like all algebraic structures we have seen, modules have their homomorphisms, which is defined as you would expect.

5.11 Definition (*Module homomorphisms*)
Let R be a ring, and let M and N be R-modules. A function $f\colon M \to N$ is an **R-module homomorphism** if it satisfies

(1) $f(x + y) = f(x) + f(y)$ for all $x, y \in M$;

(2) $f(rx) = rf(x)$ for all $x \in M$ and $r \in R$.

An **R-module isomorphism** is a bijective R-module homomorphism. When there is an R-module isomorphism between two R-modules M and N, we say that M and N are **isomorphic (as R-modules)**, and write $M \cong_R N$. When the context is clear, $M \cong N$ can be used as well.

Kernels and images are also defined analogously to groups and rings.

5.12 Definition (*Kernels and images*)
Let M and N be R-modules, and let $f\colon M \to N$ be an R-module homomorphism. The **kernel** of f, denoted by $\ker(f)$, is

$$\ker(f) = \{m \in M \mid f(m) = 0_N\} \subseteq M.$$

The **image** of f, denoted by $\operatorname{im}(f)$, is

$$\operatorname{im}(f) = \{f(m) \mid m \in M\} \subseteq N.$$

5.13 Remark
Note that an R-module is an abelian group, and an R-module homomorphism is a homomorphism between abelian groups, guaranteed by condition (1) in the definition of R-module homomorphisms. The image and the kernel of an R-module homomorphism are therefore the same as the image and the kernel of the underlying group homomorphism. It is then a straightforward exercise of the one-step submodule test that $\ker(f) \leq M$ and $\operatorname{im}(f) \leq N$ as R-submodules.

5.14 Lemma
Let M and N be R-modules, and let $f\colon M \to N$ be an R-module homomorphism. Then $\ker(f) \leq M$ and $\operatorname{im}(f) \leq N$ as R-submodules.

Proof
Omitted. □

5.15 Example (*Linear transformations*)
If $R = F$ is a field, then the definition of module homomorphisms coincide with that of **linear tranformations** on F-vector spaces. The images and kernels coincide with the images (column spaces) and kernels (null spaces) we learned in linear algebra.

5.16 Example
Let R be a ring, and consider $M = R$ itself as an R-module. The ring homomorphisms and the module homomorphisms are different. For example, if $R = \mathbb{R}[x]$, then $\varphi\colon R \to R$ defined by $\varphi\colon f(x) \to 2f(x)$ is a module homomorphism, but not a ring homomorphism, while $\psi\colon R \to R$ defined by $\psi\colon f(x) \to f(2x)$ is a ring homomorphism, but not an R-module homomorphism.

5.17 Example ($\mathbb{Z}$-*module homomorphisms*)
In a $\mathbb{Z}$-module M, we expect

$$\varphi(nx) = \varphi(x + x + \cdots + x) = \varphi(x) + \varphi(x) + \cdots + \varphi(x) = n\varphi(x).$$

But this is already guaranteed by the underlying abelian group homomorphism. Thus $\mathbb{Z}$-module homomorphisms are the same as abelian group homomorphisms.

5.18 Example (*Quotient abelian groups*)
Let M be an abelian group, and $N \leq M$ a subgroup. Then N is a normal subgroup of M, so that we have a well-defined quotient group M/N with a quotient map $f\colon M \to M/N$ defined by $f(m) = m + N$, which defines a homomorphism of abelian groups.

If M is endowed with an R-module structure and N is a submodule, then the quotient group M/N can be made into an R-module as well by defining an R-action

$$r(x + N) = rx + N$$

To verify that this action is well-defined, let $x + N = y + N$. Then $x - y \in N$, and $rx - ry = r(x - y) \in N$. Thus $rx + N = ry + N$. Additionally, M/N is an abelian group. We should further verify that this R-action indeed defines an R-module, that is, the four axioms for modules are satisfied. We leave the details as an exercise here.

The quotient map $f\colon M \to M/N$ that we just defined is also an R-module homomorphism. We only need to verify the second axiom because f is an abelian group homomorphism to begin with:

$$f(rx) = rx + N = r(x + N) = rf(x).$$

This example justifies the definition of a *quotient R-module*, given next.

