LECTURE NOTES (PART 2), MATH 500 (FALL 2022)

CHARLES REZK

1. Review: Rings

A **ring** is a set R with binary operations $+$ and $\cdot$ satisfying **F 23 Sep** ring

- $(R, +)$ is an (additive) abelian group.
- $\cdot$ is associative: $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.
- Distributative laws: $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$ and $(x + y) \cdot z + (x \cdot z) + (y \cdot z)$.

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

A **ring with idenity** has an element $1 \in R$ such that $1 \cdot a = a = a \cdot 1$ for all $a \in R$. **ring with idenity**

Remark. I'll usually write "ab" instead of "$a \cdot b$", but other symbols are also sometimes used, e.g., "$a \times b$".

Also, we assume multiplication has "higher precedence" than addition, so we can write "$ac + bc$" instead of "$(ac) + (bc)$".

Remark (About rings without identity). Our book, like most textbooks, allows for the existence of rings without identity (meaning: without *multiplicative* identity). It is true there are some contexts where you want your definition of ring to allow for this. However in many contexts (e.g., most algebra, algebraic geometry, etc) rings are assumed *by default* to have an idenity. You *must* pay attention to context to determine whether rings are assumed to have identity.

Furthermore, Dummit and Foote give virtually no examples of rings without identity (except for ideals inside ring with identity, which for DF are considered to be "subrings", but not by most people). In most sections of Chapters 7–12, rings are assumed to have identity.

I will follow D& F's definitions here, and say "ring with identity" when I want an identity element (which is basically always). But I will also avoid calling something a ring if it does not have an identity.

To complicate matters: sometimes people drop the associativity condition on multiplication, and speak of "nonassociative rings". We won't do this.

Remark. Some people who require that rings have 1, also write **rng** (pronounced "rung") for rings which do not necessarily have an identity. (Get it?) **rng**

We have the following basic facts about a ring. The proofs rely on the distributive law.

Proposition. *Let R be a ring.*

1. $a0 = 0 = 0a$ *for all* $a \in R$.
2. $(-a)b = -(ab) = a(-b)$ *for all* $a, b \in R$.
3. $(-a)(-b) = ab$ *for all* $a, b \in R$.
4. *If R has an identity, it is unique and* $(-1)a = -a = a(-1)$ *for all* $a \in R$.

Example (Trivial ring). Let $R = \{0\}$ with obvious addition and multiplication. This is a ring. In fact, it is a commutative ring, and has an identity: $1 = 0$.

Conversely: the trivial ring is the only ring with identity such that $1 = 0$. (Proof: if $1 = 0$ then $a = a1 = a0 = 0$.)

Sometimes you want to exclude the trivial ring, so you speak of a "ring with identity $1 \neq 0$".

Date: November 2, 2022.

Units and zero divisors. Fix a ring R.

- If $1 \in R$, then $a \in R$ is a **unit** if there exists $b \in R$ such that $ab = 1 = ba$. If such b exists it is unique with this property, and we write $b = a^{-1}$.

  We write $R^\times$ for the set of units, which is a group under multiplication.

  unit
- $a \in R$ is a **zero divisor** if $a \neq 0$ and there exists $b \in R \setminus \{0\}$ such that either $ab = 0$ or $ba = 0$.

  Note that a zero divisor is never a unit.

  zero divisor
- $a \in R$ is a **non-zero divisor**, or **cancellable**, if $a \neq 0$ and it is not a zero-divisor.

  If a is cancellable, then either $ab = 0$ or $ba = 0$ imply $b = 0$.

  non-zero divisor
  cancellable

2. Review: Fields and domains

We have the following special types of rings.

- A **division ring** (or **skew-field**) is a ring with $1 \neq 0$ such that every non-zero element is a unit.

  division ring
  skew-field
- A **field** is a commutative division ring.

  field
- An **integral domain** (or just **domain**) is a commutative ring with $1 \neq 0$ which has no zero divisors.

  integral domain
  domain

  That is, a domain is a commutative ring with identity such that $1 \neq 0$, and $ab = 0$ implies either $a = 0$ or $b = 0$ for all $a, b \in R$.

  Note that every field is a domain.

  Also, sometimes people talk about noncommutative domains.

Proposition. *Every finite integral domain is a field.*

Proof. Any $a \in R \setminus \{0\}$ is cancellable, so $x \mapsto ax \colon R \to R$ is injective. Since $|R| < \infty$ it is bijective, so there exists $b \in R$ such that $ab = 1$. □

A **domain** (or **integral domain**) is a commutative ring in which $1 \neq 0$, and in which $xy = 0$ implies either x or y is 0. This gives cancellation: $xy = xz$ implies $y = z$ whenever $x \neq 0$. (It is possible to talk about non-commutative domains, but this is not standard terminology.)

domain
integral domain

Thus, a commutative $R \neq 0$ is a domain iff 0 is the only zero divisor, and is a field iff every non-zero element is a unit.

Proposition. *Every finite domain is a field.*

Proof. R is a domain exactly if for all non-zero $r \in R$, multiplication by r is injective.

R is a field exactly if for all non-zero $r \in R$, multiplication by r is bijective.

If R is finite, a function $\phi \colon R \to R$ is injective iff it is bijective. □

3. Review: Subrings

A **subring** of a ring R is a subset $S \subseteq R$ which is

subring

1. a subgroup with respect to $+$, and
2. is closed under multiplication.

I'll call it a **subring with identity** if in addition $1_R \in S$. (Warning: a subring $S \subset R$ can have an identity element which is not equal to 1_R.)

subring with identity

Exercise. Show that $S \subseteq R$ is a subring iff (i) $0 \in S$, (ii) $a, b \in S$ imply $a + b \in S$, (iii) $a \in S$ implies $-a \in S$, (iv) $a, b \in S$ imply $ab \in S$.

Exercise. Let R be an integral domain and $S \subseteq R$ a subring. Show that if S has an identity element, then $1_S = 1_R$.

4. Review: Basic examples of rings

Example (Trivial rings). Take any additive group $(R, +)$, and define a multiplication by $ab = 0$. This is always a commutative ring, but never a ring with identity (unless $R = \{0\}$).

Example. The integers $\mathbb{Z}$, a commutative ring with identity.

For each $n \in \mathbb{Z}$, the subset $n\mathbb{Z} = \{\, nk \mid k \in \mathbb{Z} \,\}$ is a subring, but not a subring with identity (unless $n = \pm 1$).

(Note that $0\mathbb{Z} = \{0\}$ has an identity element, which is 0, but this is not the same as the identity element of $\mathbb{Z}$, which is 1.)

Example. The rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, are fields.

Example. For $n \geq 1$ the set $\mathbb{Z}/n$ (or $\mathbb{Z}/n\mathbb{Z}$) of integers modulo n is a commutative ring with identity. It is a field iff n is a prime number.

Example. The ring $\mathbb{H}$ of **quaternions**. Here $\mathbb{H}$ is the set $\mathbb{R}^4$ of 4-tuples of real numbers, where we write "$a + bi + cj + dk$" instead of "(a, b, c, d)". Addition is componentwise. Multiplication is defined using the distributive law and the identities **quaternions**

$$i^2 = j^2 = k^2 = -1, \quad ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik.$$

The quaternions are a division ring. Proof: mostly straightforward, though associativity is tedious to check. Multiplicative inverses are given by

$$(a + bi + cj + dk)^{-1} = \frac{a - bi - cj - dk}{a^2 + b^2 + c^2 + d^2}.$$

Example. If X is any set and R a ring, then the set

$$\mathfrak{F}(X, R) = \{\text{functions } f\colon X \to R\}$$

is a ring, using componentwise addition and multiplication: $(f + g)(x) := f(x) + g(x)$, $(fg)(x) = f(x)g(x)$. It has identity if R does, is commutative if R is.

You are familiar with the case of $\mathfrak{F}(\mathbb{R}, \mathbb{R})$, the ring of real valued functions on $\mathbb{R}$. The subset of **functions with compact support** (i.e., $f\colon \mathbb{R} \to \mathbb{R}$ such that $\exists a \leq b$ with $f(x) = 0$ when $x \notin [a, b]$) is a subring without identity. **functions with compact support**

Example (Matrix rings). Let R be any ring and $n \geq 1$. The set $M_{n\times n}(R)$ of $n \times n$ matrices with entries in R is a ring, where addition and multiplication the usual ones for matrices: if $A = (a_{ij})$ and $B = (b_{ij})$, then $A + B = (a_{ij} + b_{ij})$ and $AB = (\sum_{k=1}^{n} a_{ik} b_{kj})$.

If R has identity then so does $M_{n\times n}(R)$, namely the identity matrix I.

If R is commutative, $M_{n\times n}(R)$ is not generally commutative (unless $n = 1$).

Product rings. Let R and S be rings. Then we can make the set $R \times S$ of ordered pairs into a ring, by componentwise addition and multiplication:

$$(r, s) + (r', s') := (r + r', s + s'), \qquad (r, s)(r', s') := (rr', ss').$$

This is called the **product ring**. If R and S are commutative, then $R \times S$ are commutative. If R and S have identity, then $R \times S$ has an identity, which is $(1_R, 1_S)$. **product ring**

5. Review: Homomorphisms of rings

Let R and S be rings. A **ring homomorphism** $\phi\colon R \to S$ is a function satisfying **ring homomorphism**

- $\phi(a + b) = \phi(a) + \phi(b)$,
- $\phi(ab) = \phi(a)\phi(b)$, and

Note: if R and S are rings with identity, it is possible that a ring homomorphism does not preserve identity, i.e., that $\phi(1_R) \neq 1_S$.

Warning. In contexts where rings are assumed to have 1, ring homomorphisms are also assumed to preserve identity. I will call these **identity preserving ring homomorphisms**.

identity preserving ring homomorphisms

Example. Let R and R' be rings, and let $S = R \times R'$ be the product ring. Then the functions $\phi_1 \colon R \to S$ and $\phi_2 \colon R' \to S$ defined by $\phi_1(x) = (x, 0)$ and $\phi_2(x) = (0, x)$ are ring homomorphisms.

However, if R and R' have identity $1 \neq 0$ (and so S has identity), neither of these homomorphism preserve identity.

Example. For any two rings, the function $\phi \colon R \to S$ defined by $\phi(x) = 0$ for all $x \in R$ is a ring homomorphism (under our definitions), called the **zero homomorphism**.

zero homomorphism

Warning. The zero homomorphism does not preserve identity, unless $S = \{0\}$.

The **image** $\phi(R) \subseteq S$ of a ring homomorphism $\phi \colon R \to S$ is just the image of ϕ as a function. The image is a subring of S.

image

The **kernel** $\operatorname{Ker} \phi = \{\, x \in R \mid \phi(x) = 0 \,\}$ of a ring homomorphism ϕ is the kernel as a map of additive groups.

kernel

An **isomorphism of rings** is a ring homomorphism $\phi \colon R \to S$ which is also a bijection. In this case, the inverse function $\phi^{-1} \colon S \to R$ is also an isomorphism of rings.

isomorphism of rings

6. Review: Ideals and quotient rings

Let R be a ring, and $I \subseteq R$ a subset. For $r \in R$ we write

$$rI := \{\, rx \mid x \in I \,\}, \qquad Ir := \{\, xr \mid x \in I \,\}.$$

We say that I is

- a **left ideal** if I is an subgroup of $(R, +)$, and if $rI \subseteq I$ for all $r \in R$,
- a **right ideal** if I is a subgroup of $(R, +)$, and if $Ir \subseteq I$ for all $r \in R$,
- a **two-sided ideal** if I is both a left ideal and a right ideal.

left ideal
right ideal
two-sided ideal

We will sometimes call two-sided ideals simply **ideals**.

ideals

Note: if R is a commutative ring, all three of these notions are the same.

The **unit ideal** of R is just $I = R$. Note that this is the only ideal (of any type) in R which contains 1_R.

unit ideal

Remark. With our definition of subring, all (two-sided) ideals are subrings. With the more common notion (all rings have 1, all subrings have this same 1), only the unit ideal is also a subring.

Given an ideal $I \subseteq R$, let R/I be the (additive) quotient group of $(R, +)$, so that addition on R/I is defined by

$$(a + I) + (b + I) = (a + b) + I, \qquad a, b \in R.$$

We can define a product on R/I by

$$(a + I)(b + I) := (ab) + I, \qquad a, b \in R.$$

This is well-defined exactly because I is a two-sided ideal. We call R/I with this ring structure the **quotient ring** of R by I.

quotient ring

Note: If R is commutative so is R/I. If R has identity, then so does R/I, and $1_{R/I} = 1 + I$.

The function $\pi \colon R \to R/I$ defined by $\pi(a) = a + I$ is a ring homomorphism, called the **quotient homomorphism**.

quotient homomorphism

7. First isomorphism theorem for rings

Here's the "homomorphism theorem". **M 26 Sep**

Proposition. *Let $\phi\colon R \to S$ be a homomorphism of rings, and $I \subseteq R$ an ideal. If $I \subseteq \operatorname{Ker}(\phi)$, then there exists a unique ring homomorphism $\overline{\phi}\colon R/I \to S$ such that $\overline{\phi}(a+I) = \phi(a)$.*

$$\begin{tikzcd}
R \arrow[r, "\phi"] \arrow[d, two heads, "\pi"'] & S \\
R/I \arrow[ru, dotted, "\overline{\phi}"'] &
\end{tikzcd}$$

The proof is straightforward: the formula for $\overline{\phi}$ is well-defined exactly because $I \subseteq \operatorname{Ker}\phi$.

Theorem (First isomorphism theorem for rings)**.** *If $\phi\colon R \to S$ is a ring homomorphism, then $\operatorname{Ker}(\phi)$ is an ideal of R, $\phi(R)$ is a subring of S, and we have an isomorphism $R/\operatorname{Ker}\phi \approx \phi(R)$ of rings.*

That is, the homomorphism ϕ factors through an isomorphism $\overline{\phi}\colon R/\operatorname{Ker}\phi \xrightarrow{\sim} \phi(R)$.

$$\begin{tikzcd}
R \arrow[rrr, "\phi"] \arrow[rd, two heads] & & & S \\
& R/\operatorname{Ker}\phi \arrow[r, "\overline{\phi}", "\sim"'] & \phi(R) \arrow[ru, tail] &
\end{tikzcd}$$

Note: if the rings have identity, and ϕ preserves identity, then so does $\overline{\phi}$.

Corollary. *Every ideal is the kernel of some ring homomorphism.*

Another way to describe what this is saying: there is an injective function

$$\{\text{ring homomorphisms } R/I \to S\} \rightarrowtail \{\text{ring homomorphisms } R \to S\}$$

defined by $\overline{\phi} \mapsto \overline{\phi} \circ \pi$, whose image is the set of $\phi\colon R \to S$ such that $I \subseteq \operatorname{Ker}\phi$.

8. Other isomorphism theorems for rings

There are more isomorphism theorems. You can generate them from the ones for groups, using the analogy "subgroup : subring :: normal subgroup : ideal".

Recall that for additive subgroups $A, B \subseteq R$, we let $A + B := \{\, a + b \mid a \in A,\ b \in B \,\}$.

Theorem (Second (diamond) isomorphism theorem for rings)**.** *Let $A \subseteq R$ be a subring, and $I \subseteq R$ an ideal.*

1. *$A + I$ is a subring of R.*
2. *I is an ideal of $A + I$.*
3. *$A \cap I$ is an ideal of A,*
4. *$A/(A \cap I) \approx (A + I)/I$.*

The isomorphism of (4) sends $x + (A \cap I) \mapsto x + I$.

Theorem (Third isomorphism theorem for rings)**.** *Let $I, J \subseteq R$ be ideals with $I \subseteq J$. Then:*

1. *J/I is an ideal in R/I, and*
2. *$R/J \approx (R/I)/(J/I)$.*

The isomorphism of (2) sends $x + J \mapsto (x + I) + (J/I)$.

Theorem (Fourth (lattice) isomorphism theorem for rings)**.** *Let $I \subseteq R$ be an ideal. Then we have inverse bijections*

$$\begin{tikzcd}
\{\, \textit{subrings } A \subseteq R \mid I \subseteq A \,\} \arrow[r, leftrightarrow, "\sim"] & \{\textit{subrings } \overline{A} \subseteq R/I\} \\
A \arrow[r, mapsto] & A/I \\
\pi^{-1}\overline{A} & \overline{A} \arrow[l, mapsto]
\end{tikzcd}$$

where $\pi^{-1}\overline{A} = \{\, x \in R \mid \pi(x) \in \overline{A} \,\}$. Furthermore for subrings $A, B \subseteq R$ with $I \subseteq A \cap B$, we have

1. *$A \subseteq B$ iff $A/I \subseteq B/I$.*
2. *$(A \cap B)/I = (A/I) \cap (B/I)$.*
3. *A is an ideal in R iff A/I is an ideal in R/I.*

9. Quadratic integer rings

See DF§7.1.

Let D be a squarefree integer (i.e., its prime factorization contains no repeated factors). Let

$$\mathbb{Q}(\sqrt{D}) := \{\, a + b\sqrt{D} \in \mathbb{C} \mid a, b \in \mathbb{Q} \,\}.$$

This is a subring of $\mathbb{C}$, commutative with identity. In fact, $\mathbb{Q}(\sqrt{D})$ is a field: inverses are given by

$$(a + b\sqrt{D})^{-1} = \frac{1}{a + b\sqrt{D}} \frac{a - b\sqrt{D}}{a - b\sqrt{D}} = \frac{a - b\sqrt{D}}{a^2 - b^2 D}.$$

(Note that $a^2 - b^2 D \neq 0$ since D is not a perfect square.)

Let

$$\mathbb{Z}[\sqrt{D}] := \{\, a + b\sqrt{D} \in \mathbb{C} \mid a, b \in \mathbb{Z} \,\}.$$

This is a subring of $\mathbb{Q}(\sqrt{D})$.

Example. $\mathbb{Z}[i] = \{\, a + bi \mid a, b \in \mathbb{Z} \,\} \subseteq \mathbb{C}$ is the ring of **Gaussian integers**. **Gaussian integers**

In some cases, we can produce a slightly larger subring of $\mathbb{Q}(\sqrt{D})$. If $D \equiv 1 \pmod 4$, let

$$\omega := \frac{1 + \sqrt{D}}{2}.$$

Note that $\omega^2 = (1 + D + 2\sqrt{D})/4 = k + \omega$ where $k = (D-1)/4 \in \mathbb{Z}$. Define

$$\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{D})} := \begin{cases} \{\, a + b\sqrt{D} \mid a, b \in \mathbb{Z} \,\} & \text{if } D \equiv 2, 3 \pmod 4, \\ \{\, a + b\omega \mid a, b \in \mathbb{Z} \,\} & \text{if } D \equiv 1 \pmod 4. \end{cases}$$

In either case this is a subring of $\mathbb{Q}(\sqrt{D})$. When $D \equiv 2, 3 \pmod 4$ then $\mathcal{O} = \mathbb{Z}[\sqrt{D}]$. But when $D \equiv 1 \pmod 4$ then $\mathcal{O} \supsetneq \mathbb{Z}[\sqrt{D}]$. In either case it is called the **ring of integers** in $\mathbb{Q}(\sqrt{D})$. **ring of integers**

Example. Let $D = -3 \equiv 1 \pmod 4$, so $\omega = (1 + i\sqrt{3})/2$, and $\omega^2 = \omega - 1$. We have rings

$$\mathbb{Z}[\sqrt{-3}] \subseteq \mathcal{O}_{\mathbb{Q}(\sqrt{-3})}.$$

The larger ring $\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{-3})}$ is the ring of **Eisenstein integers**. You can show that elements of $\mathcal{O}$ are those of the form $(a + bi\sqrt{3})/2$, with $a, b \in \mathbb{Z}$ and $a \equiv b \pmod 2$. **Eisenstein integers**

Proposition. *Suppose D is squarefree and $D \equiv 1 \pmod 4$. Let $\alpha = a + b\sqrt{D} \in \mathbb{Q}(\sqrt{D})$, $a, b \in \mathbb{Q}$. The following are equivalent.*

1. *$\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.*
2. *$a - b \in \mathbb{Z}$ and $2a \in \mathbb{Z}$.*

(3) *$a = m/2$ and $b = n/2$ where $m, n \in \mathbb{Z}$ and $m \equiv n \pmod 2$.*

Proof. Exercise. □

Define the **norm map** $\mathbb{Q}(\sqrt{D}) \to \mathbb{Q}$ by

norm map

$$N(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - b^2 D \in \mathbb{Q}.$$

Here are some easily verified properties.

- $N(\alpha) = 0$ iff $\alpha = 0$.
- $N(\alpha\beta) = N(\alpha)N(\beta)$.
- If $\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ then $N(\alpha) \in \mathbb{Z}$.

Proposition. *$\alpha \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ is a unit in $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$ iff $N(\alpha) = \pm 1$.*

Proof. Since $N \colon \mathcal{O}_{\sqrt{D}} \to \mathbb{Z}$ is multiplicative, it takes units to units.

If $\alpha = a + b\sqrt{D} \in \mathcal{O}_{\sqrt{D}}$ such that $N(\alpha) = \pm 1$, then $\alpha^{-1} = (a - b\sqrt{D})/N(\alpha)$, which is also seen to be an element of $\mathcal{O}$. □

In other words, units in $\mathcal{O}_{\mathbb{Q}\sqrt{D}}$ correspond to integer solutions of Pell's equation $x^2 - Dy^2 = \pm 1$.

Exercise. If $D < 0$, then $\mathcal{O}^\times$ is a finite group, while if $D > 0$ then $\mathcal{O}^\times$ is an infinite group.

10. Group rings and monoid rings

Let G be a monoid (with mulitplication written as "gh", and identity element "e"). For instance, G could be a group. (Since I won't use existence of inverses in what follows, everything works for a monoid too.)

Let R be a commmutative ring with identity.

Let $R[G]$ (or sometimes just RG, which is what the book writes) be the set of "finite formal sums"

$$\sum_{g \in G} a_g[g], \qquad a_g \in R.$$

More precisely, an element of $R[G]$ is a tuple $a = (a_g)_{g \in G}$ of elements $a_g \in R$ indexed by elements of G, such that all but finitely many a_g are 0.