5.19 Definition (*Cosets and quotient modules*)
Let R be a ring, M be an R-module, and $N \leq M$ an R-submodule. A **coset** of N is a subset of the form

Class 23
02/25/2022

$$x + N = \{x + n \mid n \in N\}$$

for some $x \in M$. The **quotient R-module** of M by N is the R-module M/N, with the underlying abelian group being the quotient group M/N, and the R-action given by $r(x + N) = rx + N$.

5.20 Example
Let R be a ring, and let I be an ideal of R. Then $M = R$ and $N = I$ are both R-modules, with $N \leq M$ as an R-submodule. The quotient $M/N = R/I$, which consists of all cosets of N (or equivalently, I), can be used to denote the quotient abelian group, quotient ring, as well as the quotient R-module of R by I. From the set theoretical point of view, they are the same set.

5.21 Example
Let $V = \mathbb{R}^2$ be a vector space over $\mathbb{R}$. Let $W = \operatorname{span}\{(1, 2)\}$. Then $W \leq V$ as a subspace ($\mathbb{R}$-submodule). Note that W is the line $y = 2x$ in $\mathbb{R}^2$. Further, the cosets of W are of the form

$$(a, b) + W = \{(a, b) + t(1, 2) \mid t \in \mathbb{R}\}.$$

They are lines parallel to $y = 2x$. Each of these lines intersects the y-axis at a unique place—its y-intercept. Thus each coset has a unique representative of the form $(0, b)$. This suggests an $\mathbb{R}$-vector space isomorphism between $\mathbb{R}$ and V/W given by $b \mapsto (0, b)+W$, which corresponds to an *invertible linear transformation* in linear algebra. This can be further justified by the *first isomorphism theorem for modules*.

5.22 Theorem (*First isomorphism theorem for modules*)
Let R be a ring, and let M and N be R-modules. If $f\colon M \to N$ is an R-module homomorphism, then $M/\ker(f) \cong_R \operatorname{im}(f)$.

PROOF
You have already seen the proof of the analogous theorems for groups and for rings. I am sure you can imagine how the proof is going to proceed! □

5.23 Example
We revisit the quotient vector space V/W in Example 5.21. Note that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = 2x - y$ is a surjective linear transformation with $\ker(f) = \operatorname{span}\{(1, 2)\}$. By the first isomorphism theorem for modules, $V/W = \mathbb{R}^2/\ker(f) \cong \mathbb{R}$.

5.24 Example
Consider $\mathbb{R}[x]$ and $\mathbb{C}$ as vector spaces over $\mathbb{R}$ ($\mathbb{R}$-modules), and consider the evaluation map $\varphi\colon \mathbb{R}[x] \to \mathbb{C}$ defined by $\varphi\colon f(x) \mapsto f(i)$. This is the same function as the evaluation homomorphism defined in Example 2.47. Note that φ also defines a linear transformation ($\mathbb{R}$-module homomorphism):

$$\varphi(f(x) + rg(x)) = f(i) + rg(i) = \varphi(f(x)) + r\varphi(g(x))$$

for all $r \in \mathbb{R}$ and all $f(x), g(x) \in \mathbb{R}[x]$. Additionally, φ is surjective, and $\ker(\varphi) = (x^2 + 1) \subseteq \mathbb{R}[x]$ as demonstrated in Example 2.47. By the first isomorphism theorem for modules, we see that $\mathbb{R}[x]/(x^2 + 1)$ and $\mathbb{C}$ are not only isomorphic as fields (rings), but also as vector spaces over $\mathbb{R}$ ($\mathbb{R}$ modules).

5.3 Generation of submodules

We have explored ways to study the internal structure of modules in the previous section. We will now spend some time to study generation of submodules.

5.25 Definition (*Sums and generation*)
Let M be an R-module, and let $N_1, \ldots, N_k$ be submodules of M. Let $A \subseteq M$.

(1) The **sum** of $N_1, \ldots, N_k$, denoted by $N_1 + \cdots + N_k$, is the set of all finite sums of elements from the submodules:

$$N_1 + \cdots + N_k = \{n_1 + \cdots + n_k \mid n_i \in N_i\}$$

(2) The **submodule generated by** A, denoted by RA, is the set of all finite R-linear combinations:

$$RA = \{r_1 a_1 + \cdots + r_k a_k \mid k \in \mathbb{N},\ r_i \in R,\ a_i \in A\}$$

If $N \subseteq M$ is a submodule, and $N = RA$ for some subset $A \subseteq M$, then we say that A is a **generating set** for N, and that N is **generated by** A.

5.26 Remark

Note that the sum of submodules $N_1, \dots, N_k$ is the same as the submodule generated by $N_1 \cup \cdots \cup N_k$. On the other hand, if $A = \{a_1, \dots, a_k\}$, then the submodule generated by A is the same as the sum $Ra_1 + \cdots + Ra_k$.

5.27 Example

Class 24
02/28/2022

The notion of the submodule generated by a subset corresponds to the notion of the subspace spanned by a list of vectors. For example, consider the vector space $\mathbb{R}^3$ and the vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. The $\mathbb{R}$-submodule generated by $\{e_1, e_2\}$ is the same as the subspace spanned by $\{e_1, e_2\}$—both are the xy-plane. In addition, if we consider the $\mathbb{R}$-submodules $X = \mathbb{R}e_1$ and $Y = \mathbb{R}e_2$, then they are precisely the x-axis and the y-axis respectively. Their sum $X+Y$ is the same as the sum of X and Y as subspaces of $\mathbb{R}^3$—the xy-plane.

5.28 Example

Consider an abelian group G as a $\mathbb{Z}$-module. The $\mathbb{Z}$-submodules generated by $A \subseteq G$ is precisely the subgroup of G generated by A.

5.29 Definition (*Finitely-generated and cyclic submodules*)

Let M be an R-module.

(1) M is a **finitely generated R-module** if $M = RA$ for some finite subset $A \subseteq M$.
(2) M is a **cyclic R-module** if $M = Ra$ for some $a \in M$.

5.30 Example

For $\mathbb{Z}$-modules, the nomenclature corresponds exactly to finitely generated abelian groups and cyclic groups. In fact, one of the deepest theorem from group theory—the classification theorem of finitely generated abelian groups—can be generalized to a classification theorem of finitely generated modules over principal ideal domains. Because $\mathbb{Z}$ is a principal ideal domain, the theorem for abelian groups ($\mathbb{Z}$-modules) is really just a special case.

5.31 Example

Submodules of finitely generated modules need not be finitely generated. This happens when the ring R is particularly large. Let $R = \mathbb{R}[x_1, x_2, \dots]$ be the ring of polynomials over $\mathbb{R}$ on countably many indeterminates. Let $I = (x_1, x_2, \dots)$ be the ideal of R consisting of those polynomials with a zero constant term. Consider $M = R$ and $N = I$ as R-modules. Then M is finitely generated: $M = R\,1_R$. But N is not finitely generated. To see this, suppose that $N = RA$ for a finite subset $A \subseteq M = R$. Then the polynomials in A will only involve finitely many indeterminates. Thus there is some indeterminate x_k that does not occur in the polynomials in A. But $x_k \in N$, so

$$x_k = r_1 a_1 + \cdots r_n a_n$$

for some polynomials $r_1, \dots, r_n \in R$ and $a_1, \dots, a_n \in A$. At least one of r_i will

involve x_k, but then $r_i a_i$ will be degree at least 2 because a_i is non-constant, contradicting that x_k is degree 1.

5.32 The notion of modules generated by a subset corresponds to the notion of spans in vector spaces. While every module can be generated by itself, we would often times like to get a minimal set of generators. The notion of *basis* in vector spaces captures this minimality. In module theory, the corresponding concept is *free generators*.

5.33 Definition (*Free module and rank*)
Let R be a ring. An R-module M is **free** on the subset $A \subseteq M$ if every $x \in M$ is a *unique* R-linear combination of elements in A. That is, for all $x \in M$, there exist unique non-zero elements $r_1, \ldots, r_n \in R$ and unique $a_1, \ldots, a_n \in A$ such that $x = r_1 a_1 + \cdots + r_n a_n$ for some $n \in \mathbb{N}$. In this case, we say that A is a **basis** or a **set of free generators** for M. In the case that R is commutative, the cardinality of A is called the **rank** of the free module M.