We make $R[G]$ a ring by: "componentwise" addition, and with mulitplication defined using the distributive law and the formula

$$(a_i[g_i])(a_j[g_j]) := (a_i a_j)[g_k], \qquad a_i, a_j \in R, \quad g_i g_j = g_k \in G.$$

More generally,

$$\sum_{g \in G} a_g[g] \sum_{g' \in G} b_{g'}[g'] = \sum_{h \in G} c_h[h], \qquad c_h = \sum_{gg' = h} a_g b_{g'}.$$

The ring $R[G]$ has an identity element $1 = [e]$ where $e \in G$ is the identity element of G. The ring $R[G]$ is not usually commutative, but is commutative if R is a commutative ring and G has commutative multiplication. The ring $R[G]$ is called the **monoid ring** of the monoid G, and the **group ring** if G is a group.

monoid ring

group ring

Remark. If $G = \{g_1, \dots, g_n\}$ has a finite number n of elements, we can just write elements of $R[G]$ as $\sum_{k=1}^n a_k g_k$ with $a_k \in R$.0

Example. Let $G = \langle g \mid g^2 \rangle = \{e, g\}$, the group of order 2. Then $\mathbb{Q}[G]$ is the set of expressions of the form

$$a_0[e] + a_1[g], \qquad a_0, a_1 \in \mathbb{Q},$$

with operations

$$(a_0[e] + a_1[g]) + (b_0[e] + b_1[g]) = (a_0 + b_0)[e] + (a_1 + b_1)[g],$$

$$(a_0[e] + a_1[g])(b_0[e] + b_1[g]) = (a_0b_0 + a_1b_1)[e] + (a_0b_1 + a_1b_0)[g].$$

Note that $\mathbb{Q}[G]$ is not a field: $([e] + [g])([e] - [g]) = [1]^2 - [g]^2 = [e^2] - [g^2] = 0$.

11. Polynomial rings

Fix a commutative ring R with identity. Let $R[x]$ be the set of finite formal expressions

$$f = \sum_{k \in \mathbb{Z}_{\geq 0}} a_k x^k.$$

More precisely, an element of $R[x]$ is an infinite sequence $(a_0, a_1, a_2, \dots)$ of elements of R such that all but finitely many are 0. Thus, every element of $R[x]$ can be written as

$$\sum_{k=0}^{N} a_k x^k$$

for some N. Elements of $R[x]$ are **polynomials** in x with coefficents in R. **polynomials**

The **degree** of $f \in R[x]$ is the largest k such that $a_k \neq 0$. By convention we say that $\deg f = -\infty$, so $\deg\colon R[x] \to \mathbb{Z}_{\geq 0} \cup \{-\infty\}$. **degree**

We give $R[x]$ the structure of a ring, with "obvious" addition and multiplication. If $f = \sum a_k x^k$ and $g = \sum b_k x^k$, then

$$f + g = \sum_k (a_k + b_k) x^k, \qquad fg = \sum_k \left(\sum_{i=0}^{k} a_i b_{k-i} \right) x^k.$$

This is a commutative ring with identity, which is the constant polynomial 1.

Remark. Just as you can form group rings, you can form a monoid ring $R[M]$ for a monoid M. Let $M = \{\, x^k \mid k \in \mathbb{Z}_{\geq 0} \,\}$, with obvious product. Then the polynomial ring $R[x]$ is the same as the monoid ring $R[M]$.

A **constant polynomial** is one with $\deg f = 0$ or $f = 0$. These form a subring, which can be identified with the ring R. **constant polynomial**

Proposition. *Suppose R is an integral domain.*

1. *If $f, g \in R[x]$, then $\deg fg = \deg f + \deg g$.*
2. *$(R[x])^\times = R^\times$.*
3. *$R[x]$ is an integral domain.*

Proof. For (1), we use the convention $-\infty + n = -\infty = n + (-\infty)$ for any $n \in \mathbb{Z}_{\geq 0} \cup \{-\infty\}$. (This makes $\mathbb{Z}_{\geq 0} \cup \{-\infty\}$ into a *commutative semigroup* under $+$.)

The proof of (1) is straightforward, but uses that R is an integral domain: if R is not an integral domain, we only have $\deg fg \leq \deg f + \deg g$.

For (2), note that $\deg(1) = 0$, and that for $n, m \in \mathbb{Z}_{\geq 0} \cup \{-\infty\}$, if $n + m = 0$ then $n = m = 0$, and that $f \in R \smallsetminus \{0\}$ iff $\deg f = 0$.

Then (3) is clear, since $fg = 0$ iff $\deg f + \deg g = -\infty$, iff either $f = 0$ or $g = 0$. □

Since $R[x]$ is a ring, we can repeat the procedure, and consider $R[x][y]$; elements are $\sum a_j y^j$ where each $a_j = a_j(x) = \sum a_{ij} x^i$. You can show that $R[x][y]$ is isomorphic to $R[y][x]$. It is usual to call this ring $R[x, y]$, and write elements as finite sums of the form

$$\sum a_{ij} x^i y^j.$$

In particular, if R is a domain, so is $R[x_1, \dots, x_n]$.

12. Evaluation of polynomials

There is a recipe for describing homomorphisms out of a polynomial ring. **F 30 Sep**

Proposition. *Let R, S be a commutative rings with identity. Then for every pair (ϕ, a) consisting of*

(1) *a ring homomorphism $\phi\colon R \to S$ which preserves 1, and*
(2) *an element $a \in S$,*

there exists a unique ring homomorphism $\overline{\phi}\colon R[x] \to S$ which preserves 1, such that

$$\overline{\phi}(r) = \phi(r) \quad \textit{for } r \in R \subseteq R[x],, \qquad \textit{and} \qquad \overline{\phi}(x) = a.$$

Proof. Existence. Define a function $\overline{\phi}\colon R[x] \to S$ by

$$\overline{\phi}\Big(\sum_{k=0}^{n} c_k x^k\Big) := \sum_{k=0}^{n} \phi(c_k) a^k.$$

It is a straightforward exercise to show that this $\overline{\phi}$ is a ring homomorphism preserving identity.

Uniqueness. We can show from the hypotheses that any $\overline{\phi}\colon R[x] \to S$ with the given properties has the formula I gave above.

$$\begin{aligned}\overline{\phi}\Big(\sum_{k=0}^{n} c_k x^k\Big) &= \sum_{k=0}^{n} \overline{\phi}(c_k x^k) \\ &= \sum_{k=0}^{n} \overline{\phi}(c_k)\overline{\phi}(x)^k \\ &= \sum_{k=0}^{n} \phi(c_k) a^k.\end{aligned}$$

□

A special case is when $S = R$ and $\phi = \mathrm{id}$. Then $\overline{\phi}\colon R[x] \to R$ is a ring homomorphism defined by

$$f = \sum_{k=0}^{n} c_k x^k \quad \longmapsto \quad \sum_{k=0}^{n} c_k a^k.$$

We usually call the output $f(a)$, and we call the homomorphism **evaluation at** a, sometimes written $\mathrm{ev}_a\colon R[x] \to R$. That this is a ring homomorphism means you have formulas **evaluation at a**

$$(f+g)(a) = f(a) + g(a), \quad f(a)g(a) = (fg)(a), \qquad f, g \in R[x], \quad a \in R,$$

which you already know.

Recall that $\mathfrak{F}(R, R)$ is the ring of functions $f\colon R \to R$, with "pointwise" operations. Let

$$\psi\colon R[x] \to \mathcal{F}(R, R), \qquad \phi(f)\big(a\big) := \mathrm{ev}_a(f) = f(a).$$

Exercise. ψ is a ring homomorphism, preserving 1.

The function ψ turns a polynomial into a function on the ring. But ψ isn't generally injective, even if R is a field. Thus, polynomials are not really the same thing as functions.

Example. Let $R = \mathbb{F}_p = \mathbb{Z}/p$, the integers modulo p. Let $f = x^p - p \in R$. Then $\mathrm{ev}_a(f) = a^p - a = 0$ for all $a \in \mathbb{F}_p$, by Fermat's little theorem. Thus $\psi\colon \mathbb{F}_p[x] \to \mathcal{F}(\mathbb{F}_p, \mathbb{F}_p)$ is not injective.

(Exercise: describe the kernel of ψ.)

13. Ideals generated by subsets

Given a subset $A \subseteq R$, we let $(A) \subseteq R$ be the **ideal generated by** A, defined to be

ideal generated by A

$$(A) := \bigcap_{\substack{\text{ideals } I \subseteq R \\ \text{such that } A \subseteq I}} I$$

Exercise: (A) is an ideal in R, and is the smallest ideal containing the set A.

We define the following subsets of R.

- $RA = \{\, r_1a_1 + \cdots + r_ka_k \mid r_i \in R,\ a_i \in A,\ k \geq 0 \,\}$.
- $AR = \{\, a_1r_1 + \cdots + a_kr_k \mid r_i \in R,\ a_i \in A,\ k \geq 0 \,\}$.
- $RAR = \{\, r_1a_1r_1' + \cdots + r_ka_kr_k' \mid r_i, r_i' \in R,\ a_i \in A, k \geq 0 \,\}$.

In each case we always assume 0 is an element of these sets (corresponding to sums with $k = 0$).

If $A = \{a_1, \dots, a_n\}$ is a finite set, we write $(a_1, \dots, a_n)$ for A.

Proposition. *Suppose R is a ring with 1, and let A be a subset of a ring R. Then*

$$(A) = RAR.$$

If R is a commutative ring with identity, then

$$(A) = RA = AR = RAR.$$

Proof. For the first claim, note that

1. RAR is an ideal in R, since it is an additive subgroup, and $rRAR \subseteq RAR$ and $RARr \subseteq RARr$ for all $r \in R$. (This uses that $Rr \subseteq R$ since $1 \in R$.)
2. We have $A = 1A1 \subseteq RAR$ by definition, and therefore $(A) \subseteq I$ since I is an ideal.
3. We have $A \subseteq (A)$ by definition, and so $RAR \subseteq (A)$ since (A) is an ideal.

This proves $(A) = RAR$.

The second claim is clear. □

Remark. If R does not have 1, this doesn't work, because A might not be a subset of RAR. Instead, you can say

$$(A) = \{\, a_1 + \cdots + a_k - a_1' - \cdots - a_\ell' + x + y + z \mid a_i, a_j' \in A,\ x \in RA,\ y \in AR,\ z \in RAR \,\}.$$

Remark. It is easy to see that RA is the *smallest left ideal* containing A, and AR is the *smallest right ideal* containing A.

14. Principal ideals

A **principal ideal** is an ideal I such that $I = (a)$ for some single element $a \in R$.

principal ideal

In a ring with 1,

$$(a) = \{\, r_1ar_1' + \cdots + r_kar_k' \mid r_i, r_i' \in R,\ k \geq 0 \,\} = RaR.$$

When R is commutative, the distributive law lets simplifies this to

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

Not every ideal is principal.

Example. In $R = \mathbb{Z}[x]$, let $I = (2, x)$. I claim that I is not principal. The proof uses the degree function $\deg\colon \mathbb{Z}[x] \to \mathbb{Z}_{\geq 0} \cup \{-\infty\}$, and in particular the property that $\deg(fg) = \deg f + \deg g$.

We suppose $I = (p) = pR$ for some $p \in \mathbb{Z}[x]$ and derive a contradiction. Since $2, x \in I$ there must exist $f, g \in \mathbb{Z}[x]$ such that

$$2 = pf, \qquad x = pg.$$

Applying degree, we find that

$$0 = \deg p + \deg f, \quad 1 = \deg p + \deg g, \qquad \Longrightarrow \qquad \deg p = 0, \quad \deg f = 0, \quad \deg g = 1.$$

Thus both p and f are constant polynomials, so clearly $p, f \in \{\pm 1, \pm 2\}$, and $g = a + bx$ for some $a, b \in \mathbb{Z}$ with $b \neq 0$.

If $p = \pm 2$, then $x = pg = \pm 2(a + bx) = \pm(2a + 2bx)$ whence $1 = \pm 2b$, which is clearly impossible. Thus $p = \pm 1$. But then $1 \in I$, whence $1 = 2m + xn$ for some $m, n \in \mathbb{Z}[x]$. Evaluating at 0 gives $1 = 2m(0) + 0n(0) = 2m(0)$, which is impossible since $m(0) \in \mathbb{Z}$.

Example. Let $R = F[x, y]$, with F any field. Then the ideal $I = (x, y)$ is not principal. (Exercise.)

15. Ideals and fields

Proposition. *A non-zero commutative ring R with identity is a field if and only if the only $R \neq \{0\}$ and the only ideals are $\{0\}$ and R.*

Proof. We start with an observation: $a \in R$ is a unit iff $Ra = R$. Here's the proof. $\Longrightarrow$ If $a \in R^\times$, then $1 = a^{-1}a \in Ra$, so Ra is the unit ideal. $\Longleftarrow$ If $Ra = R$ then $1 \in Ra$, so there exists $b \in R$ such that $1 = ba$, so $b = a^{-1}$ and a is a unit.

Now we prove the proposition. If R is a field, then $1 \neq 0$, so $R \neq \{0\}$. If $I \subseteq R$ is an ideal and $I \neq \{0\}$, choose any $a \in I$ with $a \neq 0$. Then a is a unit, so $R = Ra \subseteq I$, so $I = R$.

Conversely, suppose $R \neq \{0\}$ with only ideals $\{0\}$ and R. Since $R \neq \{0\}$ we have $1 \neq 0$. If $a \in R \smallsetminus \{a\}$ then Ra is an ideal, and since $Ra \neq \{0\}$ then $Ra = R$, so a is a unit. □

Proposition. *Any non-zero ring homomorphism $\phi\colon F \to R$ from a field to a ring is injective.*

Proof. If $\phi \neq 0$ then $\operatorname{Ker} \phi \neq R$, and since the kernel is an ideal we must have $\operatorname{Ker} \phi = \{0\}$. □

There is a non-commutative analogue.

Exercise. A non-zero ring R with identity is a division ring if and only if the only left ideals and the only right ideals are 0 and R.

Exercise. Let F be a field and $R = M_{n \times n}(F)$. Then the only 2-sided ideals of R are 0 and R, but R is not a division ring.

16. Maximal and prime ideals

Maximal ideals. Let R be a ring with 1.

An ideal $M \subseteq R$ is **maximal** if $M \neq R$ and the only ideals containing M are M and R. (It should be called a "maximal proper ideal", but isn't.) **maximal**

Note: $M \neq R$ is equivalent to $1 \notin M$.

Proposition. *Let R be commutative ring with 1. An ideal $M \subseteq R$ is maximal if and only if R/M is a field.*

Proof. By the lattice isomorphism theorem, the ideals of R/M correspond exactly to ideals of R which contain M. So M is maximal in R iff R/M has exactly two ideals, i.e., is a field. □

Prime ideals. Let R be a commutative ring with 1.

An ideal $P \subseteq R$ is **prime** if $P \neq R$, and if $ab \in P$ implies either $a \in P$ or $b \in P$. **prime**

Example. In $\mathbb{Z}$, the ideal (n) is prime iff $n = \pm p$ where p is a prime number.

Remark. There is a formulation of "prime ideal" for non-commutative rings, but it is more complicated to state. See, e.g., wikipedia.

Proposition. *Let R be a commutative ring with 1. Then $P \subseteq R$ is prime if and only if R/P is an integral domain. All maximal ideals are prime.*

Proof. The first statement is immediate from the definition of integral domain: D is an integral domain iff $1 \neq 0$ and $xy = 0$ implies $x = 0$ or $y = 0$ for all $x, y \in D$. Applied to $D := R/P$, this recovers the the condition for the ideal P to be prime.

For the second statement, note that if M is maximal, then R/M is a field and thus an integral domain. □

Existence of maximal ideals. Here is the big theorem.

Theorem. *In any ring with identity, every proper ideal is contained in a maximal ideal.*

Note that this implies that every non-0 ring (with identity) has at least one maximal ideal. (The ring 0 has no proper ideals, so the proposition doesn't apply to it.) I will prove this soon.

Corollary. *Every non-zero commutative ring with identity has a quotient ring which is a field.*

17. Zorn's lemma

The proof of existence of maximal ideals uses "Zorn's lemma". This is a statement about partially ordered sets which is a non-trivial consequence of the axiom of choice. **M 3 Oct**

A **partial order** on a set X is a relation $\leq$ on X which is **partial order**

- *reflexive:* $x \leq x$ for all $x \in X$,
- *anti-symmetric:* $x \leq y$ and $y \leq x$ imply $x = y$ for all $x, y \in X$,
- *transitive:* $x \leq y$ and $y \leq z$ implies $x \leq z$ for all $x, y, z \in X$.

A good example of a partial order is $(\mathcal{P}S, \subseteq)$, the set of all subsets of a set S with the set-containment relation. Note that there can exist $x, y \in X$ such that neither $x \leq y$ nor $y \leq x$ hold.

For a poset $(X, \leq)$, a **chain** is a subset $C \subseteq X$ such that $x, y \in C$ implies either $x \leq y$ or $y \leq x$. **chain**

For $S \subseteq X$, an **upper bound** is $u \in X$ such that $s \leq x$ for all $s \in X$. Similarly, **lower bound**. Note that the upper/lower bounds of a subset need not be elements of the subset. **upper bound** **lower bound**

A **maximal element** of X is an $m \in X$ such that $m \leq x$ implies $m = x$ for all $x \in X$. Note that X can have multiple maximal elements, so that distinct ones are necessarily non-comparable with each other. **maximal element**

(Note: upper bounds of S do not need to be in S, only in X. Maximal elements need not be unique.)

(E.g., consider the set X of linearly independent subsets of $\mathbb{R}^n$, ordered by containment. The maximal elements of X are precisely the bases.)

Theorem (Zorn's Lemma). *Let X be a non-empty poset. If every non-empty chain in X has an upper bound, then X has a maximal element.*

I'm not going to prove it (it's really a statement about set theory, which relies essentially on the Axiom of Choice)[1].

Remark. Here is another, equivalent formulation of Zorn's lemma, which you often see.

> Let X be a poset. If every chain in X has an upper bound, then X has a maximal element.

Note that $C = \varnothing$ is a always a chain in X, and that any $u \in X$ is an upper bound of C. Applied to non-empty X, the above statement is clearly equivalent to our statement of Zorn's lemma. When X is empty, the above statement is vacuously true, since in that case the empty chain is the only chain, and it has no upper bound, so the hypothesis is not satisfied.

In practice, when checking "this chain has an upper bound", the case of the empty chain often works out differently than non-empty chains, so it can be more useful to use a formulation of Zorn's lemma which excludes the case of empty chains.

[1] See `https://faculty.math.illinois.edu/~dan/ShortProofs/Zorn.pdf` for a short proof.

18. Proof of the maximal ideal theorem

Now we apply this to the existence of maximal ideals.

Proof of the maximal ideal theorem. Let $I \subseteq R$ be a proper ideal. Let X be the set of all proper ideals of R which contain I.

We have $X \neq \varnothing$ since $I \in X$.

Consider a chain $\mathcal{C} \subseteq X$ with $\mathcal{C} \neq \varnothing$; we want to show that $\mathcal{C}$ has an upper bound. Let $J = \bigcup_{A \in \mathcal{C}} A$. I claim that J is a proper ideal of R.

J is an ideal. Since $\mathcal{C}$ is non-empty, so is J.

If $a, b \in J$, there exist $A, B \in \mathcal{C}$ such that $a \in A$ and $b \in B$. Since $\mathcal{C}$ is a chain, either $A \subseteq B$ or $B \subseteq A$, so $a \pm b \in J$. If $a \in J$ and $r \in R$, then there exists $A \in \mathcal{C}$ such that $a \in A$, and thus $ra, ar \in A \subseteq J$.

J is a proper subset. If not, then $1 \in J$, but then $1 \in A$ for some $A \in \mathcal{C}$, whence $A = R$, contradicting the hypothesis that X consists of proper ideals.

Thus, $J \in X$, so is an upper bound of the chain.

Zorn's lemma applies to show that X has a maximal element M. □

19. Rings of fractions

Note: I'm going to do things a little more generally than DF§7.5. (Because it is important.) The more general construction is in DF§15.4.

Let R be a commutative ring with 1. Let $D \subseteq R$ be a **multiplicatively closed subset**, i.e., a subset such that

multiplicatively closed subset

(1) $1 \in D$, and
(2) $a, b \in D$ implies $ab \in D$.

Example. Consider the set $D := R \smallsetminus \{0\}$ of non-zero elements in R. We have

(1) $1 \in D$ iff $1 \neq 0$, and
(2) $a, b \in D$ implies $ab \in D$ iff $ab = 0$ implies $a = 0$ or $b = 0$.

Thus, $D = R \smallsetminus \{0\}$ is multiplicatively closed iff R is an integral domain.

Given such a pair (R, D), let J be the subset

$$J := \{\, r \in R \mid \exists d \in D,\ dr = 0 \,\},$$

i.e., the elements of R which are "killed" by *some* element of D. Note that $J = \{0\}$ if and only if all elements of D are non-zerodivisors. The following exercise describes the two key properties of the subset J.

Exercise. Show that J is an ideal of R. Then show that J has the following property: for any $d \in D$ and $r \in R$, we have that $dr \in J$ implies $r \in J$.

We will construct a new ring $D^{-1}R$, called a **ring of fractions** together with a ring homomorphism $\psi\colon R \to D^{-1}R$ which preserves identity. The idea is that elements of $D^{-1}R$ are going to be *formal quotients* "r/d", where $r \in R$ and $d \in D$. The ring homomorphism $\psi\colon R \to D^{-1}R$ sends r to "$r/1$". The kernel of the homomorphism is the ideal J. In particular, when D has no zerodivisors, you can identify R with the subring $\psi(R)$ of $D^{-1}R$. This is the special case described in DF§7.5.

ring of fractions

The ring $D^{-1}R$ will have the following properties.

Proposition. *Let R be a commutative ring with 1 and $D \subseteq R$ a multiplicatively closed subset.*

(1) *If $d \in D$ then $\psi(d)$ is a unit in $D^{-1}R$.*
(2) *Every element of $D^{-1}R$ can be written $\psi(r)\psi(d)^{-1}$ for some $r \in R$, $d \in D$.*
(3) $\operatorname{Ker}\psi = J = \{\, r \in R \mid \exists\, d \in D,\ dr = 0 \,\}$.

Thus ψ identifies R/J with the subring $\psi(R) \subseteq D^{-1}R$.

I will construct the fraction ring and prove this soon.

Example (Fraction fields). For R an integral domain, let $D = R \smallsetminus \{0\}$ and define

$$Q = \operatorname{Frac}(R) := D^{-1}R.$$

Since R is an integral domain, (3) implies that $\operatorname{Ker}\psi = \{0\}$, so ψ is injective. It is convenient to identify R with its image in Q under ψ. Then every element of Q has the form ab^{-1}, where $a, b \in R$ and $b \neq 0$. In particular, Q is a field, since $(ab^{-1})(ba^{-1}) = 1$ when both $a \neq 0$ and $b \neq 0$.

We call $Q = \operatorname{Frac}(R)$ the **field of fractions** of R.

field of fractions

For instance, $\mathbb{Q} \approx \operatorname{Frac}(\mathbb{Z})$.

20. Construction of rings of fractions

Define an relation $\sim$ on the set $R \times D = \{\, (r, d) \mid r \in R,\ d \in D \,\}$, by

$$(r_1, d_1) \sim (r_2, d_2) \qquad \Longleftrightarrow \qquad \exists d \in D,\ d(r_1d_2 - r_2d_1) = 0.$$

That is, $(r_1, d_1) \sim (r_2, d_2)$ iff $r_1d_2 - r_2d_1 \in J$. Note: this is *almost* the familiar criterion for equality of fractions $\frac{r_1}{d_1} = \frac{r_2}{d_2}$, except that in general we can only require a congruence modulo J.

Remark. If all elements of D are non-zero divisors, then the condition for "$(r_1, d_1) \sim (r_2, d_1)$" simplifies to "$r_1d_2 - r_2d_1 = 0$". This special case is the construction described in DF§7.5.

Lemma. *The relation $\sim$ is an equivalence relation.*

Proof. Reflexivity and symmetry are immediate. For transitivity: if $r_1d_2 - r_2d_1, r_2d_3 - r_3d_2 \in J$, then

$$d_3(r_1d_2 - r_2d_1) + d_1(r_2d_3 - r_3d_2) \in J,$$

since J is an ideal. But this expression reduces to

$$d_3(r_1d_2 - r_2d_1) + d_1(r_2d_3 - r_3d_2) = r_1d_2d_3 - r_2d_1d_3 + r_2d_1d_3 - r_3d_1d_2 = d_2(r_1d_3 - r_3d_1)$$

so $d_2(r_1d_3 - r_3d_1) \in J$. But J has the property (see Exercise above) that $dx \in J$ implies $x \in J$ whenever $d \in D$, so we conclude that $r_1d_3 - r_3d_1 \in J$. □

Let's write "$[r/d]$" for the equivalence class of (r, d), and let $D^{-1}R$ be the set of such equivalence classes. We define operations

$$[r_1/d_1] + [r_2/d_2] := [(r_1d_2 + r_2d_1)/d_1d_2],$$
$$[r_1/d_1] \cdot [r_2/d_2] := [r_1r_2/d_1d_2].$$

Lemma. *These operations are well-defined, and give $D^{-1}R$ the structure of a commutative ring with identity, with*

$$0_{D^{-1}R} = [0/1], \qquad 1_{D^{-1}R} = [1/1], \qquad -[r/d] = [-r/d].$$

Furthermore, the function $\psi\colon R \to D^{-1}R$ defined by $\psi(r) := [r/1]$ is a ring homomorphism which preserves identity.

Proof. This is a straightforward (but lengthy) exercise. I'll just show here that the operations are well-defined.

Suppose $[r_1/d_1] = [r_1'/d_1']$ and $[r_2/d_2] = [r_2'/d_2']$, i.e., that $r_1d_1' - r_1'd_1, r_2d_2' - r_2'd_2 \in J$. Then verify that

$$(r_1d_2 + r_2d_1)d_1'd_2' - (r_1'd_2'r_2'd_1')d_1d_2 = (r_1d_1' - r_1'd_1)d_2d_2' + (r_2d_2' - r_2'd_2)d_1d_1' \in J,$$

and

$$r_1r_2d_1'd_2' - r_1'r_2'd_1d_2 = (r_1d_1' - r_1'd_1)r_2d_2' + (r_2d_2' - r_2'd_2)r_1'd_1 \in J,$$

and so $[(r_1d_2 + r_2d_1)/d_1d_2] = [(r_1'd_2' + r_2'd_1')/d_1'd_2']$ and $[r_1r_2/d_1d_2] = [r_1'r_2'/d_1'd_2']$. It remains to verify the axioms for a commutative ring with 1, and that ψ is a ring homomorphism, but this is straightforward. □

Now I'll prove the properties I promised.

Proof.

1. *$d \in D$ implies $\psi(d) \in D^{-1}R$ is a unit.* Clear, since $[d/1][1/d] = [d/d] = [1/1] = 1$.
2. *All elements of $D^{-1}R$ can be written as $\psi(r)\psi(d)^{-1}$ for some $r \in R$, $d \in D$.* Clear, since $[r/d] = [r/1][1/d] = [r/1][d/1]^{-1}$.
3. $\operatorname{Ker}\psi = J$. By construction $[r/1] = [0/1]$ iff $r1 - 01 = r$ is in J.

□

21. Examples of rings of fractions

We have seen that if R is an integral domain, then $\operatorname{Frac}(R) := (R \smallsetminus \{0\})^{-1}R$ is the fraction field of R. We note the fraction field of a polynomial ring over a field:

$$F(x_1, \dots, x_n) := \operatorname{Frac}(F[x_1, \dots, x_n]),$$

called the **field of rational functions**.

field of rational functions

Here is the easiest example where $\psi\colon R \to D^{-1}R$ is not injective

Example (Inverting 0 is deadly). If $D \subseteq R$ is a multiplicatively closed subset with $0 \in D$, then $\operatorname{Ker}\psi = R$, so $D^{-1}R = \{0\}$ is the trivial ring.

Example (Inverting nilpotent elements is deadly). Suppose there exists $a \in D$ such that $a^n = 0$ for some $n \geq 1$. Then $\operatorname{Ker}\psi = R$, so $D^{-1}R = \{0\}$ is the trivial ring.

Notation: if $a \in R$ and $D = \{\, a^n \mid n \geq 1 \,\}$, we typically write $a^{-1}R$ for $D^{-1}R$.

Example (Laurent polynomials). Let $R =$ commutative ring with 1. Let $D = \{\, x^k \mid k \in \mathbb{Z}_{\geq 0} \,\} \subseteq R[x]$. We have

$$R[x, x^{-1}] := D^{-1}R[x] = x^{-1}R[x].$$

This can be identified with the set of *finite* formal sums $\sum_{k\in\mathbb{Z}} a_k x^k$, with $a_k \in F$, and $a_k = 0$ for all but finitely many $k \in \mathbb{Z}$.

Example (Local rings). Let $P \subseteq R$ be a prime ideal (of a commutative ring with 1). Then $D := R \smallsetminus P$ satisfies $1 \in D$ and $a, b \in D$ implies $ab \in D$. Thus we can form the **localization**

localization

$$R_P := (R \smallsetminus P)^{-1}R$$

of R at P.

For example, if p is a prime number, then

$$\mathbb{Z}_{(p)} = \{\, a/b \in \mathbb{Q} \mid a, b \in Z,\ p \nmid b \,\}.$$

22. Universal property of rings of fractions

There is a recipe for constructing ring homomorphisms out of a ring of fractions.

W 5 Oct

Proposition. *Let $\phi\colon R \to S$ be a ring homomorphism between commutative rings preserving 1. If $\phi(D) \subseteq S^\times$, then there there exists a unique ring homomorphism $\overline{\phi}\colon D^{-1}R \to S$ such that $\overline{\phi} \circ \psi = \phi$.*

$$\begin{tikzcd}
R \arrow[r, "\phi"] \arrow[d, "\psi"'] & S \\
D^{-1}R \arrow[ur, dotted, "\exists!", "\overline{\phi}"']
\end{tikzcd}$$

In other words, for every ring S with 1 there is a bijection

$$\left\{ \begin{array}{c} \text{ring homomorphisms} \\ \overline{\phi}\colon D^{-1}R \to S \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{ring homomorphisms} \\ \phi\colon R \to S \\ \text{such that } \phi(D) \subseteq S^{\times} \end{array} \right\}$$

by $\overline{\phi} \mapsto \overline{\phi} \circ \psi$.

Proof. Suppose given $\phi\colon R \to S$ such that $\phi(D) \subseteq S^{\times}$. For existence, define $\overline{\phi}([r/d]) := \phi(r)\phi(d)^{-1}$. We need to check that this is well-defined, i.e., if $[r_1/d_1] = [r_2/d_2]$ then $\phi(r_1)\phi(d_1)^{-1} = \phi(r_2)\phi(d_2)^{-1}$. There exists $d \in D$ such that $dr_1d_2 = dr_2d_1$, so $\phi(d)\phi(r_1)\phi(d_2) = \phi(d)\phi(r_2)\phi(d_1)$. Since $\phi(d)$ is a unit we can cancel it, so $\phi(r_1)\phi(d_2) = \phi(r_2)\phi(d_1)$. Then

$$\phi(r_1)\phi(d_1)^{-1} = \phi(r_1)\phi(d_2)\phi(d_2)^{-1}\phi(d_1)^{-1} = \phi(r_2)\phi(d_1)\phi(d_2)^{-1}\phi(d_1)^{-1} = \phi(r_2)\phi(d_2)^{-1}$$

as desired.

For uniqueness, since every element $x \in D^{-1}R$ has the form $x = \psi(r)\psi(d)^{-1}$ for some $r \in R$ and $d \in D$, the hypothesis on $\overline{\phi}$ implies that $\overline{\phi}(x) = \phi(r)\phi(d)^{-1}$. □

Proposition. *Suppose F is a field and $R \subseteq F$ is a non-trivial subring with $1_F \in R$. Then R is an integral domain, and the fraction field $Q = \operatorname{Frac}(R)$ of R is isomorphic to the smallest subfield of F containing R.*

Proof. It is clear that R is a commuative ring with 1, and has cancellation since F does.

Let $\phi\colon R \rightarrowtail F$ be the inclusion homomorphism. Since ϕ is injective, every non-zero element of R is sent to a unit in F, so ϕ extends (uniquely) to a homomorphism $\overline{\phi}\colon Q \to F$ (preserving 1). Since Q is a field and $\overline{\phi} \neq 0$, $\overline{\phi}$ is injective, so we get an isomorphism $Q \approx \overline{\phi}(Q)$.

It remains to show that if $F' \subseteq F$ is a subfield containing R, then $\overline{\phi}(Q) \subseteq F'$, but this is clear from the fact that elements of $\overline{\phi}(Q)$ have the form $\phi(r)\phi(d)^{-1} = rd^{-1}$ for $r, d \in R$, $d \neq 0$. □

Exercise. Let $R = R_1 \times R_2$ be a product of two commutative rings with 1. Let $e_1 = (1, 0), e_2 = (0, 1) \in R$. Show that

$$e_1^{-1}R \approx R_1, \qquad e_2^{-1}R \approx R_2.$$

23. Prime fields

Let R be a ring with identity. Then the map $\phi\colon \mathbb{Z} \to R$ defined by $\phi(1 + \cdots + 1) = 1 + \cdots + 1$ is a ring homomorphism, which preserves 1. The kernel of ϕ is an ideal of $\mathbb{Z}$. I usually won't notate this map: in a ring R, the symbol "n" means "$1 + \cdots + 1$", but note that it could be the case that $n = 0$ in R, even if $n \neq 0$ in $\mathbb{Z}$.

If F is a field, then the kernel of $\mathbb{Z} \to F$ must be a prime ideal. Let (p) be the kernel, where p is either a prime integer or 0. Then $\mathbb{Z}/(p) \approx \phi(\mathbb{Z}) \subseteq F$, is a subdomain of F.

- If $p > 0$, then $\mathbb{Z}/(p)$ is a finite set, and is thus a field, denoted $\mathbb{F}_p$.
- If $p = 0$, then $\mathbb{Z}/(0) \approx \mathbb{Z}$, and therefore $\mathbb{Q} \approx \operatorname{Frac}(\mathbb{Z}) \subseteq F$.

The fields $\mathbb{F}_p$ for p prime and $\mathbb{Q}$ are called **prime fields**. Every field contains a unique prime subfield.

prime fields

The **characteristic** of F is the non-negative integer p such that $\operatorname{Ker}(\mathbb{Z} \to F) = (p)$. So any field containing a copy of $\mathbb{F}_p$ has "characteristic p", while any field containing a copy of $\mathbb{Q}$ has "charactersitic 0".

characteristic

24. Chinese Remainder Theorem

Here rings are commutative with 1.

Recall that if $R_1, \dots, R_n$ are rings, we can define the product ring $R_1 \times \cdots \times R_n$.

If $A, B \subseteq R$ are ideals, then there is a homomorphism (preserving 1),

$$\phi\colon R \to R/A \times R/B, \qquad \phi(r) = (r + A, r + B).$$

The kernel of ϕ is $A \cap B$.

Question: when is this an isomorphism?

If $A, B \subseteq R$ are ideals, let

$$A + B := \{\, a + b \mid a \in A,\ b \in B \,\} \subseteq R$$

and

$$AB := \{\, a_1b_1 + \cdots + a_kb_k \mid a_i \in A,\ b_i \in B,\ k \geq 0 \,\}.$$

Then both $A + B$ and AB are ideals in R (exercise).

Exercise. If $A = (a_1, \dots, a_m)$ and $B = (b_1, \dots, b_n)$, then $A + B = (a_1, \dots, a_m, b_1, \dots, b_n)$ and (if R is commutative $AB = (a_1b_1, \dots, a_mb_n) = (a_ib_j,\ 1 \leq i \leq m,\ 1 \leq j \leq n)$.

Say that ideals A and B are **comaximal** if $A + B = R$. (Some people call this "coprime", but DFs term is better.) Since the unit ideal is the only one which contains 1, we have that A and B are comaximal iff $1 = a + b$ for some $a \in A$ and $b \in B$. Note that comaximal does not imply maximal.

comaximal

Example. $(a), (b) \subseteq \mathbb{Z}$ are comaximal if and only if a, b are relatively prime, in the usual sense (no common divisors greater than 1).

To see this, observe that $(a) + (b) = \mathbb{Z}$ if and only if there exist $m, n \in \mathbb{Z}$ such that $ma + nb = 1$. It is standard that a, b are relatively prime iff there exist $m, n \in \mathbb{Z}$ such that $1 = ma + nb$ (proof given below).

Proof. Note that since ideals in $\mathbb{Z}$ are additive subgroups, all ideals in $\mathbb{Z}$ are principal. Thus $(a) + (b) = (c)$ for some $c \geq 0$.

Since $a, b \in (c)$, we see that c is a common divisor of a, b. If d is any common divisor of a, b, we must have $d \mid c$ as well since $c = ma + nb$ for some $m, n \in \mathbb{Z}$. Thus $c = \gcd(a, b)$ (or is 0 if $a = b = 0$). So $(a) + (b) = \mathbb{Z}$ iff a, b are relatively prime.

Proposition. *If A, B are comaximal ideals in R, then $AB = A \cap B$, and ϕ induces an isomorphism $R/(AB) \to R/A \times R/B$.*

Proof. In general, $AB \subseteq A \cap B$. Conversely, if $x \in A \cap B$, then using $1 = a + b$, we have $x = ax + xb \in AB$, whence $AB = A \cap B$.

If $(\overline{r}_1, \overline{r}_2) \in R/A \times R/B$, let $r_1, r_2 \in R$ such that $r_k \in \overline{r}_k$. Then using $1 = a + b$ consider $r = r_2a + r_1b$. The image of r in R/A is equal to that of r_1, and the image in R/B is equal to that of r_2. Thus $\phi\colon R \to R/A \times R/B$ is surjective, and the kernel is clearly $A \cap B$, which we have shown is the same as AB. □

Exercise. This theorem has a converse. If R is a commutative ring, and we have a ring isomorphism $R \approx S \times T$, show that there are ideals $A, B \subseteq R$ which are comaximal, such that $R/A \approx S$ and $R/B \approx T$.

We can generalize:

Proposition. *If $A_1, \dots, A_n$ are pairwise comaximal, then $A_1 \cdots A_n = A_1 \cap \cdots \cap A_n$, and $R/A_1 \cdots A_n \to (R/A_1) \times \cdots \times (R/A_n)$ is an isomorphism.*

Proof. We prove this by induction on n, noting that the cases of $n = 1, 2$ are already done.

Let $A = A_1$ and $B = A_2 \cdots A_n$. I claim that $A + B = R$. Since $A_1 + A_k = R$ for $k = 2, \dots, n$, there exist $x_k \in A_1$ and $a_k \in A_k$ such that $x_k + a_k = 1$. Thus

$$1 = (a_2 + x_2)(a_3 + x_3) \cdots (a_k + x_k) = (a_2 \cdots a_k) + (a_2 \cdots a_k) + x_2(\text{stuff}) + \cdots + x_k(\text{stuff}) \in B + A.$$

Thus, we have an isomorphism of rings $R/AB \approx R/A \times R/B$, and thus

$$R/A_1 \cdots A_n \approx R/A_1 \times R/A_2 \cdots A_n \approx R/A_1 \times R/A_2 \cdots A_n \approx R/A_1 \times R/A_2 \times \cdots \times R/A_n$$

by induction. □

Thus, if $a_1, \dots, a_n$ are integers which are pairwise relatively prime, and $a = a_1 \cdots a_n$, then $\mathbb{Z}/(a) \approx \mathbb{Z}/(a_1) \times \cdots \times \mathbb{Z}/(a_n)$. In particular, if $n = p_1^{e_1} \cdots p_r^{e_r}$ is a factorization into distinct primes, we have

$$\mathbb{Z}/(n) \approx \mathbb{Z}/(p^{e_1}) \times \cdots \times \mathbb{Z}/(p^{e_r}).$$

Since this is also an isomorphism of additive groups, this gives the primary decomposition of finite cyclic groups which we described earlier.

25. Euclidean domains

Many of the most familiar integral domains have a "Euclidean algorithm" of some sort. Such domains are called *Euclidean domains.*

A **Euclidean domain** is an integral domain R such that there exists a function $N \colon R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ such that **Euclidean domain**

- for any $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that

$$a = qb + r \quad \text{with either } r = 0 \text{ or } N(r) < N(b).$$

Note: This is not exactly the definition given in DF: they extend N to all of R by setting $N(0) = 0$. However, the definition I have given is very common (it is the one on wikipedia, for instance). These variations don't make any essential difference in how this notion is used.

Note: the function N is not unique: there can be many such functions for any Euclidean domain.

Example. Here are some standard examples of Euclidean domains.

- Any field with $N(a) = 1$ for all $a \neq 0$.
- $\mathbb{Z}$ with $N(a) = |a|$.
- $F[x]$ with $N(f) = \deg(f)$.
  (This example explains why we don't always want to define $N(0)$.)

Exercise. A pair (R, N) consisting of an integral domain R and a function $N \to R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ is a Euclidean domain iff for every non-trivial principal ideal $0 \neq (b) \subseteq R$, every non-trivial coset of (b) contains a representative r such that $N(r) < N(b)$.

Example (Gaussian integers are a Euclidean domain). Let $\mathcal{O} = Z[i] \subseteq \mathbb{C}$, the ring of Gaussian integers. If you picture $\mathbb{C}$ as the plane, then the subset $\mathbb{Z}[i]$ is the "integer lattice" in the plane.

Define

$$N(a + bi) := |a + bi|^2 = (a + bi)(a - bi) = a^2 + b^2.$$

Note that $N(\alpha\beta) = N(\alpha)N(\beta)$.

Recall that $\mathrm{Frac}(\mathbb{Z}[i]) = \mathbb{Q}(i) \subseteq \mathbb{C}$. Given $\alpha = a + bi$ and $\beta = c + di$ in $\mathbb{Z}[i]$, we have

$$\frac{\alpha}{\beta} = r + si = \frac{ac - bd}{c^2 + d^2} + \frac{ad + bc}{c^2 + d^2} i \in \mathbb{Q}(i) \subseteq \mathbb{C}, \qquad r = \frac{ac - bd}{c^2 + d^2},\ s = \frac{ad + bc}{c^2 + d^2} \in \mathbb{Q}.$$

This element α/β is, at most, a distance of $1/\sqrt{2}$ from some element of $\mathbb{Z}[i]$, which are exactly the points in the integer lattice inside $\mathbb{C}$.

In fact, choose $p, q \in \mathbb{Z}$ such that $|r - p|, |s - q| \leq 1/2$. Then

$$|(\alpha/\beta) - (p + qi)|^2 = |r - p|^2 + |s - q|^2 \leq 1/2,$$

and thus

$$|\alpha - (p+qi)\beta|^2 \le |\beta|^2/2.$$

Therefore, setting $\gamma = \alpha - (p+qi)\beta \in \mathbb{Z}[i]$, we have

$$\alpha = (p+qi)\beta + \gamma, \qquad |\gamma|^2 < |\beta|^2.$$

Thus $\mathbb{Z}[i]$ is a Euclidean domain.

On the other hand, not every quadratic integer ring is a Euclidean domain, e.g., $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}]$, as we will see. I'll prove this indirectly, but note that the argument I gave for $\mathbb{Z}[i]$ doesn't work here: an element in $\mathbb{C}$ can as far as $\sqrt{3/2} > 1$ from any element of $\mathbb{Z}[\sqrt{-5}]$.

26. Principal ideal domains

A **principal ideal domain** (or just **PID**) is a domain in which every ideal is principal. As we have already seen, $\mathbb{Z}$ is a PID, since ideals are additive subgroups, and all additive subgroups of $\mathbb{Z}$ are cyclic groups.

principal ideal domain
PID

Proposition. *Every Euclidean domain is a PID.*

Proof. Let I be a non-0 ideal in R. Let $d \in I$ be a non-zero element of minimal norm. We claim that $I = (d)$. Clearly $(d) \subseteq I$.

If $a \in I$, the Euclidean property gives $a = qd + r$ with either $r = 0$ or $N(r) < N(d)$. If $r = 0$, then $a = qd \in (ad)$. If $N(r) < N(d)$, then $r = a - qd \in I$, contadicting the minimality of $N(a)$. □

Thus, we get some examples of PIDs: Any field, $\mathbb{Z}$, $F[x]$ for F a field, and $\mathbb{Z}[i]$.

Remark. There exist PIDs which are not Euclidean domains. For instance, $\mathcal{O}_{\mathbb{Q}(\sqrt{-19})}$ (see Example at end of DF§8.2).