5.34 Example
If $R = F$ is a field, and V is a vector space over F, then a set of free generators for V is precisely a basis in the usual sense of linear algebra. The rank of V over F is precisely the dimension $\dim_F(V)$.

5.35 Example
We saw in Example 5.24 that $\mathbb{C}$ is a vector space over $\mathbb{R}$. Evidently, $\{1, i\}$ is a basis, so $\operatorname{rank}_{\mathbb{R}}(\mathbb{C}) = \dim_{\mathbb{R}}(\mathbb{C}) = 2$.

5.36 Example
Not every module can be freely generated. For example, consider the $\mathbb{Z}$-module $M = \mathbb{Z}_2$. The additive identity $0 \in \mathbb{Z}_2$ does not need to be included in a set of free generators, because $0 = 0_{\mathbb{Z}} \cdot a$ for all $a \in M$. If we use $1 \in \mathbb{Z}_2$ as the generator, we notice that $1_2 = 1_{\mathbb{Z}} \cdot 1_2 = 3_{\mathbb{Z}} \cdot 1_2$. Thus $1 \in \mathbb{Z}_2$ is not a unique $\mathbb{Z}$-linear combination of generators, and $\mathbb{Z}_2$ is not freely generated.

5.37 Example
If M is a free $\mathbb{Z}$-module with a certain set $A = \{a_1, \ldots, a_n\}$ of free generators, then every $x \in M$ is a unique $\mathbb{Z}$-linear combination:

$$x = k_1 a_1 + \cdots + k_n a_n$$

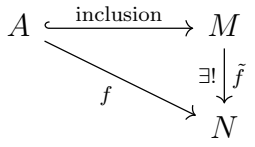
The coefficients $k_1, \ldots, k_n \in \mathbb{Z}$ can be placed in coordinates, thus giving an isomorphism $M \cong_{\mathbb{Z}} \mathbb{Z}^n$. This corresponds to *the free abelian group on n generators*.

In general, free R-modules takes the isomorphism class R^n for some $n \in \mathbb{N}$.

Class 25
03/02/2022

5.38 The existence of a free generating set for an R-module M vastly simplifies the definition of homomorphisms from M. We have seen this property at play in linear algebra—to define a linear transformation T from V to W, all we need to do is to specify the images $T(v_i)$ of some basis vectors $v_i \in V$, then use linearity to extend the definition to every vector $x \in V$. This also justify the use of matrices to define linear transformations with respect to bases.

5.39 Theorem (*Universal property of free modules*)
Let R be a ring, and let M be a free R-module with a basis $A \subseteq M$. The inclusion map $A \hookrightarrow M$ satisfies the following *universal property*: for every R-module N and for every function $f\colon A \to N$, there exists a unique R-module homomorphism $\tilde{f}\colon M \to N$ such that $\tilde{f}(a) = f(a)$ for all $a \in A$, that is, the following diagram commutes.



Proof
Because M is generated freely by A, then every $m \in M$ can be written as a *unique* finite R-linear combination of elements in A:

$$m = r_1 a_1 + \cdots + r_n a_n$$

for unique $r_i \in R$ and $a_i \in A$ for some $n \in \mathbb{N}$. Define the function $\tilde{f}$ by

$$\tilde{f}(r_1 a_1 + \cdots + r_n a_n) = r_1 f(a_1) + \cdots + r_n f(a_n).$$

$\tilde{f}$ is an R-module homomorphism. To see this let $m, n \in M$ and $t \in R$. Write

$$m = r_1 a_1 + \cdot + r_k a_k \text{ and } n = s_1 a_1 + \cdot s_k a_k,$$

allowing $m_i = 0$ or $n_j = 0$ if necessary. Then

$$\begin{aligned}
\tilde{f}(m + tn) &= \tilde{f}(r_1 a_1 + \cdots + r_k a_k + t(s_1 a_1 + \cdots + s_k a_k)) \\
&= \tilde{f}((r_1 + ts_1)a_1 + \cdots + (r_k + ts_k)a_k) \\
&= (r_1 + ts_1) f(a_1) + \cdots + (r_k + ts_k) f(a_k) \\
&= r_1 f(a_1) + \cdots + r_k f(a_k) + t(s_1 f(a_1) + \cdots + s_k f(a_k)) \\
&= \tilde{f}(m) + t\tilde{f}(n).
\end{aligned}$$