In an integral domain R, a **greatest common divisor (gcd)** of $a, b \in R$ with $b \neq 0$ is any element $d \in R$ such that (i) d divides both a and b, and (ii) if e divides both a and b, then $e|d$.

greatest common divisor (gcd)

Note: Defined this way, the gcd is only unique up-to-units. For instance, in $\mathbb{Z}$ both 3 and -3 are greatest common divisors of the pair $6, 15$.

Note: a in general a gcd might not exist. We will give an example below.

Proposition. *In a PID, the greatest common divisor d of a and b always exists, and $(d) = (a,b) = (a) + (b)$.*

Proof. In a PID, (a,b) is a principal ideal, so there is d such that $(a,b) = (d)$. Clearly $a, b \in (d)$, so d is a common divisor, while if $a, b \in (e)$ for some e, then $(d) = (a,b) \subseteq (e)$. □

Note: If the PID is actually a Euclidean domain, then the Euclidean algorithm can be used to compute gcds.

Proposition. *In a PID, every non-zero prime ideal is a maximal ideal.*

Proof. Let $(0) \neq (p) \subsetneq R$ be a prime ideal. I will show that if $(p) \subseteq (a) \subseteq R$ for some $a \in R$, then either $(a) = (p)$ or $(a) = R$.

If $(p) \subseteq (a)$ then $p = ab$ for some $b \in R$. If $a \in (p)$ then $(a) = (p)$ so we are done, so suppose $a \notin (p)$. Then since (p) is a prime ideal, $ab = p \in (p)$ and $a \notin (p)$ imply $b \in (p)$, so $b = cp$ for some $c \in R$. Thus

$$p = ab = acp \quad \Rightarrow \quad 1 = ac,$$

so $1 = ac \in (a)$ whence $(a) = R$. □

27. A quadratic integer ring which is not a PID

Most integral domains are *not* PIDs. For instance, neither $\mathbb{Z}[x]$ or $F[x,y]$ ($F =$ field) are PIDs, **F 7 Oct** as we have seen.

In fact, not all quadratic integer rings are PIDs.

Example ($\mathcal{O}_{\mathbb{Q}(\sqrt{-5})}$ is not a PID). Let $\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}] = \{\, a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\,\} \subset \mathbb{C}$. Let $I = (3, 2+\sqrt{-5})$. I claim that I is not principal, so $\mathcal{O}$ is not a PID (and therefore not a Euclidean domain). To do this, I'm going to need the function

$$N\colon \mathcal{O} \to \mathbb{Z}_{\geq 0}, \qquad N(a+b\sqrt{-5}) := (a+b\sqrt{-5})(a-b\sqrt{-5}) = a^2+5b^2, \qquad a,b\in\mathbb{Z},$$

and the properties:

$$N(\alpha) = 0 \iff \alpha = 0, \qquad N(\alpha\beta) = N(\alpha)N(\beta).$$

(Note: this will not be a Euclidean function.)

Claim. $I \neq \mathcal{O}$. If $1 \in I = (3,\, 2+\sqrt{-5})$, then there exist $\alpha, \beta \in \mathcal{O}$ such that

$$1 = 3\alpha + (2+\sqrt{-5})\beta.$$

Note that $(2+\sqrt{-5})(2-\sqrt{-5}) = 9$. Thus multiplying the above by $2-\sqrt{-5}$ gives

$$2-\sqrt{-5} = 3(2-\sqrt{-5})\alpha + 9\beta = (2-\sqrt{-5})\alpha + 3\beta),$$

and thus $2-\sqrt{-5} \in (3)$. But this is not possible, because principal ideals in $\mathcal{O}$ which are generated by *integers* have a very simple form: $(3) = \{\, 3a + 3b\sqrt{-5} \mid a,b\in\mathbb{Z}\,\}$.

Claim. $N(\alpha) = 1$ iff $\alpha = \pm 1$. This is because the only integer solutions for $a^2+5b^2 = 1$ are $a = \pm 1$, $b = 0$. This means that $\mathcal{O}^\times = \{\pm 1\}$.

Now suppose I is a principal ideal, i.e., suppose $I = (3, 2+\sqrt{-5}) = (a+b\sqrt{-5})$, for some $a, b\in\mathbb{Z}$. Then $3 = (a+b\sqrt{-5})\alpha$ and $2+\sqrt{-5} = (a+b\sqrt{-5})\beta$ for some $\alpha,\beta\in R$. Thus,

$$9 = N(3) = N(a+b\sqrt{-5})N(\alpha) = (a^2+5b^2)N(\alpha),$$
$$9 = N(2+\sqrt{-5}) = N(a+b\sqrt{-5})N(\beta) = (a^2+5b^2)N(\beta).$$

We have $a,b\in\mathbb{Z}$ and a^2+5b^2 divides 9, and the only possibilities are $a^2+5b^2\in\{1,3,9\}$.

- If $a^2+5b^2 = 1$, the only integer solutions are $a = \pm 1$ and $b = 0$, and so $I = R$, which we have already disallowed.
- There are no integer solutions to $a^2+5b^2 = 3$.
- If $a^2+5b^2 = 9$, then $N(\alpha) = N(\beta) = 1$, whence $\alpha,\beta\in\{\pm1\}$. Thus $a+b\sqrt{-5} =$ both ± 3 and $\pm(2+\sqrt{-5})$, which is not possible.

Therefore, I is not principal.

Example (Elements without a gcd). Let $\mathcal{O} = \mathbb{Z}[\sqrt{-5}]$, and consider the elements

$$6, \qquad 2+2\sqrt{-5}.$$

I claim these have no gcd in $\mathcal{O}$. To show this I'll just find *all* the common divisors.

If $a+b\sqrt{-5}$ is any common divisor, then there are $\alpha,\beta\in\mathcal{O}$ such that

$$6 = \alpha(a+b\sqrt{-5}), \qquad 2+2\sqrt{-5} = \beta(a+b\sqrt{-5}).$$

Taking norm gives

$$36 = N(\alpha)(a^2+5b^2), \qquad 24 = N(\beta)(a^2+5b^2).$$

Therefore a^2+5b^2 is a common divisor of $24, 36$ in $\mathbb{Z}$, i.e.,

$$a^2+5b^2 \in \{1,2,3,4,6,12\}.$$

We can easily find all possible solutions with $a,b\in\mathbb{Z}$, which turn out to all be common divisors:

$$\pm 1, \qquad \pm 2, \qquad \pm 2 \pm 2\sqrt{-5}.$$

None of these can be the gcd, because none of these elements is divisible in $\mathcal{O}$ by both 2 and $2+\sqrt{-5}$.

28. Irreducible elements

Let R be a domain. We can classify elements of R into exactly one of the following types.

- *Zero.* Just 0.
- *Units.* Elements which have a multiplicative inverse.
- *Reducible elements.* $r \in R$ which is not 0 or a unit, such that $r = ab$ for some a, b which are not 0 or units.
- *Irreducible elements.* $r \in R$ which are not 0 or a unit or reducible.

We say $a, b \in R$ are **associate** (or **same up to units** if there exists a unit $u \in R^\times$ such that $b = ua$. Being associate is an equivalence relation on R (exercise).

associate
same up to units

We say that $a \mid b$ iff $(a) \subseteq (b)$. Equivalently, iff there is $c \in R$ such that $b = ac$.

Proposition. *Let $a, b \in R$ a domain. TFAE.*

1. *a and b are associate.*
2. *$a \mid b$ and $b \mid a$.*
3. *$(a) = (b)$.*

Proof. Straightforward. □

So when we talk about elements up to units, we are really talking about principal ideals.

Example. For $R = F$ a field, we have

- 0.
- $F^\times = F \setminus \{0\}$.
- Irreducible elements = none.
- Reducible elements = none.

Example. For $R = \mathbb{Z}$, we have

- 0.
- $\mathbb{Z}^\times = \{\pm 1\}$.
- Irreducible elements = $\{\, \pm p \mid p \in \mathbb{N} \textit{ is a prime number} \,\}$.
- Reducible elements = composite integers (positive and negative).

Example. For $R = F[x]$ with F a field, we have

- 0.
- $(F[x])^\times = F^\times$, the non-zero constant polynomials.
- Irreducible elements = *irreducible polynomials*, i.e., polynomials f which cannot be written as a product of polynomials of strictly smaller degree.
- Reducible elements = polynomials of degree ≥ 1 which do factor as a product of polynomials of strictly smaller degree.

Example. For $R = \mathbb{C}[x]$, we have

- 0.
- $(\mathbb{C}[x])^\times = \mathbb{C}^\times$.
- Irreducible elements = $\{\, ax + b \mid a, b \in \mathbb{C},\ a \neq 0 \,\}$.
- Reducible elements =everything else.

Example. For $R = \mathbb{R}[x]$, we have

- 0.
- $(\mathbb{R}[x])^\times = \mathbb{R}^\times$.

- Irreducible elements $= \{\, ax+b \mid a,b\in\mathbb{R},\ a\neq 0\,\}\cup\{\, ax^2+bx+c \mid a,b,c\in\mathbb{R},\ a\neq 0,\ b^2<4ac\,\}$.
- Reducible elements = everything else.

Sketch proof: use that (i) $f\in\mathbb{R}[x]\smallsetminus\mathbb{R}$ factors as a product of degree 1 factors in $\mathbb{C}[x]$ ("fundamental theorem of algebra"), (ii) all non-real complex roots of a real polynomial come in conjugate pairs, so if $\lambda\in\mathbb{C}\smallsetminus\mathbb{R}$ is a root of $f\in\mathbb{R}[x]$, then

$$g=(x-\lambda)(x-\overline{\lambda})=x^2-2\operatorname{Re}(\lambda)x+|\lambda|^2\in\mathbb{R}[x]$$

divides f, so $f=gh$ for some $h\in\mathbb{R}[x]$ with $\deg h<\deg f$.

Gaussian integers.

The notion of irreducibility is related to (but not the same as) that of maximal ideal.

Lemma. *Let $p\in R$ which is not zero and not a unit. Then $p\in R$ is irreducible iff for all $a\in R$, $(p)\subsetneq(a)$ implies $(a)=R$. That is, p is irreducible iff (p) is maximal among proper* principal *ideals.*

Proof. If p is irreducible and $(p)\subsetneq(a)$, then $p=ab$ for some $b\in R$ not a unit. Thus a is a unit since p is irreducible, so $(a)=R$. On the other hand, if p is a unit then $(p)=R$, and if p is reducible, then $p=ab$ for non-units a,b whence $(p)\subsetneq(a)\subsetneq R$. □

Thus if R is a PID, $p\in R$ is irreducible iff $p\neq 0$ and (p) is a maximal ideal.

Corollary. *If p,q are irreducible elements in a domain, then $p|q$ iff p and q are the same up to units.*

Proof. It is clear that associate elements divide each other. Conversely, if $q\in(p)$, then $(q)\subseteq(p)\subsetneq R$; since q is irreducible, we can only have $(q)=(p)$. □

29. Prime elements

In a domain R, an element $p\in R$ is **prime** iff $p\neq 0$ and (p) is a prime ideal. That is, iff p is non-zero and not a unit, and if $p\mid ab$ implies either $p\mid a$ or $p\mid b$. **prime element**

Remark. If $p\in R$ with $p\neq 0$, then p is prime iff $R/(p)$ is an integral domain.

Proposition. *In any domain, prime elements are irreducible.*

Proof. Let p be prime, and suppose $(p)\subsetneq(a)$ for some a. Then $a\notin(p)$, while $p=ab$ for some $b\in R$. Since p is prime, $b\in(p)$, so $b=pr$ for some $r\in R$, whence

$$p=ab=apr \qquad\Longrightarrow\qquad 1=ar,$$

whence a is a unit. □

It is not the case that irreducibles are always prime.

Example. In $R=\mathbb{Z}[\sqrt{-5}]$, the element 3 is irreducible but not prime. *Irreducibility:* if $3=\alpha\beta$, then $9=N(\alpha)N(\beta)$, whence either α or β is a unit, since there is no element of norm 3. *Not prime:* because $(2+\sqrt{-5})(2-\sqrt{-5})=3^2\in(3)$.

Remark. Recall that a *prime number* is a positive integer p such that $p\neq 1$ and the only positive divisors of p are 1 and p. This is basically the same as saying p is a positive *irreducible* element. It is not the same statement as saying p is a *prime element*, in the above sense.

This can be a real point of confusion.

It turns out that the primes (in the above sense) in $\mathbb{Z}$ are the same as the irreducibles: If $p\in\mathbb{Z}$ is irreducible, its only factors are $\{\pm1,\pm p\}$, so if $p\mid ab$ but $p\nmid a$, then $\gcd(p,a)=1$, so there are $m,n\in\mathbb{Z}$ such that $1=mp+na$, so $b=mpb+n(ab)$ is divisible by p.

In fact, this proof works in any PID.

Proposition. *In a PID, prime and irreducible are equivalent.*

Proof. I have shown that in general prime implies irreducible, so I only need to show that irreducibles are prime in a PID. We know that a non-zero p is irreducible if and only if (p) is maximal among proper *principal* ideals. In a PID, all ideals are principal, so this says (p) is a maximal ideal, and therefore is also a prime ideal, so p is a prime element.

Here is a more elementary proof that irreducibles are prime in a PID R. Let p be irreducible, and suppose $ab \in (p)$. I need to show that $a \notin (p)$ implies $b \in (p)$. The ideal (p, a) must be principal (since we are in a PID), so $(p, a) = (r)$ for some r. Then $p = rs$ for some s. If s is a unit, then $(p, a) = (r) = (p)$, and thus $a \in (p)$, contradicting the hypothesis. Since p is irreducible, it follows that since s is not a unit, r is a unit. Thus $(p, a) = R$, whence

$$1 = pu + av$$

for some $u, v \in R$. Multiplying by b gives

$$b = bpu + bav \in (p)$$

since $ab \in (p)$. □

30. Unique factorization domains

A **Unique factorization domain** (UFD) is a domain such that every non-zero non-unit $r \in R$ satisfies

Unique factorization domain

(1) $r = p_1 \cdots p_n$ for some irreducibles $p_1, \dots, p_n \in R$, $n \geq 1$, and
(2) this decomposition is unique up to associates, i.e., if $r = p_1 \cdots p_n = q_1 \cdots q_m$ for irreducibles p_i, q_j, then $m = n$ and there is a permutation $\sigma \in S_n$ such that $q_k = u_k p_{\sigma(k)}$ for some unit u_k.

Note: in a UFD, prime and irreducible are equivalent. Prime always implies irreducible, whereas if p is irreducible and $ab \in (p)$, then $ab = pr$, so p must appear (up to associate) in the irreducible factorization of either a or b, whence either a or b is in (p).

We will soon prove that any PID is a UFD, which when applied to $\mathbb{Z}$ proves the "Fundamental theorem of arithmetic". On the other hand, $\mathbb{Z}[\sqrt{-5}]$ is not a UFD, since $3^2 = (2 + \sqrt{-5})(2 - \sqrt{-5})$.

Theorem. *Every PID is a UFD.*

31. PIDs are UFDs: Existence of factorizations

I'm going to do this as a consequence of a more general fact.

M 10 Oct

Let R be an integral domain. Say that R has the **ascending chain condition (acc) for principal ideals** if for any collection $\{I_k\}_{k \in \mathbb{Z}_{>0}}$ of principal ideals in R such that

ascending chain condition (acc) for principal ideals

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

there exists n such that $I_k = I_n$ for all $n \geq k$.

Lemma. *Every PID has the acc for principal ideals.*

Proof. Let $\{I_k\}_{k \geq 1}$ be a countable chain of principal ideals, so $I = (a_k)$ for some $a_k \in R$. Let $J := \bigcup_{k=1}^{\infty} I_k$. Then J is an ideal. (The things you have to check involve only a finite number of elements of J, and these will always be in some I_k.)

Since R is a PID, then $J = (b)$ for some b. But then $b \in I_n$ for some n, whence $J = I_n$ and so $I_k = I_n$ if $k \geq n$. □

Warning: in general, a union of ideals is not an ideal (e.g., $2\mathbb{Z} \cup 3\mathbb{Z}$ is not an ideal). A union of a *chain* of ideals is an ideal however.

Proposition. *Let R be a integral domain. If R has the acc for principal ideals, then every non-zero non-unit in R is equal to a finite product of irreducible elements.*

Proof of existence of irreducible factorization. Let R be an integral domain with the acc for principal ideals. Say $a \in R$ is *bad* if it is a non-zero non-unit which cannot be writen as a product of finitely many irreducibles. We want to show R has no bad elements. Clearly, an irreducible element is not bad, since it is a product of one irreducible.

Note that if a, b are not bad, then ab is also not bad. Thus, if a is bad, it must be reducible and for any factorization $a = bc$ at least one of b or c is bad. Thus, for bad a we can always write $a = a'b$ with a' also bad and b a non-unit.

Using this observation, we can produce an infinite chain of factorizations of a bad element a, of the form

$$a = a_1b_1 = a_2b_2b_1 = a_3b_3b_2b_1 = \cdots ,$$

where in each step we factor the bad element a_{k-1} as $a_{k-1} = a_kb_k$ with a_k bad and b_k a non-unit. This gives us a strictly ascending chain of principal ideals.

$$(a) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots .$$

But this is impossible, since R has the acc for principal ideals. □

32. PIDs are UFDs: Uniqueness of factorizations

Proposition. *Let R be a integral domain. If all irreducible elements in R are prime elements, then factorization in irreducibles (when it exists) is unique up to units and reordering.*

The proof of this part is exactly the same as the proof in the case $R = \mathbb{Z}$, which you have probably seen before.

Proof. If a non-zero non-unit r has an irreducible factorization, it has one of shortest length, say $r = p_1 \cdots p_n$. We will show that this is the only irreducible factorization up to equivalence, by induction on n.

If $n = 1$, and $r = p_1 = q_1 \cdots q_m$ are two irreducible factorizations, then $q_1 \cdots q_m \in (p_1)$; since p_1 is irreducible, and so prime by hypothesis, some $q_k \in (p_1)$, whence p_1 and q_k are associate, and thus $q_1 \cdots \widehat{q_k} \cdots q_m$ is a unit, so this cannot happen unless $m = 1$.

More generally, if $r = p_1 \cdots p_n = q_1 \cdots q_m$ are irreducible factorizations with $n \leq m$, then since $q_1 \cdots q_m \in (p_n)$, we have that p_n is be associate to some q_k, whence $p_1 \cdots p_{n-1} = uq_1 \cdots \widehat{q_k} \ldots q_m$ for some unit u. By induction, $n - 1 = m - 1$, and the remaining q_is are associate to the remaining p_js. □

33. Factorization in the Gaussian integers

Let $\mathcal{O} = \mathbb{Z}[i]$. We are going to identify all the irreducible elements in $\mathcal{O}$, up to associates. By what we have proved, this is a PID, and thus a UFD.

Recall the norm function $N \colon \mathcal{O} \to \mathbb{Z}$ defined by $N(a+bi) = a^2 + b^2$ if $a, b \in \mathbb{Z}$. The norm function actually takes only non-negative values, and is multiplicative: $N(\alpha\beta) = N(\alpha)N(\beta)$.

We have already shown that $\alpha \in \mathcal{O}$ is a unit iff $N(\alpha) = 1$. From this it is easy to see that

$$\mathcal{O}^\times = \{\pm 1, \pm i\},$$

by solving the equation $a^2 + b^2 = 1$ in $\mathbb{Z}$.

Next we want to determine the irreducible=prime elements of $\mathcal{O}$, up to associates.

Here is a criterion for finding some irreducible elements in $\mathcal{O}$.

Lemma. *Let $\alpha \in \mathcal{O}$. If $N(\alpha) \in \mathbb{Z}$ is a prime number, then α is irreducible in $\mathcal{O}$.*

Proof. If $\alpha = \beta_1\beta_2$ for some $\beta_,\beta_2 \in \mathcal{O}$, then $N(\alpha) = N(\beta_1)N(\beta_2)$ with $N(\beta_1), N(\beta_2) \in \mathbb{Z}$. If $N(\alpha)$ is prime, then one of $N(\beta_1), N(\beta_2)$ is 1, so one of β_1, β_2 is a unit in $\mathcal{O}$. □

E.g., $N(2 + i) = N(2 - i) = 5$, so $2 + i$ and $2 - i$ are primes in $\mathcal{O}$.

Note that $\mathbb{Z}$ is a subring of $\mathcal{O}$ (with 1).

Proposition. *If R is a commutative ring with 1, and $S \subseteq R$ is a subring (with 1), and $P \subseteq R$ is a prime ideal of R, then $S \cap P$ is a prime ideal of S.*

Proof. If $a, b \in S$ such that $ab \in S \cap P$, then $ab \in P$, so either $a \in S \cap P$ or $b \in S \cap P$ since P is a prime ideal. □

Note: in this case we say that the prime ideal P in R "lies over" the prime ideal $S \cap P$ in S.

As a conseqquence, an irreducible element in $\mathcal{O}$ divides *exactly one* "rational prime" in $\mathbb{Z}$.

Proposition. *Let $p \in \mathbb{Z}$ be a prime number, and let $\alpha \in \mathcal{O}$ be an irreducible element. TFAE.*

1. *α is an irreducible divisor of p in $\mathcal{O}$.*
2. *$p\mathbb{Z} = (\alpha) \cap \mathbb{Z}$, i.e., (α) lies over $p\mathbb{Z}$.*

Proof. Since α is irreducible in $\mathcal{O}$, the ideal (α) is maximal among proper principal ideals in $\mathcal{O}$. Since $\mathcal{O}$ is a PID, this means that (α) is a maximal ideal, and so prime.

Therefore $(\alpha) \cap \mathbb{Z} = q\mathbb{Z}$ for some prime number q.

(1) $\implies$ (2). If α divides p, then $p \in (\alpha) \cap \mathbb{Z} = q\mathbb{Z}$ and thus $p = q$.

(2) $\implies$ (1). If $q = p$, then $p \in (\alpha)$ and thus $p = \alpha\beta$ for some $\beta \in \mathcal{O}$, i.e., α is a divisor of p. □

Now we classify irreducibles in $\mathcal{O}$ which lie over a given rational prime p. Suppose $\alpha \in \mathcal{O}$ is irreducible with $(\alpha) \cap \mathbb{Z} = p\mathbb{Z}$ with p a prime integer. We have $p = \alpha\beta$ for some $\beta \in \mathcal{O}$. Taking norms gives

$$p^2 = N(p) = N(\alpha)N(\beta).$$

Since α is not a unit, there are two cases:

- $N(\alpha) = p^2$, $N(\beta) = 1$, whence β is a unit and thus p and α are associate, so $\alpha \in \{\pm p, \pm pi\}$.
- $N(\alpha) = p$, $N(\beta) = p$, so both α and β are irreducible, and $p = \alpha\beta$ is an irreducible factorization of p. Thus these are the only two irreducible divisors up to associates, by uniqueness of irreducible factorization.