To see that $\tilde{f}$ is unique, suppose that $f'\colon M \to N$ is an R-module homomorphism that satisfies $f'(a) = f(a)$ for all $a \in A$. Let $m = r_1 a_1 + \cdots + r_n a_n$ Then being a homomorphism guarantees that

$$\begin{aligned}
f'(m) &= f'(r_1 a_1 + \cdots + r_n a_n) \\
&= r_1 f'(a_1) + \cdots + r_n f'(a_n) \\
&= r_1 f(a_1) + \cdots + r_n f(a_n) \\
&= \tilde{f}(m).
\end{aligned}$$

So $f' = \tilde{f}$ as desired. □

5.40 Corollary
Let R be a commutative ring. Every free R-module of rank n is isomorphic, as an R-module, to R^n.

Proof

This is left as an exercise in your homework assignment. □

5.41 Remark

A remarkable consequence of Theorem 5.39 is that homomorphisms between free modules of finite rank can be represented using matrices, just like we are used to doing for linear transformations between vector spaces.

Specifically, let R be a ring, let M and N be free R-modules of finite rank k and ℓ, respectively, and let $A = \{a_1, \ldots, a_k\} \subseteq M$ and $B = \{b_1, \ldots, b_\ell\} \subseteq N$ be bases. We can represent elements of M and N as vectors—specifically, an element $m \in M$ can be represented uniquely by a vector $(m_1, m_2, \ldots, m_k)_A$, where $m_i \in R$ and $m = m_1 a_1 + \cdots + m_k a_k$. The subscript A tells us that the components of the vector are the coefficients of the elements of the basis A. Likewise, elements $n \in N$ can be expressed uniquely as $(n_1, n_2, \ldots, n_\ell)_B$, where the components of the vector tell us that $n = n_1 b_1 + n_2 b_2 + \cdots + n_\ell b_\ell$.

An R-module homomorphism $f : M \to N$ is uniquely determined by the values of $f(a_i) \in N$ for each $1 \leq i \leq k$. Because B is a basis of N, we can write $f(a_i)$ uniquely as

$$f(a_i) = r_{1i} b_1 + r_{2i} b_2 + \cdots + r_{\ell i} b_\ell = (r_{1i}, r_{2i}, \ldots, r_{\ell i})_B$$

for some $r_{ij} \in R$. Extending to all of M, and using vector notation for elements of M and N, gives

$$f\colon \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{bmatrix}_A \mapsto \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\ell 1} & r_{\ell 2} & \cdots & r_{\ell k} \end{bmatrix}}_{\Phi} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{bmatrix}_A$$

where the operation on the right is the usual notion of matrix multiplication; the resulting vector is with respect to the basis B. Thus f can be represented using an $\ell \times k$ matrix Φ whose coefficients are elements of R.

5.42 Example

For a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$, with the standard basis vectors, the $m \times n$ matrix $[T]$ is the usual standard matrix from linear algebra—that is, the columns of $[T]$ are the images of the standard basis vectors $e_i \in \mathbb{R}^n$ under T.

5.43 Example

Let $\mathbb{Z}[x]_n$ denote the abelian group of all polynomials over $\mathbb{Z}$ of degree at most n under addition, viewed as a $\mathbb{Z}$-module. Note that $A = \{1, x, x^2, \ldots, x^n\}$ is a basis of $\mathbb{Z}[x]_n$ over $\mathbb{Z}$, so that $\operatorname{rank}_{\mathbb{Z}}(\mathbb{Z}[x]_n) = n + 1$.