Conclusion: If p is a prime number, then an element $\alpha = a + bi \in \mathcal{O}$, $a, b \in \mathbb{Z}$, is an irreducible divisor of p iff one of the following mutually exclusive cases occurs:

1. $\alpha = \pm p$ or $\alpha = \pm pi$, or
2. $a^2 + b^2 = p$.

In case (1) there is only one irreducible divisor, up to associates. In case (2) there are at most two irreducible divisors up to associates (depending on whether $\alpha = a + bi$ and $\beta = a - bi$ are associtate).

In particular, understanding how p factors in $\mathcal{O}$ amounts to knowing whether the equation $a^2 + b^2 = 1$ has solutions in $\mathbb{Z}$. If it does not, we are in case (1), while if it does we are in case (2).

Examples:

- Up to associates, $1 + i$ is the only prime over 2, since $1 - i = -i(1 + i)$.
- 3 is the only prime over 3 (since $N(a + bi) = a^2 + b^2 = 3$ has no integer solutions).
- $2 + i$ and $2 - i$ are non-associate primes which lie over 5.

Note that $\alpha = a + bi$ and $\overline{\alpha} = a - bi$ are associate iff (a) $b = 0$, (b) $a = 0$, (c) $a = b$, or (d) $a = -b$. If α is also irreducbile, this can only happen when $a, b \in \{\pm 1\}$. Thus 2 is the *only* prime number with two irreducible factors which are associate.

34. Fermat's theorem on sums of squares

Lemma (Lagrange). *Let p be a prime number of the form $p = 4m + 1$, with $m \in \mathbb{Z}$. There exists $n \in \mathbb{Z}$ such that $p \mid (n^2 + 1)$.* **W 12 Oct**

Proof. This is really a statement about the field $\mathbb{Z}/p$: if p is a prime congruent to 1 mod p, then $\mathbb{Z}/p$ contains a square root of -1. This is left as a (non-obvious) exercise. (It will also be proved later on.) □

Theorem (Fermat)**.** *A rational prime p is a sum of two squares iff $p = 2$ or $p \equiv 1 \mod 4$.*

Proof. The case of $p = 2$ is clear, so suppose p is odd.

If an odd p is a sum of squares, it is easy to show that $p \equiv 1 \mod 4$, because squares of integers are congruent to either 0 or 1 mod 4.

Conversely, if $p \equiv 1 \mod 4$, we have shown $p \mid n^2 + 1$ for some n. Thus, in the Gaussian integers $\mathcal{O}$ we have $p \mid (n + i)(n - i)$. If p is a prime element of $\mathcal{O}$, then p must divide either $n + i$ or $n - i$ in $\mathcal{O}$, but this is clearly not possible, since $p(a + bi) = (pa) + (pb)i$. Thus p is not prime in $\mathcal{O}$, and therefore p is reducible in $\mathcal{O}$. (This is because irreducible elements must be prime in a PID like $\mathcal{O}$.) The claim follows, since we have shown that if p is reducible in $\mathcal{O}$ then $p = (a+bi)(a-bi) = a^2 + b^2$ for some $a, b \in \mathbb{Z}$. □

This gives an essentially complete understanding of irreducible elements in $\mathbb{Z}[i]$.

Corollary. *A positive integer n has the form $n = a^2 + b^2$ for some $a, b \in \mathbb{Z}$ iff its prime factorization (in $\mathbb{Z}$) $n = p_1^{k_1} \cdots p_r^{k_r}$, (primes p_i pairwise distinct) is such that: if $p_i \equiv -1 \pmod 4$, then k_i is even.*

Proof. It is immediate that the the set of S integers which are a sum of two squares is precisely the image of the norm function $N \colon \mathbb{Z}[i] \smallsetminus \{0\} \to \mathbb{Z}$. Since N is multiplicative, the subset S is multiplicatively closed.

Let $T \subseteq \mathbb{Z}$ be the subset described in the statement of the Corollary. Note that T is also multiplicatively closed. We want to show $S = T$.

For any prime p we have $p^2 \in S$, and by Fermat's theorem we have $p \in S$ if $p \not\equiv -1 \pmod 4$. Since S is multiplicatively closed we see that $T \subseteq S$.

We show $m = a^2 + b^2 \in S$ implies $m \in T$ by induction on m. The statement is clearly true for $m = 1, 2$, which are both in S and T. Given $m \in S$ with $m \geq 3$, either

- m has no prime factors in $\mathbb{Z}$ which are $\equiv -1 \pmod 4$, whence $m \in T$, or
- some prime $p \mid m$ with $p \equiv -1 \pmod 4$. In this case we will show that $p^2 \mid m$, whence $m' = m/p^2 \in S$ and thus $m' \in T$ by induction, and therefore $m = m'p^2 \in T$ since $p^2 \in T$.

Since $m = a^2 + b^2$, we have that

$$(a + bi)(a - bi) = p\beta \qquad \text{for some } \beta \in \mathbb{Z}[i].$$

But p is irreducible=prime in $\mathbb{Z}[i]$ by Fermat, so either $p \mid a + bi$ or $p \mid a - bi$ (divisibility in $\mathbb{Z}[i]$). WLOG suppose $a + bi = p(c + di)$ for some $c, d \in \mathbb{Z}$, whence $a - bi = p(c - di)$, so

$$m = (a + bi)(a - bi) = p(c + di)p(c - di) = p^2(c^2 + d^2).$$

□

Thus, the integers ≤ 50 which are not a sum of two squares are:

3, 6, 7, 11, 12, 14, 15, 19, 21, 22, 23, 24, 27, 28, 30, 31, 33, 35, 38, 39, 42, 43, 44, 46, 47, 48.

On the other hand, we have

$$\begin{array}{ccccc}
1 = 1^2 + 0^2, & 9 = 3^2 + 0^2, & 18 = 3^2 + 3^2, & 32 = 4^2 + 4^2, & 41 = 5^2 + 4^2, \\
2 = 1^2 + 1^2, & 10 = 3^2 + 1^2, & 20 = 4^2 + 2^2, & 34 = 5^2 + 3^2, & 45 = 6^2 + 3^2, \\
4 = 2^2 + 0^2, & 13 = 3^2 + 2^2, & 25 = 5^2 + 0^2, & 36 = 6^2 + 0^2, & 49 = 7^2 + 0^2, \\
5 = 2^2 + 1^2, & 16 = 4^2 + 0^2, & 26 = 5^2 + 1^2, & 37 = 6^2 + 1^2, & 50 = 7^2 + 1^2. \\
8 = 2^2 + 2^2, & 17 = 4^2 + 1^2, & 29 = 5^2 + 2^2, & 40 = 6^2 + 2^2, &
\end{array}$$

35. GCDs in UFDs

Recall that for $a, b \in R$ in an integral domain, we say $a \mid b$ iff there exists $c \in R$ such that $b = ac$. We have $a \mid b$ iff $b \in (a)$ iff $(b) \subseteq (a)$.

Given a finite set $\{a_1, \dots, a_n\}$ of elements in an integral domain R, say that $d \in R$ is a **greatest common divisor** of $\{a_1, \dots, a_n\}$ if

greatest common divisor

(1) $d \mid a_k$ for $k = 1, \dots, n$, and
(2) if $e \mid a_k$ for $k = 1, \dots, n$, then $e \mid d$.

As we have seen, GCDs can fail to exist: e.g., there is no GCD for $\{6, 2 + 2\sqrt{-5}\}$ in $\mathbb{Z}[\sqrt{-5}]$.

Proposition. *Let $\{a_1, \dots, a_n\}$ be a finite subset of R. Then $d \in R$ is a GCD of the set iff*

(1) *$(a_1, \dots, a_n) \subseteq (d)$, and*
(2) *if $(a_1, \dots, a_n) \subseteq (e)$ for some $e \in R$, then $(d) \subseteq (e)$.*

Proof. Immediate. □

Thus, d is a GCD of $\{a_1, \dots, a_n\}$ iff (d) is the "smallest principal ideal containing $a_1, \dots, a_n$".

This immediately implies that GCDs exist in PIDs: $(d) = (a_1, \dots, a_n)$. Note that this "formula" fails in general, even if a GCD exists.

Example. In $R = \mathbb{Z}[x]$, the element 1 is a GCD of $\{2, x\}$, but $(2, x) \neq (1)$.

We have already shown that $I = (2, x) \neq R$, and that I is not a principal ideal. It is straightforward to show that R/I has only two elements, so if $I \subsetneq (f)$ then $(f) = R$.

Exercise. If every pair $a_1, a_2 \in R$ of elements in R has a GCD, then every finite subset of R has a GCD.

Proposition. *If R is a UFD, then every finite subset of R has a GCD.*

Proof. This is the proof you imagine. Given $\{a_1, \dots, a_n\}$, factor each a_k into irreducibles. It is convenient to choose a fixed representative p for associate classs of prime elements, so that we get $a_k = u_k \prod_p p^{m_{p,k}}$, with $u_k \in R^\times$, $m_{p,k} \geq 0$, and $m_{p,k} = 0$ for all but finitely many p. Then $d = \prod_p p^{n_p}$ with $n_p = \min(m_{k,p},\ k = 1, \dots, n)$ is a GCD. □

Finally, we have the following "cancellation formula" for GCDs in a UFD, which will be needed later.

Proposition. *Let R be a UFD, $\{a_1, \dots, a_n\}$ a finite set of elements in R, and $d, c \in R$ with $c \neq 0$. Then d is a GCD of $\{a_1, \dots, a_n\}$, if and only if dc is a GCD of $\{a_1 c, \dots, a_n c\}$.*

Proof. This is straightforward using irreducible factorizations. □

We say a subset $\{a_1, \dots, a_n\}$ is **relatively prime** if 1 is a GCD. If d is a GCD of a subset $\{a_1, \dots, a_n\}$ of a UFD, then $\{a_1/d, \dots, a_n/d\}$ is a relatively prime subset.

relatively prime

Warning. This is not the same as "pairwise relatively prime", which is the hypothesis we saw in the Chinese Remainder Theorem.

Fractions in lowest terms. Let $F = \operatorname{Frac}(R)$ be the fraction field of R. If R is a UFD, then we can always write elements of F as fractions in "lowest terms" in an essentially unique way.

Proposition. *If $c \in F \smallsetminus \{0\}$, then we can write $c = a/b$ for $a, b \in R$ with $\{a, b\}$ relatively prime. Furthermore, any two such expressions $c = a/b = a'/b'$ differ by a unit: i.e., there exist $u' \in R^\times$ such that $a' = ua$ and $b' = ub$.*

Proof. Existence: straightforward using unique factorization in R. Given c, write $c = a/b$ for some $a, b \in R$, factor numerator and denominator into irreducibles. If an irreducible p divides both a and b, replace with a/p and b/p, until you cannot proceed any further.

Uniqueness: If $c = a/b = a'/b'$ in lowest terms, then $ab' = a'b$. If an irreducible p divides a, it must divide $a'b$ and therefore a'. □

Example. In $R = \mathbb{Z}[\sqrt{-5}]$ (not a UFD), we have $\frac{1+\sqrt{-5}}{2} = \frac{3}{1-\sqrt{-5}}$. These are both in "lowest terms", but they are not "same-up-to-units".

36. Polynomial rings over UFDs

We are going to prove the following theorem.

Theorem. *If R is a UFD, then $R[x]$ is a UFD.*

This immediately generalizes to polynomials in several variables.

Corollary. *If R is a UFD, then $R[x_1, \dots, x_n]$ is a UFD for any $n \geq 1$.*

We are actually going to do more: we will give a kind of classification of irreducibles in $R[x]$, and a kind of algorithm for factoring in $R[x]$. This will involve the following diagram of subrings:

$$\begin{tikzcd} R \arrow[r, tail] \arrow[d, tail] & R[x] \arrow[d, tail] \\ F \arrow[r, tail] & F[x] \end{tikzcd}$$

In particular, any element of $R[x]$ is also an element of $F[x]$. Note that $R^\times = (R[x])^\times$ and $F^\times = (F[x])^\times$.

(It may be helpful to think about the special case $R = \mathbb{Z}$ and $F = \mathbb{Q}$.)

The basic idea of the proof is to factor elements of $R[x]$ into a product of two types of elements: *scalars* and *primitives*, and then to prove unique factorization separately for each type.

The following example gives an idea of how things will work.

Example. Let's consider $S = \mathbb{Z}[x]$, and the element $f = 60x^3 + 30x^2 - 140x - 70$. Its irreducible factorization in S is

$$f = (10)(6x^3 + 3x^2 - 14x - 7) = (2)(5)\ (2x+1)(3x^2 - 7).$$

I did this in two steps. *Step 1.* Factor out the scalar 10, which was the GCD of the coefficients of f. The other factor $g = 6x^3 + 3x^2 - 14 - 7$ is a polynomial whose coefficients form a relatively prime set, called a "primitive" polynomial. *Step 2.* Factor 10 into prime integers, and factor g into irreducible polynomials, which are also primitive.

We will prove that to show the primitive polynomials $2x + 1$ and $3x^2 - 7$ are irreducible in $\mathbb{Z}[x]$, it suffices to show that they are irreducible in $\mathbb{Q}[x]$.

Unique factorization for constant polynomials. For a domain S, we write $\operatorname{Irred}(S) \subset S$ for the subset of irreducible elements.

Regarding R as a subring of $R[x]$, we have the following fact.

Proposition. *If $f, g, h \in R[x]$ are such that $f = gh$, then $f \in R \smallsetminus \{0\}$ iff $g, h \in R \smallsetminus \{0\}$.*

Proof. Use the equation $\deg f = \deg g + \deg h$. □

This implies that an element $f \in R \subseteq R[x]$ is a unit/irreducible/reducible in $R[x]$ iff it is a unit/irreducible/reducible in R, and if reducible it only factors into a product of elements of R. Thus, $\operatorname{Irred}(R) = R \cap \operatorname{Irred}(R[x])$, and the factorizations of an element of R into irreducibles in R are the same as factorizations of it into irreducibles in $R[x]$.

37. Primitive polynomials over UFDs

Let $f = \sum_{k=0}^{n} c_k x^k \in R[x]$. We say that f is **primitive** if the set of its coefficients is relatively prime. That is, if 1 is a GCD for $\{c_0, \ldots, c_n\}$. **primitive**

I'm going to write $\mathrm{Prim}(R[x])$ for the set of primitive polynomials in $R[x]$.

Example. A **monic** polynomial is $f = \sum_{k=0}^{n} c_k x^k$ with $c_n = 1$. Monic polynomials in $R]x]$ are always primitive. **monic**

Example. A degree 0 polyomial $f = a_0$ is primitive iff it is a unit in $R[x]$. Thus $R \cap \mathrm{Prim}(R[x]) = R^\times$.

The first observation is that we can always factor polynomials into a product (scalar) × (primitive), and this is unique up to units in R.

Proposition. *Let $f \in R[x]$ with $f \neq 0$. Then there exist $a \in R$ and $g \in \mathrm{Prim}(R[x])$ such that*

$$f = ag.$$

Furthermore, this factorization is unique up-to-units in R. That is, if

$$f = ag = a'g', \qquad a, a' \in R, \qquad g, g' \in \mathrm{Prim}(R[x]),$$

then there exists $u \in R^\times$ such that $a' = ua$, $g' = u^{-1}g$.

Proof. Consider non-zero $f = \sum_{k=0}^{n} c_k x^k$, $c_k \in R$. Let a be a GCD of $\{c_0, \ldots, c_n\}$. Then $c_k = ad_k$ for some $d_k \in R$, and we can define $g = \sum_{k=0}^{n} d_k x^k$. Thus $f = ag$, and g is primitive since $\{d_0, \ldots, d_n\}$ has GCD 1.

Conversely, if $f = ag$ is *any* factorization with $a \in R$ and $g \in \mathrm{Prim}(R[x])$, the same argument shows that a is a GCD of the coefficients of g. Since GCDs are unique up to units, the claim follows. □

We can extend this to elements of $F[x]$.

Proposition. *Let $f \in F[x]$ with $f \neq 0$. Then there exist $c \in F^\times$ and $g \in \mathrm{Prim}(R)$ such that $f = cg$, and furthermore this factorization is unique up-to-units in R. That is, if*

$$f = cg = c'g' \in F[x], \qquad c, c' \in F, \qquad g, g' \in \mathrm{Prim}(R[x]),$$

then there exists $u \in R^\times$ such that $c' = uc$ and $g' = u^{-1}g$.

Proof. For existence, write $f = \sum_{k=0}^{n} (c_k/d_k) x^k$ with $c_k, d_k \in R$, and let $d = d_0 \cdots d_n \in R \smallsetminus \{0\}$. Then $df \in R[x]$, and by the previous proposition we can write

$$df = ag, \qquad a \in R, \qquad g \in \mathrm{Prim}(R).$$

So $f = cg$ with $c = a/d \in F$.

For uniqueness, suppose $f = cg = c'g'$ with $c, c' \in F$, $g, g' \in \mathrm{Prim}(R[x])$. Then $g' = (c/c')g$. Write $c/c' = a/b$ for some $a, b \in R$, so

$$bg' = ag.$$

This writes an element of $R[x]$ as a product of a scalar (a, b) and a primitive (g, g') in two ways, so by the previous proposition there exist $u \in R^\times$ such that $b = ua$, $g' = u^{-1}g$, and so $c' = uc$. □

Example. In $\mathbb{Q}[x]$, we have

$$\frac{1}{2}x^3 - \frac{1}{3}x + \frac{1}{4} = \frac{1}{12}(6x^3 - 8x + 3).$$

38. Gauss lemma

Here is the key statement. **F 14 Oct**

Proposition (Gauss lemma)**.** *The product of two primitive polynomials is primitive.*

Proof. Let

$$f = a_0 + \cdots + a_m x^m, \qquad g = b_0 + \cdots + b_m x^n, \qquad a_i, b_j \in R.$$

WLOG assume $a_m, b_n \in R \smallsetminus \{0\}$. Then

$$\begin{aligned} fg &= c_0 + c_1 x + c_2 x^2 + \cdots + c_{m+n-1} x^{m+n-1} + c_{m+n} x^{m+n} \\ &= (a_0 b_0) + (a_0 b_1 + a_1 b_0)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \cdots \\ &\qquad (a_{m-1} b_n + a_m b_{n-1})x^{m+n-1} + (a_m b_n)x^{m+n}. \end{aligned}$$

We want to show $\{c_0, \dots, c_{m+n}\}$ is relatively prime.

We suppose $p \in R$ is a prime element such that $p \mid c_k$ for all k, and derive a contradiction. Since f and g are primitive, p cannot divide at least one of each of their coefficients. Let s be minimal such that $p \nmid a_s$, and t minimal such that $p \nmid b_t$. We have

$$c_{s+t} = \underbrace{a_0 b_{s+t} + \cdots + a_{s-1} b_{t+1}}_{p \text{ divides } a_0, \dots, a_{s-1}} + a_s b_t + \underbrace{a_{s+1} b_{t-1} + \cdots + a_{s+t} b_0}_{p \text{ divides } b_{t-1}, \dots, b_0}.$$

Therefore $p \mid a_s b_t$, so p divides one of a_s or b_t, a contradiction. □

Thus $\mathrm{Prim}(R[x])$ is multiplicatively closed. In fact, we have a better property.

Proposition. *Let $f = gh \in R[x]$. Then $f \in \mathrm{Prim}(R[x])$ iff $g, h \in \mathrm{Prim}(R[x])$.*

Proof. We just need the other direction. Write $g = ag'$ and $h = bh'$, with $a, b \in R$ and $g', h' \in \mathrm{Prim}(R[x])$. Then $f = (ab)(g'h')$ is a decomposition into a scalar times a primitive, which is unique up to units in R. Since $f = 1f$ is another such, we see that ab is a unit and so $a, b \in R^\times$, and thus $g = ag'$ and $h = bh'$ are primitive. □

Note: this is actually enough to prove that primitive polynomials have factorizations into irreducible primitives, via induction on degree.

39. Proof that polynomials over a UFD is a UFD

Now we relate factorization in $\mathrm{Prim}(R[x])$ to factorization in $F[x]$. The rule is: if a primitive factors in $F[x]$, you can adjust the factorization by a unit in R to get a factorization in $\mathrm{Prim}(R[x])$.

Proposition. *If*

$$f = gh \in \mathrm{Prim}(R[x]), \qquad g, h \in F[x],$$

then there exist

$$c \in F^\times, \quad g', h' \in \mathrm{Prim}(F[x]) \qquad \text{such that} \qquad g = c^{-1} g', \quad h = ch', \qquad f = g'h'.$$

Proof. By an earlier proposition, we can choose

$$b, c \in F, \quad g_1, h_1 \in \mathrm{Prim}(R[x]) \qquad \text{such that} \qquad g = bg_1, \quad h = ch_1.$$

Then we have two factorizations

$$1f = (bc)(g_1 h_1)$$

into scalar-times-primitive, so they are the same up to units in R. In particular $u = bc \in R^\times$. Set $g' = ug_1$ and $h' = h_1$. Then both $g', h' \in \mathrm{Prim}(R)$, and $g = c^{-1} g'$ and $h = ch'$ with $c \in F^\times$. □

Corollary. *An $f \in \mathrm{Prim}(R[x])$ is irreducible in $R[x]$ iff it is irreducible in $F[x]$.*

Corollary. *A non-unit $f \in \mathrm{Prim}(R[x])$ admits a factorization $f = p_1 \cdots p_r$ into primitive irreducibes $p_1, \dots, p_r$, and this factorization is unique up to reordering and units.*

Proof. Existence. Since non-unit primitives have degree ≥ 1, we can work by induction on degree. If f is an irreducible primitive we are done. Otherwise $f = gh$ for some g, h of degree strictly between 0 and $\deg f$. The elemnets g and h are necessarily primitive, whence g, h can be factored into primitive irreducibles by induction.

(Alternate proof: factor f into irreducibles in $F[x]$, and then "adjust" the factors by units in F to get a product of primitives.)