Fix $n \geq 1$. Define $\varphi : \mathbb{Z}[x]_n \to \mathbb{Z}[x]_n$ by $\varphi(f(x)) = f'(x)$, where $f(x)$ is the (formal) derivative of $f(x)$—that is,

$$\varphi(a_0 + a_1 x + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}.$$

It is easy to verify that φ is a homomorphism, and so it is represented by an $(n+1)\times(n+1)$ matrix Φ. To find the matrix of Φ with respect to the basis A, note that $\Phi(x^j) = jx^{j-1}$ for all $0 \le j \le n$, so that we have

$$\Phi_{ij} = \begin{cases} j & \text{if } i = j-1, \\ 0 & \text{otherwise.} \end{cases}$$

For example, when $n = 3$, the matrix of φ with respect to A is

$$\Phi = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5.44 Remark

Many of the usual results from linear algebra carry over to R-module homomorphisms between free R-modules. For example, we know from linear algebra that a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ represented by a matrix A is invertible if and only if $\det(A) \neq 0$. The corresponding result in module theory tells us that an R-module homomorphism $f : M \to N$ between free R-modules of equal rank, with associated matrix Φ, is invertible if and only if $\det(\Phi)$ is a unit in R.

5.4 Direct sums

Class 26
03/04/2022

Recall that the *sum* of submodules $N_1, \dots, N_k$ of a module M is the set

$$N_1 + \cdots + N_k = \{n_1 + \cdots + n_k \mid n_i \in N_i\}$$

A special case of sums of submodules occurs when every element in the sum can be expressed as a *unique* list of summands.

5.45 Definition (*Direct sums*)

Let M be an R-module, and let $N_1, \dots, N_k$ be submodules of M. Let $N = N_1 + \cdots + N_k$. We say that N is the **(internal) direct sum** of $N_1, \dots, N_k$ if every element $n \in N$ can be written uniquely in the form $n = n_1 + n_2 + \cdots + n_k$ with $n_i \in N_i$. In this case, we write

$$N = N_1 \oplus N_2 \oplus \cdots \oplus N_k = \bigoplus_{i=1}^{k} N_i.$$

5.46 Example

Let's begin with one of our most familiar vector spaces—$\mathbb{R}^3$. Consider the subspaces $V = \operatorname{span}\{e_1, e_2\}$, $W = \operatorname{span}\{e_2, e_3\}$, and $X = \operatorname{span}\{e_1\}$. Evidently, $\mathbb{R}^3 = V + W$ and $\mathbb{R}^3 = X + W$. The first sum is not direct, while the second is.

- In $\mathbb{R}^3 = V + W$, we have

$$(1,1,1) = (1,0,0) + (0,1,1) = (1,1,0) + (0,0,1).$$

- In $\mathbb{R}^3 = X \oplus W$, because $\{e_1, e_2, e_3\}$ is a basis for $\mathbb{R}^3$, every $x \in \mathbb{R}^3$ is a unique linear combination $x = c_1e_1 + c_2e_2 + c_3e_3 = (c_1e_1) + (c_2e_2 + c_3e_3)$.

5.47 Example

The direct sum in Example 5.46 is a special case of *orthogonal complements*. If $V \leq \mathbb{R}^n$ is a subspace, and $V^\perp$ is its orthogonal complement, then $\mathbb{R}^n = V \oplus V^\perp$. This follows from the orthogonal decomposition theorem, which claims existence and uniqueness of vectors $x^\parallel$ and $x^\perp$ for every $x \in \mathbb{R}^n$ such that $x^\parallel \in V$, $x^\perp \in V^\perp$, and $x = x^\parallel + x^\perp$.

5.48 Example

Direct sums are related to free-ness of modules. If an R-module M is freely generated by a set $\{a_1, a_2, \ldots, a_k\}$, then $M = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_k$, because every element of M can be expressed uniquely as an R-linear combination of the elements $a_1, \ldots, a_k$.

However, the converse is not true. For example, consider $M = (\mathbb{Z}_2)^2$, viewed as a $\mathbb{Z}$-module. Then $M = R(1,0) \oplus R(0,1)$, because every element of M can be expressed uniquely as a sum of $(1,0)$ and $(0,1)$. But M is not freely generated by $(1,0)$ and $(0,1)$. For example we have $(1,0) = 1\cdot(1,0)+0\cdot(0,1) = 3\cdot(1,0)-4\cdot(0,1)$. So although the *elements* in the sum are unique, the $\mathbb{Z}$-*linear combination* (that is, the integer coefficients) do not have to be.

The following lemma is an alternative classification of direct sums.

5.49 Lemma

Let $N_1, \ldots, N_k$ be submodules of an R-module M. The sum $N_1 + \cdots + N_k$ is a direct sum if and only if $N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$ for all $1 \leq j \leq k$.