Uniqueness. If $f = p_1 \cdots p_r = q_1 \cdots q_s$ with p_i, q_j primitive irreducibes, then these are also irreducible factorizations in $F[x]$, so are the same up to reordering and units in F. In particular, $r = s$, and by reordering we may assume $q_k = c_k p_k$ for some $c_k \in F^\times$. But p_k, q_k are primitive, so $c_k \in R^\times$. □

We obtain the desired theorem.

Theorem. *Let R be a UFD. Then $R[x]$ is also a UFD.*

Furthermore, every irreducible $f \in R[x]$ is one of exactly two types.

1. *$f \in R$ and irreducible in R.*
2. *$f \in \mathrm{Prim}(R[x])$ and irreducible in $F[x]$.*

Proof. We prove the second claim first. Every non-zero $f \in R[x]$ can be written $f = ag$ for some $a \in R$ and $g \in \mathrm{Prim}(R[x])$. If f is irreducible then one of a or g is a unit in $R^\times$. If g is a unit then f is irreducible in R, while if a is a unit then f is primitive and irreducible in $F[x]$.

Now we show that F is a UFD. Let $f \in R[x]$ be a non-zero element. *Existence.* $f = ag$ for some $a \in R$, $g \in \mathrm{Prim}(R[x])$. We know that $a = p_1 \cdots p_r$ for some irreducibles $p_i \in R$, and $g = q_1 \cdots q_s$ for some primitive irreducibles $q_j \in \mathrm{Prim}(R[x])$.

Uniqueness. If $f = p_1 \cdots p_r q_1 \cdots q_s = p'_1 \cdots p'_k q_1 \cdots q'_\ell$ with p_i, p'_i irreducible in R, and q_j, q'_j irreducible in $\mathrm{Prim}(R[x])$, then we know there exists a $u \in R^\times$ such that

$$p'_1 \cdots p'_k = u(p_1 \cdots p_r), \qquad q'_1 \cdots q'_\ell = u^{-1}(q_1 \cdots q_s).$$

The result follows from unique factorization in R and unique factorization in $\mathrm{Prim}(R[x])$. □

40. Irreducibility criteria for polynomials

Given $f \in F[x]$ with F a field, when is f irreducible or reducible? How can we find factors of f?

Proposition. *If $f(x) \in F[x]$ and $a \in F$ is such that $f(a) = 0$, then $f = (x-a)g$ for some $g \in F[x]$.*

Proof. By the division algorithm applied to $f \div (x-a)$, there exist $g, r \in F[x]$ with

$$f = (x-a)g + r, \qquad \deg r < \deg(x-a) = 1.$$

Thus $r \in F \subseteq F[x]$. Apply evaluation at a, which is a ring homomorphism:

$$f(a) = \mathrm{ev}_a\big((x-a)g + r\big) = (a-a)g(a) + r(a) = r.$$

Since $f(a) = 0$, we have $f = (x-a)g$. □

Corollary. *If $f \in F[x]$ with $\deg(f) \in \{2,3\}$, then f is reducible iff it has a root in f.*

Let $f \in F[x]$. Say that $c \in F$ is a **root of multiplicity** m if $m \in \mathbb{Z}_{\geq 0}$ is the largest integer such that $(x-c)^m \mid f$ in $F[x]$.

root of multiplicity m

Proposition. *If $f \in F[x]$ with $\deg f = n$, then f has at most n roots in F, even if counted "up to multiplicity".*

Proof. This is just a consequence of the fact that $F[x]$ is a UFD. □

Proposition. *Suppose F is the fraction field of a UFD R, and consider a polynomial in $R[x]$ of the form*

$$f = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \qquad a_k \in R, \quad \deg f = n.$$

If $c \in F$ is a root of f, and if $c = r/s$ with $r, s \in R$ is a fraction in lowest terms, then:

$$r \mid a_0 \qquad \text{and} \qquad s \mid a_n.$$

In particular, if f is monic*, then any roots of f in F are elements $c \in R$ which divide a_0.*

Proof. We have

$$0 = f(r/s) = a_n(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots + a_1(r/s) + a_0.$$

Multiply through by s_n to get

$$0 = a_n r^n + a_{n-1}r^{n-1}s + \cdots a_1 r s^{n-1} + a_0 s^n.$$

an equation in R. Everything except the left-hand term is divisible by s in R, and since $\{r, s\}$ is relatively prime, we get $s \mid a_n$. Likewise, everything except the right-hand term is divisible by r in R, so we get $r \mid a_0$.

If $a_n = 1$, then $s \in R^\times$ and so $c = r/s \in R$ with $c \mid a_0$. □

Example. The polynomial $f = x^3 - 3x - 1 \in \mathbb{Z}[x]$ is irreducible in $\mathbb{Q}[x]$, since by the above the only possible roots are ± 1, but $f(\pm 1) \neq 0$. Because f is monic and thus primitive, it is also irreducible in $\mathbb{Z}[x]$.

Let $I \subseteq R$ be an ideal. Then we can form the ring $(R/I)[x]$ of polynomials over the quotient ring, and we get a surjective ring homomorphism

$$R[x] \twoheadrightarrow (R/I)[x], \qquad \sum c_k x^k \mapsto \pi(c_k)x^k,$$

which I will usually also call π. I'll also write $\overline{f} := \pi(f) \in (R/I)[x]$.

Note: the kernel of this homomorphism is

$$(I)_{R[x]} = I\,R[x] = \{\, \sum c_k x^k \in R[x] \mid c_k \in I \,\},$$

the ideal in $R[x]$ generated by I, so $(R/I)[x] \approx R[x]/IR[x]$.

Proposition. *Let R be an integral domain, and $I \subsetneq R$ a proper ideal. Let $f \in R[x]$ be a monic polynomial of positive degree. If its image $\overline{f} \in (R/I)[x]$ is irreducible in $(R/I)[x]$, then f is irreducible in $R[x]$.*

Proof. Note that since f is monic and $I \neq R$, we have $\overline{f} \neq 0$, and in fact $\deg(\overline{f}) = \deg(f) > 0$. In particular, $\overline{f}$ is not a unit in $(R/I)[x]$.

So to prove the claim, it suffices to show that f reducible implies $\overline{f}$ reducible.

Suppose $f = gh \in R[x]$ with g and h non-units. Then we have $\overline{f} = \overline{g}\overline{h} \in (R/I)[x]$. To show $\overline{f}$ is reducible it suffices to show that both $\overline{g}, \overline{h}$ are non-zero and non-units. It suffices to show $\deg(\overline{g}) = \deg(g)$ and $\deg(\overline{h}) = \deg(h)$, since then both $\overline{g}, \overline{h}$ will be non-constant.

This follows because since f is monic, the leading coefficients of g and h must be units, and so project to non-zero elements in R/I since $I \neq R$. □

Example. Let $R = \mathbb{Q}[x]$ and $I = (x)$. Consider $f = x^3 + y^2 + 3x^2y + 17xy + 1 \in R[y] = \mathbb{Q}[x, y]$. As a polynomial with coefficients in $\mathbb{Q}[x]$ this is monic. Note that $(R/I)[y] \approx \mathbb{Q}[y]$, and reducing mod I amounts to setting $x = 0$, and gives $\overline{f} = y^2 + 1$, which is irreducible in $\mathbb{Q}[y]$, so f is irreducible.

41. Eisenstein criterion

The Eisenstein criterion is sometimes handy for producing irreducible polynomials in $\mathbb{Q}[x]$. It's not always applicable, but it is easy to use when it is.

Proposition (Eisenstein criterion)**.** *Let R be an integral domain with prime ideal $P \subseteq R$, and let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in R[x]$ be a monic polynomial over R. If $a_0, \ldots, a_{n-1} \in P$ and $a_0 \notin P^2$, then f is irreducible in $R[x]$.*

Proof. Suppose $f = gh$ with g, h non-units, whence $\deg g, \deg h < \deg f$. We can project to the quotient ring $(R/P)[x]$, which is an integral domain since R/P is, because P is prime. In the quotient we have

$$x^n = \overline{f} = \overline{g}\overline{h}.$$

This implies $0 = \overline{f} = \overline{g}(0)\overline{h}(0)$, and thus $\overline{g}(0) = \overline{h}(0) = 0$, which means that $g(0), h(0) \in P$. Thus $a_0 = g(0)h(0) \in P^2$, contradicting the hypothesis. (In fact, we must have $\overline{g} = bx^k$, $\overline{h} = cx^{n-k}$ with $b, c \in (R/P)^\times$ and $0 < k < n$.) □

The following example will be important for us.

Example (The cyclotomic polynomial Φ_p). Let $R = \mathbb{Z}$ and $P = (p)$ for some prime p. Then if $f = a_n x^n + \cdots + a_0$ is a monic polynomial in $\mathbb{Z}[x]$ such that $p \mid a_k$ for all $k = 0, \ldots, n-1$, and $p \nmid a_0$, then f is irreducible.

For instance, consider $\Phi_p(x) = \sum_{k=0}^{p-1} x^k \in \mathbb{Z}[x]$. This is a factor of

$$x^p - 1 = (x-1)\Phi_p(x),$$

so roots of Φ_p in $\mathbb{C}$ are $\lambda \in \mathbb{C}$ such that $\lambda^p = 1$ but $\lambda \neq 1$.

Let

$$f(x) = \Phi_p(x+1) = \sum_{k=0}^{p-1} \binom{p}{k+1} x^k = x^p + px^{p-1} + \cdots + \frac{p(p-1)}{2}x + p.$$

This has the Eisenstein property for p, so f is irreducible in $\mathbb{Z}[x]$, and thus in $\mathbb{Q}[x]$. Therefore $\Phi_p(x)$ is also irreducible, since if $\Phi_p(x) = g(x)h(x)$ then $f(x) = g(x+1)h(x+1)$ would be reducible.

(Note: this argument uses the fact that the function $\mathbb{Q}[x] \to \mathbb{Q}[x]$ defined by $f(x) \mapsto f(x+1)$ is an *isomorphism of rings*, and thus takes irreducible elements to irreducible elements.)y

42. Finite subgroups of units in a field

Proposition. *Let F be a field, and $G \leq F^\times$ a finite subgroup of its abelian group of units. Then G is a cyclic group.* **M 17 Oct**

Example. In $\mathbb{C}^\times$, for every n there is a unique cyclic subgroup of n, generated by $\zeta := e^{2\pi i/n}$. (You can also take ζ^k as a generator for this subgroup if k is relatively prime to n.)

Example. If F is a finite field, then $F^\times$ is a cyclic group. In particular, $\mathbb{F}_p^\times = (\mathbb{Z}/p)^\times$ is a cyclic group of order $p-1$.

For instance, $(\mathbb{Z}/7)^\times$ is cyclic of order 6, and in fact $(\mathbb{Z}/7)^\times = \langle [3] \rangle = \langle [5] \rangle$.

We need the following fact: for any $n \geq 1$, the set $\{\, c \in F^\times \mid c^n = 1 \,\}$ has at most n distinct elements. Proof: this set is precisely the roots of the polynomial $f = x^n - 1 \in F[x]$, which has degree n. Then the claim is a consequence of the following.

Proposition. *Let G be a finite abelian group such that, for every prime number p, there are at most p elements $g \in G$ such that $g^p = 1$. Then G is cyclic.*

Proof. In this proof I'm going to write $G^m := \{\, g^m \mid g \in G \,\}$. Because G is abelian, this subset is always a subgroup.

I argue by induction on $n = |G|$. So let G be a non-trivial group satisfying the hypothesis. In any non-trivial finite group there exists an element of G with order > 1, and hence (by taking an appropriate power) we can choose an element $c \in G$ with $|c| = p$ for some prime p.

Because G is abelian, the function $\phi(x) := x^p$ defines a surjective homomorphism of groups

$$\phi\colon G \twoheadrightarrow G^p \leq G.$$

By hypothesis, $\operatorname{Ker}(\phi) = \{\, x \in G \mid x^p = 1 \,\}$ has at most p elements, and since it contains an element c with order p, we have $\operatorname{Ker}(\phi) = \langle c \rangle$, a subgroup of order p. Thus the function ϕ is a "p-to-1" surjective function: every $a \in G^p$ has exactly p distinct "pth roots", i.e., elements $b \in G$ such that $b^p = a$. (The point is that each preimage $\phi^{-1}(a)$ is a coset of $\operatorname{Ker}(\phi) = \langle c \rangle$.)

By the first isomorphism theorem, $G^p \approx G/\operatorname{Ker}\phi$ and so $|G^p| = n/p < n$. Since G^p is a subgroup of G, it also satisfies the hypothesis of the proposition, so by induction is cyclic on some generator we will call $a \in G^p$.

Since $a \in G^p$, there exists a $b \in G$ such that $b^p = a$. In fact, the set $S := \{\, b \in G \mid b^p = \phi(b) = a \,\}$ has size exactly p, since it is a coset of $\operatorname{Ker}(\phi) = \langle c \rangle$. That is, a has exactly p distinct "pth roots" in G.

Claim. Any $b \in S \smallsetminus G^p$ is a cyclic generator of G. In fact, given a $b \in G \smallsetminus G^p$ with $b^p = a$, we have

$$G^p = \langle a \rangle \lneq \langle b \rangle \leq G,$$

and since $|G : G^p| = p$ is prime, we must have $G = \langle b \rangle$.

So we are reduced to showing that $S \smallsetminus G^p$ is non-empty. In fact, I can show that *at most one* of these pth roots b of a is in G^p. Clearly if $S \cap G^p = \varnothing$ we are done, since S is non-empty. So suppose $b \in G^p$ with $b^p = a$. Then the restricted homomorphism

$$\phi|_{G^p} \colon G^p \to G^p = \langle a \rangle$$

is surjective, and thus a bijection since both domain and codomain are finite sets of the same order. Thus in this case there is exactly one $b \in G^p$ with $b^p = a$.

Therefore, there are either p or $p-1$ distinct elements $b \in G \smallsetminus G^p$ such that $b^p = a$, so there is at least one, as desired. □

43. Noetherian rings

Let R be a commutative ring with 1. We say that R is **Noetherian** if it has the ascending chain condition for ideals. That is, if $\{I_k\}_{k \in \mathbb{Z}_{>0}}$ is a sequence of ideals with $I_k \subseteq I_{k+1}$, then there exists $n > 0$ such that $I_k = I_n$ for all $k \geq n$. (Note: this is similar to, but different than, the ascending chain condition for *principal ideals* which we saw in the proof that a PID is a UFD.) **Noetherian**

This is not the definition as given in the book, but it is equivalent to it.

Theorem. *Let R be a commutative ring with 1. Then R has the acc for ideals iff every ideal in R is finitely generated.*

Proof. This is the same idea as something we saw earlier: a group has the acc for subgroups iff every subgroup is finitely generated.

$\Longrightarrow$. I'll prove the converse. If $I \subseteq R$ is a non-finitely generated ideal, then we can choose a sequence $f_k \in I$, $k \geq 1$, such that $f_{k+1} \notin I_k := (f_1, \ldots, f_k)$ for all k. This given a strictly increasing chain $(f_1) \subsetneq (f_1, f_2) \subsetneq \cdots$, which implies R does not have the acc for ideals.

$\Longleftarrow$. Suppose every ideal is finitely generated. Then for a chain $I_1 \subseteq I_2 \subseteq \cdots$ of ideals, let $J := \bigcup_{k=1}^{\infty} I_k$. This J is also an ideal, so by hypothesis $J = (f_1, \ldots, f_m)$ for some finite subset $\{f_1, \ldots, f_m\}$. But then there exists n such that $f_1, \ldots, f_m \in I_n$, whence $I_n = J$. □

Clearly, any PID is Noetherian.

44. Hilbert basis theorem

Theorem (Hilbert basis theorem)**.** *If R is a commutative ring with 1 which is Noetherian, then $R[x]$ is Notherian.*

Corollary. *If R is Noetherian, then $R[x_1, \dots, x_n]$ is Noetherian.*

In particular, all ideals in polynomial rings over fields are finitely generated.

Proof. Suppose $I \subseteq R[x]$ is an ideal which is not finitely generated, and derive a contradiction. We can choose a sequence of elements $f_k \in I$, $k \geq 1$, such that f_k has minimal degree among elements of $I \smallsetminus (f_1, \dots, f_{k-1})$. (That is, take f_1 to be an element of minimal degree in $I \smallsetminus (0)$, f_2 an element of minimal degree in $I \smallsetminus (f_1)$, etc.)

Write $d_k := \deg f_k \geq 0$, and note that $d_k \leq d_{k+1}$ for all k. Write $a_k \in R \smallsetminus \{0\}$ for the leading coefficient of the polynomial f_k, so that

$$f_k = a_k x^{d_k} + (\text{lower degree terms}).$$

Consider the ideal $J := (a_k,\ k \geq 1) \subseteq R$ generated by leading coefficients. Since R is Noetherian, the chain of ideals

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

must terminate, so there exists n such that $a_n \in (a_1, \dots, a_{n-1})$, i.e.,

$$a_n = c_1 a_1 + \cdots + c_{n-1} a_{n-1}, \qquad c_1, \dots, c_{n-1} \in R.$$

Let

$$g := c_1 x^{d_n - d_1} f_1 + \cdots + c_{n-1} x^{d_n - d_{n-1}} f_{n-1} = \sum_{k=1}^{n-1} c_k x^{d_n - d_k} f_k \in R[x]$$

Note that

$$x^{d_n - d_k} f_k = a_k x^{d_n} + (\text{lower degree terms}), \qquad 1 \leq k < n,$$

so that by construction g has the same leading term as f_n, i.e., $g = a_n x^{d_n} + \text{lower degree terms}$, and also $g \in (f_1, \dots, f_{n-1})$. Thus $h := f_n - g \in I \smallsetminus (f_1, \dots, f_{n-1})$, but $\deg(h) < \deg(f_n)$, contradicting minimiality of $\deg(f_n)$ in $I \smallsetminus (f_1, \dots, f_{n-1})$.

□

45. Modules

(DF 10.1, 10.2, 10.3)

Let R be a ring (not necessarily commutative, but with 1). A **(left) R-module** M is a triple $(M, +, \cdot)$ consisting of **(left) R-module**

1. an abelian group $(M, +)$, together with
2. a function $R \times M \to M$ (denoted $(r, m) \mapsto rm$), satisfying
   - (a) $(r_1 + r_2)m = r_1 m + r_2 m$,
   - (b) $r(m_1 + m_2) = r m_1 + r m_2$,
   - (c) $r_1(r_2 m) = (r_1 r_2)m$,
   - (d) $1m = m$.

Note that:

- $0m = 0$ for $0 \in R$ and $m \in M$, since $(0+0)m = 0m + 0m = 0m$ by (a).
- $(-1)m = -m$, since $0m = (1 + (-1))m = 1m + (-1)m$.

Example. If F is a field, then an F-module is exactly the same thing as an F-**vector space**. A homomorphism of F-modules is exactly the same thing as an F-linear map between vector spaces. **vector space**

Example. An abelian group M admits a unique structure of a $\mathbb{Z}$-module: the axioms imply that $nm = \underbrace{m + \cdots + m}_{n \text{ times}}$ and $(-n)m = -(nm)$ if $n \in \mathbb{Z}_{>0}$. Thus $\mathbb{Z}$-modules are the same thing as abelian groups, and homomorphisms of $\mathbb{Z}$-modules are the same things as abelian groups.

If you think of modules as a common generalization of abelian groups and vector spaces, then you can predict most of the basic theory.

Example (Free module of rank n). For $n \geq 0$, let $R^n = \{ (a_1, \ldots, a_n) \mid a_i \in R \}$. Then R^n is an R-module, with componentwise addition and $r(a_1, \ldots, a_n) := (ra_1, \ldots, ra_n)$. This is called a **free module of rank** n.

free module of rank n

In particular, we can take $n = 1$, so R is an R-module.

46. Submodules

W 19 Oct

A **submodule** of an R-module M is a subset $N \subseteq M$ such that $(N, +)$ is a subgroup of $(M, +)$, and $r \in R$ and $n \in N$ imply $rn \in N$.

submodule

Exercise. A subset $N \subseteq M$ is a submodule iff (i) $(N, +)$ is a subgroup of $(M, +)$ and (ii) for all $r \in R$ and $n \in N$, we have $r \cdot n \in N$.

Example. A submodule of a $\mathbb{Z}$-module is the same as a subgroup. For a field F, a submodule of an F-module is the same as a vector subspace.

Example. Let V be a $F[x]$-module, thought of as an F-vector space equipped with a linear map $T \colon V \to V$ corresponding to $Tv = xv$. Then $W \subseteq V$ is a submodule if and only if it is a T-invariant subspace of V.

Example. Consider R as a module over itself. Then a submodule of R is the same thing as a left ideal.

47. Right modules

Let R be a ring (not necessarily commutative, but with 1). A **right R-module** M is

right R-module

(1) an abelian group $(M, +)$, together with
(2) a function $M \times R \to M$ (denoted $(m, r) \mapsto mr$), satisfying
   (a) $m(r_1 + r_2) = mr_1 + mr_2$,
   (b) $(m_1 + m_2)r = m_1 r + m_2 r$,
   (c) $(mr_1)r_2 = m(r_1 r_2)$,
   (d) $m = m1$.

Consider a ring $(R, +, \mu)$, where for the moment we choose to write multiplication as a function $\mu(a, b) = ab$. Let $\mu^{\mathrm{op}} \colon R \times R \to R$ be the function $\mu^{\mathrm{op}}(a, b) := \mu(b, a)$. Then the triple $(R, +, \mu^{\mathrm{op}})$ is also a ring, called the **opposite ring**, and usually denoted R^{op}. Note that its underlying abelian group is the same as that of R, but the multiplication is different.

opposite ring

Example. If R is commutative ring, then $\mu = \mu^{\mathrm{op}}$, and thus $R = R^{\mathrm{op}}$.

Exercise. For any ring, $(R^{\mathrm{op}})^{\mathrm{op}} = R$.

Let $(M, +, \lambda)$ be a left R-module, where $\lambda \colon R \times M \to M$ is the action $\lambda(r, m) := rm$. Define $\lambda^{\mathrm{op}} \colon M \times R \to M$ by $\lambda^{\mathrm{op}}(m, r) = \lambda(m, r)$.

Proposition. *The data $(M, +, \lambda)$ is a left R-module iff $(M, +, \lambda^{\mathrm{op}})$ is a right R^{op}-module.*

Thus, left R-modules are the same thing as right R^{op}-modules.

In particular, if R is a commutative ring, then every left R-module can also be made into a right R-module, via $mr := rm$.

48. Module homomorphisms

An **module homomorphism** is a function $\phi\colon M \to M'$ between left R-modules which satisfies

module homomorphism

(1) $\phi(m_1 + m_2) = \phi(m_1) + \phi(m_2)$,
(2) $\phi(rm) = r\phi(m)$.

It is an isomorphism if it is bijective, in which case the inverse map $\phi^{-1}\colon N \to M$ is also an isomorphism.

The objects $\mathrm{Ker}(\phi) = \{\, m \in M \mid \phi(m) = 0 \,\}$ and $\phi(N)$ are submodules of M and N respectively.

We write $\mathrm{Hom}_R(M, N)$ for the set of (left) R-module homomorphisms.

Remark. We can also define homomorphisms of right R-modules, and we obtain a set $\mathrm{Hom}_R^{\mathrm{right}}(M, N)$ of homomorphisms of right R-modules.

We have the following, which are straightforward verifications.

(1) If $\phi, \psi \in \mathrm{Hom}_R(M, N)$, then the function $\chi = \phi + \psi$ defined by $(\phi + \psi)(m) = \phi(m) + \psi(m)$ is also an R-module homomorphism $M \to N$. The + operation makes $\mathrm{Hom}_R(M, N)$ into an abelian group.
(2) If $\phi \in \mathrm{Hom}_R(M, N)$ and $\psi \in \mathrm{Hom}_R(L, M)$, then $\phi \circ \psi \in \mathrm{Hom}_R(L, M)$. We have
$$\phi \circ (\psi_1 + \psi_2) = \phi \circ \psi_1 + \phi \circ \psi_2, \qquad (\phi_1 + \phi_2) \circ \psi = \phi_1 \circ \psi + \phi_2 \circ \psi.$$
(3) If M is a left R-module, then $\mathrm{Hom}_R(M, M)$, equipped with addition and with composition, is ring with 1, called the **endomorphism ring** of M and denoted $\mathrm{End}_R(M)$.

endomorphism ring

(4) If R is a *commutative* ring with 1, then for $r \in R$ and $\phi \in \mathrm{Hom}_R(M, N)$ the function $r\phi\colon M \to N$ defined by $(r\phi)(m) := r\phi(m)$ is an R-module homomorphism. This operation makes $\mathrm{Hom}_R(M, N)$ into an R-module.

Example. Consider $\phi \in \mathrm{Hom}_R(R^m, R^n)$. If we write $u_i = (0, \dots, 1, \dots, 0) \in R^m$ and $v_i = (0, \dots, 1, \dots, 0) \in R^n$, then
$$\phi(u_i) = (a_{i1}, \dots, a_{in}) = \sum_j a_{ij} v_j$$
for some $a_{ij} \in R$. Then if $x = \sum_i x_i u_i \in R^m$, we have
$$\phi(x) = \sum_i x_i \phi(u_i) = \sum_{i,j} x_i a_{ij} v_j = (\sum_i x_i a_{i1}, \dots, \sum_i x_i a_{in}).$$
In other words, $\phi(x) = xA$ where x is thought of as a row vector in R and $A = (a_{ij}) \in M_{m\times n}(R)$.

Conversely, given $A = (a_{ij}) \in M_{m\times n}(R)$ we can use these formulas to define a homomorphism $\rho_A\colon R^m \to R^n$. In other words, there is an evident correspondence $\mathrm{Hom}_R(R^m, R^n) \approx M_{m\times n}(R)$, so that if we identify elements of R^m with $1 \times m$ row vectors x, then an $m \times n$ matrix A corresponds to a homomorphism defined by $x \mapsto xA$.

Furthermore, composition of linear maps corresponds to multiplication of matrices (but with a change of order): $\rho_A \circ \rho_B = \rho_{BA}$.

If we think about *right* modules and right module homomorphisms, then we can identity $A \in M_{m\times n}(R)$ with $\lambda_A \in \mathrm{Hom}_R^{\mathrm{right}}(R^n, R^m)$ so that $\lambda_A(x) = Ax$, thinking of elements in R^n as column vectors.

If R is commutative, right and left modules are the same thing, so we can use either formalism. In particular, $\mathrm{End}_R(R^n) \approx M_{n\times n}(R)$.

The **automorphisms** of a module $\mathrm{Aut}_R(M) \subseteq \mathrm{End}_R(M)$ is the set of homomorphisms which are isomorphisms. This is the same thing as the group of units $\mathrm{End}_R(M)^\times$ in the endomorphism ring. It is not an abelian group under addition, but it is a (usually non-abelian) group under composition.

automorphisms

49. Quotient modules and isomorphism theorems

Let M be an R-module and $N \subseteq M$ a submodule. Then the quotient group M/N obtains the structure of an R-module, via

$$r(m+N) := (rm) + N, \qquad r \in R, \quad m \in M.$$

This is the **quotient module** of M by N. There is a corresponding **quotient homomorphism** $\pi\colon M \to M/N$.

quotient module

quotient homomorphism

We have a "homomorphism theorem" for modules.

Proposition. *Let $\phi\colon M \to N$ be a homomorphism of R-modules, and $A \subseteq M$ a submodule. If $A \subseteq \operatorname{Ker}(\phi)$ then there exists a unique module homomorphism $\overline{\phi}\colon M/A \to N$ such that $\overline{\phi}(m+A) = \phi(m)$.*

$$\begin{tikzcd}
M \arrow[r, "\phi"] \arrow[d, two heads, "\pi"'] & S \\
M/A \arrow[ur, dotted, "\overline{\phi}"'] &
\end{tikzcd}$$

Theorem (First isomorphism theorem for modules)**.** *If $\phi\colon M \to N$ is an R-module homomorphism, then $\operatorname{Ker}(\phi)$ is a submodule of M, $\phi(M)$ is a submodule of N, and we have an isomorphism $R/\operatorname{Ker}(\phi) \approx \phi(M)$ of R-modules.*

That is, the homomorphism ϕ factors through an isomorphism $\overline{\phi}\colon M/\operatorname{Ker}(\phi) \xrightarrow{\sim} \phi(M)$.

$$\begin{tikzcd}
M \arrow[rr, "\phi"] \arrow[dr, two heads] & & N \\
& M/\operatorname{Ker}(\phi) \arrow[r, "\overline{\phi}", "\sim"'] & \phi(M) \arrow[u, hook]
\end{tikzcd}$$

Theorem (Second (diamond) isomorphism theorem for modules)**.** *Let $A, B \subseteq M$ be submodules.*

1. *$A+B$ is a submodule of M.*
2. *B is a submodule of $A+B$.*
3. *$A \cap B$ is a submodule of A.*
4. *$A/(A\cap B) \approx (A+B)/B$.*

The isomorphism of (4) sends $x + (A \cap B) \mapsto x + B$.

Theorem (Third isomorphism theorem for modules)**.** *Let $A, B \subseteq M$ be submodules with $A \subseteq B$. Then*

1. *B/A is a submodule of M/A, and*
2. *$M/B \approx (M/A)/(B/A)$.*

The isomorphism of (2) sends $x + B \mapsto (x+A) + (B/A)$.

Theorem (Fourth (lattice) isomorphism theorem for modules)**.** *Let $N \subseteq M$ be a submodule. Then we have inverse bijections*

$$\{\, \textit{submodules } A \subseteq M \mid N \subseteq A \,\} \overset{\sim}{\longleftrightarrow} \{\textit{submodules } \overline{A} \subseteq M/N\}$$

$$\begin{tikzcd}
A \arrow[r, mapsto] & A/N \\
\pi^{-1}\overline{A} & \overline{A} \arrow[l, mapsto]
\end{tikzcd}$$

where $\pi^{-1}\overline{A} = \{\, x \in M \mid \pi(x) \in \overline{A} \,\}$. Furthermore for submodules $A, B \subseteq M$ with $N \subseteq A \cap B$, we have

1. *$A \subseteq B$ iff $A/N \subseteq B/N$.*
2. *$(A \cap B)/N = (A/N) \cap (B/N)$.*

50. Finitely generated modules

Let $A \subseteq M$ be a subset of an R-module M. Let

$$RA = \{\, r_1 a_1 + \cdots + r_k a_k \in M \mid r_i \in R,\ a_i \in A,\ k \geq 0 \,\}.$$

Then $RA \subseteq M$ is a submodule of M.

Proposition. *We have that*

$$RA = \bigcap_{\substack{\text{submodule } N \subseteq M \\ A \subseteq N}} N.$$

Thus, RA is the smallest *submodule of M containing the set A.*

If $M = RA$, we say that M is **generated** as a module by the set A.

generated

Example. Let F be a field and V an F-module, i.e., an F-vector space. Then $A \subseteq V$ generates V iff it spans V in the sense of linear algebra.

Say M is a **finitely generated** module if it admits a finite generating set, and a **cyclic** if it admits a generating set of size 1.

finitely generated
cyclic

Example. Let R be a ring and $I \subseteq R$ a *left ideal.* Then $I \subseteq R$ is a submodule, and thus we can form the quotient module R/I. The module R/I is cyclic: every element of R/I is of the form $r(1+I) = r + I$.

Proposition. *If M is a cyclic R-module, then there is an isomorphism of R-modules $M \approx R/I$ for some left ideal $I \subseteq R$.*

Proof. Let $\{a\} \subseteq M$ be a singleton subset such that $R\{a\} = M$. Define $\phi\colon R \to M$ by $\phi(r) := ra$. This is a surjective R-module homomorphism, and so by the 1st isomorphism theorem we get an isomorphism $\overline{\phi}\colon R/I \to M$ where $I = \operatorname{Ker}(\phi)$. □

51. Direct sum and product of modules

Let $(M_i)_{i\in I}$ be an indexed collection of R-modules.

The **product** (or **direct product**) of these is the module

product
direct product

$$\prod_{i\in I} M_i := \{\, (x_i)_{i\in I} \mid x_i \in M_i \,\},$$

where the operations are defined componentwise: $(x_i) + (y_i) = (x_i + y_i)$ and $r(x_i) = (rx_i)$.

The **direct sum** (or **coproduct**) of these is the submodule

direct sum
coproduct

$$\bigoplus_{i\in I} M_i := \{\, (x_i) \in \prod_i M_i \mid |\{\, i \in I \mid x_i \neq 0 \,\}| < \infty \,\},$$

consisting of tuples such that only finitely many entries are 0.

Note that if $|I| < \infty$, then $\bigoplus_{i\in I} M_i = \prod_{i\in I} M_i$. In this case we usually write

$$M_1 \oplus \cdots \oplus M_n = M_1 \times \cdots \times M_n.$$

Proposition. *Let $N_1, \ldots, N_k \subseteq M$ be submodules, and let $N = N_1 + \cdots + N_k \subseteq N$. TFAE.*

1. *The map $\phi\colon N_1 \oplus \cdots \oplus N_k \to N$ defined by $\phi(x_1, \ldots, x_k) := x_1 + \cdots + x_k$ is an isomorphism of modules.*

(2) $N_j \cap (N_1 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$ *for all* $j = 1, \ldots, k$.
(3) *Every* $x \in N$ *can be written* uniquely *in the form* $x = x_1 + \cdots + x_k$ *with* $x_j \in N_j$.

In the above setting, we say that the submodule N is an **internal direct sum** of the submodules $N_1, \ldots, N_k$.

internal direct sum

52. Free modules

F 21 Oct

free R-module

Let R be a ring with 1. A **free R-module** on a set S is a pair (M, e), consisting of an R-module M, and a function $e\colon S \to M$ (whose values I will write as $e_s \in M$ for $s \in S$), with the following property: for every $x \in M$, there exists a unique expression of the form

$$x = \sum_{s \in S} a_s e_s, \qquad |\{\, s \in S \mid a_s \neq 0 \,\}| < \infty.$$

That is, for each x there exists a unique S-indexed tuple $(a_s)_{s \in S}$ of elements of R, only finitely many of which are non-zero, making the above identity hold.

This is easier to discuss when S is finite. If $S = \{1, \ldots, n\}$, then (M, e) is free if every $x \in M$ has the form $x = \sum_{k=1}^{n} a_s e_s$ for unique $a_1, \ldots, a_n \in R$.

Example. If $S = \{1, \ldots, n\}$, let $M := R^n$ and let $e_k := (\delta_{k1}, \ldots, \delta_{kn})$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Then this (M, e) is a free module on S.

Example. If $S = \varnothing$, then $M = 0$ is a free module on S.

Proposition. *A free R-module exists for every set S.*

Proof. First, suppose given a set S. Let

$$M := \bigoplus_{s \in S} R = \{\, (a_s \in R)_{s \in S} \mid |\{\, s \in S \mid a_s \neq 0 \,\}| < \infty \,\}.$$

Write $e_s \in M$ for the element $(\delta_{st})_{t \in S} \in M$. Then it is immediate that (M, e) defines a free module. □

A function $e\colon S \to M$ such that (M, e) is a free module on S is called a **basis** for M. A module can have many different bases (or no bases, if it is not a free module).

basis

Example. If V is a vector space over a field F, then a basis for V in this sense is exactly the same as a basis of the vector space.

Exercise. With our definition, it is *not* in general the case that the basis elements are pairwise distinct, i.e., that $e_s \neq e_t$ when $s \neq t$. But there really is only one counterexample: the trivial ring $R = \{0\}$. For the trivial ring there is only one module (up to isomorphism), namely the trivial module $M = \{0\}$. In this case, M is free on *every* set S, via the function $e\colon S \to M$ with $e_s = 0$ for all s.

If R is a ring with $1 \neq 0$, then for any free module (M, e) on a set S, we do have that $s \neq t$ implies $e_s \neq e_t$. Thus, as long as R is not the trivial ring, we can identify the basis of a free module with the *subset* $\{\, e_s \mid s \in S \,\}$ of basis elements.

Note: the defintion of free module is phrased slightly differently in DF §10.3, basically in order to avoid this weird counterexample, so they can always regard the indexed collection $\{e_i\}$ as a *subset* of the free module.)

Universal property of free modules.

Proposition. *Let (M, e) be a free R-module on some set S. Then for every R-module N and function $\phi\colon S \to M$, there exists a* unique *R-module homomorphism $\overline{\phi}\colon M \to N$ such that $\overline{\phi} \circ e = \phi$.*

$$\begin{tikzcd} S \arrow[r, "\phi"] \arrow[d, "e"'] & N \\ M \arrow[ur, dotted, "\exists!\ \overline{\phi}"'] & \end{tikzcd}$$

That is, for any R-module N there is a bijection

$$\begin{tikzcd} \operatorname{Hom}_R(M,N) \arrow[r, leftrightarrow] & \{\textit{functions } \phi\colon S \to N\} \\ \overline{\phi} \arrow[r, mapsto] & \overline{\phi} \circ e \end{tikzcd}$$

Proof. Given ϕ, define $\overline{\phi}$ by

$$\overline{\phi}\big(\sum_{s\in S} a_s e_s\big) := \sum_{s \in S} a_s \phi(s).$$

This is an R-module homomorphism such that $\overline{\phi} \circ e = \phi$, homomorphism, and is the only one sending $e_s \mapsto \phi(s)$. □

Corollary. *If M and N are both R-modules which are free on a set S then M and N are isomorphic.*

Proof. Let $(e_s)_{s\in S}$ and $(f_s)_{s \in S}$ be free bases for M and N respectively. By the universal property, there exist homomorphisms $\phi\colon M \to N$ and $\psi\colon N \to M$ such that $\phi(e_s) = f_s$ and $\psi(f_s) = e_s$. It is clear that $\psi \circ \phi = \mathrm{id}_M$ nad $\phi \circ \psi = \mathrm{id}_N$. □

53. Bilinear maps

Let M, N, A be abelian groups ($= \mathbb{Z}$-modules). A **bilinear function** $\beta\colon M \times N \to A$ is a function such that

bilinear function

(1) $\beta(m_1 + m_2, n) = \beta(m_1, n) + \beta(m_2, n)$ for $m_1, m_2 \in M$, $n \in N$, and
(2) $\beta(m, n_1 + n_2) = \beta(m, n_1) + \beta(m, n_2)$ for $m \in M$, $n_1, n_2 \in N$.

In other words, a function is bilinear iff it is bilinear separately in each variable. Note that this implies that $\beta(m, 0) = 0 = \beta(0, n)$.

Exercise. If $\beta\colon M \times N \to A$ is bilinear, then for any integer r we have

$$\beta(mr, n) = \beta(m, rn).$$

Example. Let $M = \mathbb{Z}^m$ and $N = \mathbb{Z}^n$ (free abelian groups), with free bases $\{u_1, \dots, u_m\}$ and $\{v_1, \dots, v_n\}$ respectively. Then it is easy to see that for any $m \times n$-tuple (a_{ij}) of elements of some abelian group A, we can define a bilinear function by

$$\beta\colon M \times N \to A, \qquad \beta\big(\sum_i x_i u_i,\ \sum_j y_j v_j\big) = \sum_{i,j} x_i y_j a_{ij}.$$

Furthermore, this is the unique bilinear function such that $\beta(u_i, v_j) = a_{ij}$.

Example. Let $M = \mathbb{Z}/m$ and $N = \mathbb{Z}/n$, with $m, n \in \mathbb{Z}_{>0}$, and consider a bilinear function $\beta\colon M \times N \to A$. For $x, y \in \mathbb{Z}$ we have

$$\beta([x]_m,\ [y]_n) = \beta(x[1]_m,\ y[1]_n) = x\,\beta([1]_m,\ y[1]_n) = xy\,\beta([1]_m,\ [1]_n).$$

Write $a := \beta([1]_m,\ [1]_n)$. Note that if $d = my + xn$ for some $d, x, y \in \mathbb{Z}$, then

$$\beta([m]_m,\ [y]_n) + \beta([x]_m,\ [n]_n) = (my)a + (xn)a = da,$$

but also

$$\beta([m]_m,\, [y]_n) = \beta([0]_m,\, [y]_n) = 0, \qquad \beta([x]_n,\, [n]_n) = \beta([x]_n,\, [0]_n) = 0,$$

and thus $da = 0$.

In particular, if $\gcd(m, n) = 1$, the only bilinear function $\mathbb{Z}/m \times \mathbb{Z}/n \to A$ is the constant homomorphism 0. More generally, there is a bijective correspondence

$$\big\{\text{bilinear functions } \mathbb{Z}/m \times \mathbb{Z}/n \to A\big\} \leftrightarrow \big\{a \in A \text{ such that } da = 0\big\},$$

where $d = \gcd(m, n)$.

Let R be a ring with 1 (but possibly non-commutative). Suppose a right R-module, N is a left R-module, and A is an abelian group. Then an ***R*-balanced bilinear function** $\beta\colon M \times N \to A$ is a function is a bilinear function such that also

R-balanced bilinear function

(3) $\beta(mr, n) = \beta(m, rn)$ for $m \in M$, $n \in N$, $r \in R$.

Note: if $R = \mathbb{Z}$, then any bilinear map is already $\mathbb{Z}$-balanced.

If R is commutative, then right and left modules are the same. In this case, if A is also an R-module, that $\beta\colon M \times N \to A$ is ***R*-bilinear**, if

R-bilinear

(3) $\beta(mr, n) = \beta(m, rn) = r\,\beta(m, n)$ for $m \in M$, $n \in N$, $r \in R$.

Exercise. Let M, N, A be modules over a commutative ring R, with submodules $M' \subseteq M$ and $N' \subseteq N$. Then there is a bijection between

(1) R-bilinear maps $\overline{\beta}\colon M/M' \times N/N' \to A$, and
(2) R-bilinear maps $\beta\colon M \times N \to A$ such that $\beta(M, N') = 0 = \beta(M', N)$.

54. Tensor products

Let M be a right R-module and N be a left R-module. We define an abelian group $M \otimes_R N$, called the **tensor product** of M and N over R, together with an bilinear map $\alpha\colon M \times N \to M \otimes_R N$, as follows.

tensor product

Let $\mathbb{Z}[M \otimes N]$ denote free abelian group on the set $M \times N$. We write $[m, n] \in \mathbb{Z}[M \times N]$ for the element in the free basis corresponding to $(m, n) \in M \times N$. Let $G \subseteq \mathbb{Z}[M \times N]$ be the subgroup generated by all elements of the forms:

(1) $[m_1, n] + [m_2, n] - [m_1 + m_2, n]$, for $m_1, m_2 \in M$, $n \in N$,
(2) $[m, n_1] + [m, n_2] - [m, n_1 + n_2]$, for $m \in M$, $n_1, n_2 \in N$.
(3) $[mr, n] - [m, rn]$, for $m \in M$, $n \in N$, $r \in R$.

We define

$$M \otimes_R N := \mathbb{Z}[M \times N]/G.$$

Remark. If $R = \mathbb{Z}$, then the subgroup generated by elements of types (1) and (2) automatically contains those elements of type (3).

Write "$m \otimes n$" for the coset of $[m, n]$ in $M \otimes_R N$. Define

$$\alpha\colon M \times N \to M \otimes_R N, \qquad \alpha(m, n) := m \otimes n.$$

The function α is seen to be R-bilinear, by construction.

Proposition. *The function $\alpha\colon M \times N \to M \otimes_R N$ defined by $(m, n) \mapsto m \otimes n$ is the universal R-bilinear map out of $M \times N$. That is, for any abelian group A and R-bilinear map $\beta\colon M \times N \to A$, there exists a unique group homomorphism $\phi\colon M \otimes_R N \to A$ such that $\phi \circ \alpha = \beta$.*

$$\begin{tikzcd}
M \times N \arrow[r, "\beta"] \arrow[d, "\alpha"'] & A \\
M \otimes N \arrow[ur, dotted, "\exists!\ \phi"'] &
\end{tikzcd}$$

In other words, there is a bijection

$$\begin{tikzcd}\big\{\text{group homomorphisms } \phi\colon M \otimes_R N \to A\big\} \arrow[r, leftrightarrow] & \big\{R\text{-balanced bilinear maps } \beta\colon M \times N \to A\big\} \\ \phi \arrow[r, mapsto] & \phi \circ \alpha\end{tikzcd}$$

Proof. By the property of free abelian groups,

$$\{\text{homomorphisms } \psi\colon \mathbb{Z}[M \times N] \to A\} \longleftrightarrow \{\text{functions } \beta\colon M \times N \to A\}.$$

Verify that $\psi(G) = \{0\}$ if and only if β is R-bilinear. By the homomorphism theorem for abelian groups, in this case there exists a unique homomorphism $\phi\colon M \otimes_R N \to A$ extending ψ. □

Remark. Because $M \otimes_R N$ is a quotient of the free abelian group on $M \times N$, every element of $M \otimes_R N$ can be written

$$\sum_{k=1}^{r} m_k \otimes n_k, \qquad m_k \in M, \quad n_k \in N, \quad k \geq 0.$$

An element which can be written as $m \otimes n$ is sometimes called a *simple tensor.* It is rarely true that all elements of the tensor product are simple tensors.