PROOF

($\Rightarrow$) We prove the contrapositive. Without loss of generality, suppose that $N_1 \cap (N_2 + \cdots + N_k) \neq 0$. Let $a \neq 0$ be an element in the intersection. Then $a = n_2 + \cdots + n_k$, where $n_i \in N_i$. But now 0 is simultaneously the sum of two different sets of summands:

$$\begin{aligned} 0 &= \quad 0 \quad + 0 + \cdots + 0, \quad \text{and} \\ 0 &= (-a) + n_2 + \cdots + n_k. \end{aligned}$$

Thus the sum $N_1 + \cdots + N_k$ is not direct.

($\Leftarrow$) Suppose that $N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$ for all $1 \leq j \leq k$. Let $a_i, b_i \in N_i$ such that $a_1 + \cdots + a_k = b_1 + \cdots + b_k$. Then for each j, we have

Class 27
03/07/2022

$$a_j - b_j = (b_1 - a_1) + \cdots + (b_{j-1} - a_{j-1}) + (b_{j+1} - a_{j+1}) + \cdots + (b_k - a_k)$$

The the right-hand side is in $N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k$, and the left-hand side is in N_j. By assumption, both sides are 0. Thus $a_j = b_j$ for all j. Hence, the sum $N_1 + \cdots + N_k$ is direct. □

5.50 Example

In Example 5.46, $V \cap W \neq 0$, while $X \cap W = 0$. Thus the sum $V + W$ is not direct, while $X \oplus W$ is direct.

More generally, $V \cap V^\perp = 0$, thus $\mathbb{R}^n = V \oplus V^\perp$ for any subspace V.

5.51 Example

As a $\mathbb{Z}$-module, $\mathbb{Z}_6$ can be expressed as the direct sum $\mathbb{Z}\{2\} \oplus \mathbb{Z}\{3\}$. Translating to the language of abelian groups, this is the same as saying that $\mathbb{Z}_6 \cong \mathbb{Z}_3 \times \mathbb{Z}_2$.

Direct sums are closely related to direct products. In fact, if there are only finitely many summands, the direct sum is isomorphic to the direct product.

5.52 Theorem

Let $N_1, \ldots, N_k$ be submodules of an R-module M. Define $\sigma\colon N_1 \times \cdots \times N_k \to M$ by $\sigma(n_1, \ldots, n_k) = n_1 + \cdots + n_k$. Then $M = \bigoplus_{i=1}^{k} N_i$ if and only if σ is an isomorphism of R-modules.

Proof

First verify that σ is an R-module homomorphism: for all $m_i, n_i \in N_i$ and for all $r \in R$,

$$\begin{aligned}\sigma(m + rn) &= (m_1 + rn_1) + \cdots + (m_k + rn_k) \\ &= (m_1 + \cdots + m_k) + r(n_1 + \cdots + n_k) \\ &= \sigma(m) + r\sigma(n).\end{aligned}$$

Now σ is an isomorphism if and only if it is a bijection. But σ being a bijection says precisely that every element $m \in M$ can be expressed uniquely as $\sigma(n) = n_1 + \cdots + n_k$ for $n = (n_1, \ldots, n_k) \in N_1 \times \cdots \times N_k$, which is what it means for M to be the direct sum of $N_1, \ldots, N_k$. □

5.53 Remark

The direct product allows us to define an "external direct sum," which does not require the summands to be submodules of some large module.

Given modules $M_1, M_2, \ldots, M_k$, define the (external) direct sum $M_1 \oplus \cdots \oplus M_k$ by identifying the module M_i with the submodule

$$\{0\} \times \{0\} \times \cdots \times M_i \times \cdots \times \{0\} \leq_R M_1 \times M_2 \times \cdots \times M_i \times \cdots \times M_k$$

This automatically forces the map $\sigma : M_1 \times M_2 \times \cdots \times M_k \to M_1 + M_2 + \cdots + M_k$ to be an isomorphism. Thus when $M_1, M_2, \ldots, M_k$ are modules that are *not* considered to be submodules of some ambient module, we may still define their direct sum to be equal to their product, but identify an element $m_i \in M_i$ with the corresponding element $(0, 0, \ldots, m_i, \ldots, 0)$ of the direct product.