Remark. When R is commutative, we can assume both M and N are left R-modules (by equating rm with mr). In this case, we can give $M \otimes_R N$ an R-module structure, by the formula

$$r \sum_{k=1}^{n} m_k \otimes n_k = \sum_{k=1}^{n} rm_k \otimes n_k = \sum_{k=1} m_k \otimes rn_k.$$

Exercise: verify that this is well-defined, and makes $M \otimes_R N$ into an R-module. Furthermore, for any R-module A, there is a bijection

$$\begin{tikzcd}\big\{R\text{-module homomorphisms } \phi\colon M \otimes_R N \to A\big\} \arrow[r, leftrightarrow] & \big\{R\text{-bilinear maps } \beta\colon M \times N \to A\big\} \\ \phi \arrow[r, mapsto] & \phi \circ \alpha\end{tikzcd}$$

55. Computing tensor products

There is a lot to say about tensor products, but we don't have a lot of time. However, here are some simple recipes for computing tensor products of modules over a *commutative* ring R.

Proposition. *If M and N are free R-modules on bases $\{u_1, \ldots, u_m\}$ and $\{v_1, \ldots, v_n\}$ respectively, then $M \otimes_R N$ is a free R-module on the basis $\{u_i \otimes v_j\}_{i=1,\ldots,m,\, j=1,\ldots,n}$.*

Proof. Deduce this using the universal properties of tensor products and free modules: for any $m \times n$-tuple of elements (a_{ij}) in A, there is a unique R-bilinear map

$$\beta\colon M \times N \to A, \qquad \text{such that } \beta(u_i, v_j) = a_{ij}.$$

Thus $M \otimes_R N$ is a free module on the subset $\{u_i \otimes v_j\}$. □

Thus $R^m \otimes_R R^n \approx R^{mn}$.

Proposition. *If M and N are R-modules, generated by subsets S and T respectively, and $M' \subseteq M$ and $N' \subseteq N$ are submodules generated by subsets $U \subseteq M'$ and $V \subseteq N'$, then*

$$M/M' \otimes_R N/N' \approx (M \otimes_R N)/R\{\, s \otimes v,\ u \otimes t \mid s \in S,\, t \in T,\, u \in U,\, v \in V\,\}.$$

Proof. First show that R-bilinear maps $\beta\colon M/M' \times N/N' \to A$ are in bijective correspondence with R-bilinear maps $\widetilde{\beta}\colon M \times N \to A$ such that $\widetilde{\beta}(M', N) = 0 = \widetilde{\beta}(M, N')$. Then note that $\widetilde{\beta}(M', N) = 0$ iff $\widetilde{\beta}(S, V) = 0$, and $\widetilde{\beta}(M, N') = 0$ iff $\widetilde{\beta}(T, U) = 0$. □

Example. Suppose $M = R/(a)$ and $N = R/(b)$ are cyclic modules. Since $R \otimes_R R$ is free of rank 1 with basis $\{1 \otimes 1\}$, we find that

$$R/(a) \otimes_R R/(b) \approx R \otimes_R R/R\{a \otimes 1,\ 1 \otimes b\} \approx R/(a,b).$$

56. Torsion modules

We are now going to talk about modules over integral domains.

Let R be an integral domain.

An element $x \in M$ in an R-module is **torsion** if there exists a non-zero $r \in R$ such that $rx = 0$. We say a module M is **torsion** if $M_{\text{tors}} = M$ and is **torsionfree** if $M_{\text{tors}} = \{0\}$.

torsion
torsion
torsionfree

Lemma. *The collection $M_{\text{tors}} \subseteq M$ of torsion elements is a submodule. The quotient module M/M_{tors} is torsionfree.*

Proof. Clearly $0 \in M_{\text{tors}}$. If $x, y \in M_{\text{tors}}$ and $r \in R$, then there are $a, b \in R \setminus \{0\}$ such that $ax = 0 = by$, whence $(ab)(x+y) = 0$ with $ab \neq 0$ and $a(rx) = 0$ with $a \neq 0$.

If there exist $\overline{x} \in M/M_{\text{tors}}$ and $a \in R \setminus \{0\}$ such that $a\overline{x} = 0$, then for any lift $x \in M$ of x we have $ax \in M_{\text{tors}}$, whence there exists $b \in R \setminus \{0\}$ such that $bax = 0$. Since $ba \neq 0$ we have that $x \in M_{\text{tors}}$, so $\overline{x} = 0$. □

The following will often be useful.

Proposition. *Given a submodule $N \subseteq M$, the quotient module M/N is torsion iff for all $x \in M$ there exists $c \in R \setminus \{0\}$ such that $cx \in N$.*

Proof. It's enough to show that an element $\overline{x} \in M/N$ is a torsion element iff for some $x \in M$ with $\overline{x} = x + N$, there exists $c \in R \setminus \{0\}$ such that $cx \in N$. But this is immediate. □

Example. A cyclic module of the form R/I is torsion iff $I \neq 0$. If there exists $a \in I \setminus \{0\}$, then for any $b \in R$ we have $ab \in I$, whence R/I is torsion by the previous proposition. Conversely, if $I = 0$ then $R/I = R$, and $1 \in R$ is certainly not a torsion element.

Example. If F is a field and V an F-vector space, then $V_{\text{tors}} = \{0\}$, so every vector space is torsionfree, and V is a torsion module iff $V = 0$.

57. R-linear independence

Let R be an integral domain, and M an R-module. We are going to define a notion of "linear dependence" which generalizes that for vector spaces.

Say that an indexed collection $(x_i \in M)_{i \in I}$ is **R-linearly dependent** (or just **R-dependent**) if there exists an indexed collection $(r_i \in R)_{i \in I}$ with $0 < |\{\, i \in I \mid r_i \neq 0 \,\}| < \infty$ and $\sum_i r_i x_i = 0$. Otherwise the collection is **R-linearly independent**, or just **R-independent**.

R-linearly dependent
R-dependent
R-linearly independent
R-independent

In particular, $(x_i)_{i \in I}$ is R-independent iff for all distinct $i_1, \dots, i_n \in I$, we have that $a_1 x_1 + \cdots + a_n x_n = 0$ with $a_k \in R$ implies $a_k = 0$ for $k = 1, \dots, n$.

Note: elements of an R-independent collection $(x_i)_{i \in I}$ are always pairwise distinct (using that $1 \neq 0$ in R). Thus in practice we often speak only of the *R-independent subset* $\{\, x_i \mid i \in I \,\} \subseteq M$.

Lemma. *A subset $S \subseteq M$ is R-independent iff the submodule $N = RS$ generated by S is free, with S a free basis of N.*

Proof. Immediate. □

Example. A module M with subset $S \subseteq M$ is free with basis S iff (i) S is R-independent and (ii) $M = RS$.

Example. If M is a torsion module, then the only R-independent subset is the empty set.

The collection of R-independent subsets $S \subseteq M$ is ordered by $\subseteq$. Say that an R-independent subset $S \subseteq M$ is **maximally R-independent** if it is maximal with respect to this ordering, i.e., if whenever $S \subseteq T \subseteq M$ with T an R-independent subset, then $S = T$.

maximally R-independent

Example. The basis of a free module is always maximally R-independent.

Warning: a maximally R-independent set need not be a basis, even if the module is free, and can exist if the module is not free.

Here is a handy criterion for maximally independent sets.

Lemma. *An R-independent subset $S \subseteq M$ is maximal iff M/N is a torsion module where $N = RS$.*

Proof. Let $y \in M$, with image $\overline{y} = y + N \in M/N$. Observe $\overline{y}$ is a torsion element in M/N if and only if there exists

$$by = a_1x_1 + \cdots + a_nx_n, \qquad x_1, \dots, x_n \in S \text{ pairwise distinct}, \qquad a_1, \dots, a_n, b \in R, \quad b \neq 0.$$

Thus $\overline{y}$ is torsion iff either $y \in S$ or $S \cup \{y\}$ is R-dependent (since any R-dependence among $y, x_1, \dots, x_n$ must involve $b \neq 0$, since S is itself R-independent). So all $\overline{y}$ are torsion iff S is maximally R-independent. □

Proposition. *Every module over an integral domain admits a maximal R-independent set.*

Proof. Let $\mathcal{P}$ be the collection of all R-independent subsets of M, ordered by the subset relation. Note that

- $\mathcal{P}$ is non-empty, since $\varnothing \subseteq M$ is R-independent.
- If $\mathcal{C} \subseteq \mathcal{P}$ is a non-empty chain, then $T := \bigcup_{S \in \mathcal{C}} S$ is R-independent, since this is a condition verified on finite subsets, all of which are a subset of some $S \in \mathcal{C}$ because $\mathcal{C}$ is a chain.

Thus we can apply Zorn's lemma to $\mathcal{P}$, giving a maximal R-independent set S. □

Corollary. *Every vector space V over a field F has a basis.*

Proof. There exists a maximally independent subset $S \subseteq V$, whence V/FS is torsion, so $V/FS = 0$ since F is a field and thus $V = FS$. □

58. Rank of modules over integral domains

M 24 Oct

Let R be an integral domain and M an R-module. We define an invariant of R-modules called the *rank*, which generalizes the notion of dimension of a vector space. However, unlike for vector spaces, rank is not a complete isomorphism invariant. I will concentrate on the case of finite rank.

The **rank** of a module M over an integral domain is defined to be the largest size of any R-independent subset.

rank

Proposition (Interchange lemma)**.** *Let R be an integral domain, and M an R-module. Suppose we have sequences of elements $v_1, \dots, v_m$, $w_1, \dots, w_n$ in M such that*

- *$v_1, \dots, v_m$ is R-independent, and*
- *$M/R\{w_1, \dots, w_n\}$ is a torsion module.*

Then

1. *$m \leq n$, and*
2. *after reordering $w_1, \dots, w_n$, we have that $M/R\{v_1, \dots, v_m, w_{m+1}, \dots, w_n\}$ is a torsion module.*

Proof. We prove, by induction on $k \in \{0, \dots, m\}$, that we have

(a) $k \leq n$, and
(b) $M/R\{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ is torsion (after possibly reordering the w_js).

If $k = 0$ this is immediate.

Suppose the k case is true for some $k \in \{0, \dots, m-1\}$. By (b) there exists

$$bv_{k+1} = a_1v_1 + \cdots + a_kv_k + a_{k+1}w_{k+1} + \cdots + a_nw_n, \qquad a_1, \dots, a_n, b \in R, \quad b \neq 0.$$

Since $v_1, \dots, v_{k+1}$ is R-independent, we must have $k < n$ and at least one one of the $a_{k+1}, \dots, a_n$ must be non-zero. Reorder $w_{k+1}, \dots, w_n$ so that $a_{k+1} \neq 0$. We thus have

$$a_{k+1}w_{k+1} = a_1v_1 + \cdots + a_kv_k + (-b)v_{k+1} + a_{k+2}w_{k+2} + \cdots + a_nw_n, \qquad a_{k+1} \neq 0.$$

I claim that $M/R\{v_1, \dots, v_{k+1}, w_{k+2}, \dots, w_m\}$ is torsion. Suppose $\overline{x}$ is an element of this quotient, represented by some $x \in M$. Then by (b) there exists

$$dx = c_1v_1 + \cdots + c_kv_k + c_{k+1}w_{k+1} + \cdots + c_nw_n, \qquad c_1, \dots, c_n, d \in R, \quad d \neq 0.$$

Then

$$\begin{aligned} a_{k+1}dx &= a_{k+1}c_1v_1 + \cdots + a_{k+1}c_kv_k + \underbrace{a_{k+1}c_{k+1}w_{k+1}} + a_{k+1}c_{k+2}w_{k+2} + \cdots + a_{k+1}c_nw_n \\ &= (a_{k+1}c_1 + c_{k+1}a_1)v_1 + \cdots + (a_{k+1}c_k + c_{k+1}a_k)v_k \\ &\qquad + (-c_{k+1}b)w_{k+1} + (a_{k+1}c_{k+2} + c_{k+1}a_{k+2})w_{k+2} + \cdots + (a_{k+1}c_n + c_{k+1}a_n)w_n, \end{aligned}$$

where we substitute in the previous expression for $a_{k+1}w_{k+1}$. Since $a_{k+1}d \neq 0$, this implies $\overline{x}$ is a torsion element. □

Proposition. *Let $S \subseteq M$ be a finite subset of size n such that M/RS is torsion. Then there exists a maximal R-independent subset of size $m \leq n$, and every maximal R-indpendent subset of M has size m.*

Proof. By the interchange lemma applied with $S = \{w_1, \dots, w_n\}$, every finite R-independent subset of M has size $m \leq n$, whence there are no infinite R-independent subsets.

Now apply the interchange lemma to two maximal R-independent subsets $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$, giving that $m \leq n$ for any pair of such subsets, and thus by symmetry that $m = n$. □

We call this m the **rank** of M. **rank**

59. Properties of rank of modules

Example. Every torsion module has rank 0.

Example. The free module R^n has rank n, since its basis is clearly an R-independent spanning set. As a consequence, we get *invariance of rank* for finitely generated free modules over an integral domain: if $R^n \approx R^m$, then $n = m$.

Note: this can fail if R is not an integral domain.

Example. If F is a field and V an F-vector space which admits a finite spanning set of size m, then it admits a basis of some size $n \leq m$, and every basis has size n.

Finally, we can add ranks.

Proposition. *Let R be an integral domain, M an R-module with $N \subseteq M$ a submodule. If N has finite rank n, and M/N has finite rank m, then M has finite rank $m+n$.*

In particular, if A, B are modules of finite rank, then $\operatorname{rank}(A \oplus B) = \operatorname{rank} A + \operatorname{rank} B$.

Proof. Pick maximal R-independent subsets $y_1, \dots, y_n$ of N, and $\overline{x}_1, \dots, \overline{x}_m$ of M/N. Choose elements $x_k \in M$ such that $x_k + N = \overline{x}_k$. I claim that $x_1, \dots, x_m, y_1, \dots, y_n$ is a maximal R-independent subset of M.

R-independence. Suppose

$$a_1x_1 + \cdots a_mx_m + b_1y_1 + \cdots b_ny_n = 0, \qquad a_1, \dots, a_m, b_1, \dots, b_n \in R.$$

Then $a_1\overline{x}_1 + \cdots + a_m\overline{x}_m = 0$, whence all $a_i = 0$ since the $\overline{x}_i$ are R-independent in M/N. So $b_1y_1 + \cdots + b_ny_n = 0$, so all $b_j = 0$ since the y_j are R-independent in N.

Maximality. Suppose $z \in M$, and let $\overline{z} = z + N \in M/N$. Then there exists

$$c\overline{z} = a_1\overline{x}_1 + \cdots + a_m\overline{x}_m, \qquad a_1, \ldots, a_m, c \in R, \quad c \neq 0.$$

Thus $u := cz - (a_1x_1 + \cdots + a_nx_n) \in N$, so there exists

$$du = b_1y_1 + \cdots + b_ny_n, \qquad b_1, \ldots, b_n, d \in R, \quad d \neq 0.$$

From this we have $(dc)z \in R\{x_1, \ldots, x_m, y_1, \ldots, y_n\}$ with $dc \neq 0$. □

Remark. There is another way to talk about all this, using tensor products to reduce to dimension theory for vector spaces. Let $F = \mathrm{Frac}(R)$ be the fraction field. Then for any R-module M the tensor product $V := F \otimes_R M$ is naturally an F-module. It turns out that $\mathrm{rank}_R M = \dim_F V$.

60. Annihilator ideals and cyclic modules

Let R be a ring with 1 (but not assumed to be commutative). Given a (left) R-module M, the **annihilator** of M is the subset

annihilator

$$\mathrm{Ann}(M) := \{\, x \in R \mid xM = 0 \,\} = \{\, x \in R \mid xm = 0 \text{ for all } m \in M \,\}$$

of the ring R.

Proposition. *$\mathrm{Ann}(M)$ is a two-sided ideal in R.*

Proof. Straightforward: $0M = 0$, if $xM = 0 = yM$, then $(x + y)M = 0$, if $xM = 0$ then $xrM = 0$ for all $r \in R$, and also $rxM = 0$ for all $r \in R$. □

Proposition. *If $M \approx N$ are isomorphic left R-modules, then $\mathrm{Ann}(M) = \mathrm{Ann}(N)$.*

Proof. Clear: if $\phi\colon M \to N$ is an isomorphism, then $rm = 0$ for all $m \in M$ implies $rn = 0$ for all $n \in N$ since $n = \phi(m)$ for some m, and conversely using $\phi^{-1}\colon N \to M$. □

Example. Suppose $I \subseteq R$ is a 2-sided ideal. Then $\mathrm{Ann}(R/I) = I$. To see this, note (i) if $x(R/I) = 0$, then in particular $x1 \in I$, so $x \in I$, and (ii) if $x \in I$, then for any $y \in R$ we have $xy \in I$ since I is also a right ideal.

Remark. If $I \subseteq R$ is merely a left ideal, then $\mathrm{Ann}(R/I) \subseteq I$, but they are not necessarily equal. For instance, consider $R = M_{2\times 2}(F)$ with F a field, and let I be the subset of all matrices whose second column is 0. Then $\mathrm{Ann}(R/I) = 0$ but $I \neq 0$.

Proposition. *Let R be ring, and $I, J \subseteq R$ 2-sided ideals. Then $R/I \approx R/J$ as left R-modules iff $I = J$.*

Proof. If $I = J$ then R/I and R/J are identical modules, and so are certainly isomorphic. Conversely, if there is a left module isomorphism $R/I \approx R/J$, then

$$I = \mathrm{Ann}(R/I) = \mathrm{Ann}(R/J) = J.$$

□

Example. This can fail if the ideals are not 2-sided ideals. For instance, consider the left ideals $I, J \subseteq R = M_{2\times 2}(F)$, consisting of matrices whose first or second column is 0 respectively. Then $I \neq J$ but $R/I \approx R/J$ as modules.

If R is commutative, then this implies that R/I and R/J are isomorphic modules iff $I = J$.

61. Classification of modules over a PID

Let R be a PID. We are going to classify finitely generated modules over R. In the case of $R = \mathbb{Z}$, this will give the classification of finitely generated abelian groups. In the case of $R = F[x]$, this will lead to theorems on canonical forms of linear operators.

Since the PID R is commutative ring, we have that $R/(a) \approx R/(b)$ as modules iff $(a) = (b)$, i.e., iff a and b are the same up-to-units.

The cyclic modules for a domain come in three types.

- The *trivial* cyclic module $R/(a) = \{0\}$, when $a \in R^\times$.
- The *nontrivial torsion* cyclic module $R/(a)$, when $a \neq 0$ and $a \notin R^\times$.
- The *free* cyclic module $R = R/(0)$.

Here is the key fact.

Proposition. *Every finitely generated module over a PID is isomorphic to a finite direct sum of cyclic modules.*

Remark. This is certainly false for general rings. For instance, suppose $R = \mathbb{Z}[x]$ and consider the submodule $M = R\{2, x\} \subseteq R$ (which is also the ideal $(2, x)$). Then M is not a cyclic module (exactly because it is not a principal ideal), but it is also not isomorphic to a direct sum of any two non-trivial modules. (See exercise.)

Remark. This fails for infinitely generated modules over a PID. For instance, as a $\mathbb{Z}$-module, $M = \mathbb{Q}$ is not a direct sum of cyclic modules.

We can sharpen the above to a classification of finitely generated modules over a PID.

Theorem (Modules over a PID: Invariant factor form)**.** *Let R be a PID, and M a finitely generated R-module.*

- *There exists $t \geq 0$ and a chain of proper ideals $R \supsetneq (a_1) \supseteq \cdots \supseteq (a_t)$ such that*
$$M \approx R/(a_1) \oplus \cdots \oplus R/(a_t).$$
- *The number t and the sequence $(a_1), \dots, (a_t)$ of ideals are unique, in the sense that if also $M \approx R/(a'_1) \oplus \cdots \oplus R/(a'_{t'})$ with $R \supsetneq (a'_1) \supseteq \cdots \supseteq (a'_{t'})$, then $t = t'$ and $(a_k) = (a'_k)$ for all k.*

Remark. Write $t = s+r$ with $0 \leq s, r \leq t$ where $(a_1), \dots, (a_s) \neq (0)$ and $(a_{s+1}) = \cdots = (a_{s+r}) = (0)$. Then this becomes
$$M \approx R/(a_1) \oplus \cdots \oplus R/(a_s) \oplus R^r,$$
where each $R/(a_1), \dots, R/(a_s)$ is a torsion cyclic module, and $\operatorname{rank} M = r$. This is how the invariant factor decomposition is usually presented.

The ideals $(a_1), \dots, (a_s)$ are called **invariant factors**, and $r = \operatorname{rank} M$. **invariant factors**

Example. If $R = \mathbb{Z}$, then we get the invariant factor classification of finitely generated abelian groups:
$$M \approx \mathbb{Z}/(a_1) \oplus \cdots \oplus \mathbb{Z}/(a_s) \oplus \mathbb{Z}^r, \qquad a_1 \mid \cdots \mid a_s.$$

There is another form of this classification.

Theorem (Modules over a PID: Elementary divisor form)**.** *Let R be a PID, and M a finitely generated R-module.*

- *There exist $r, u \geq 0$, and a sequence of elements $p_1^{k_1}, \dots, p_u^{k_u} \in R$ (not necessarily distinct) with p_i prime and $k_i \geq 1$, such that*
$$M \approx R^r \oplus R/(p_1^{k_1}) \oplus \cdots \oplus R/(p_u^{k_u}).$$

- *The numbers r and u are unique, and the sequence $p_1^{k_1}, \dots, p_u^{k_u}$ is unique up to reordering and units, in the sense that if also $M \approx R^{r'} \oplus R/(q_1^{\ell_1}) \oplus \cdots \oplus R/(q_{u'}^{\ell_{u'}})$, then $r = r'$, $u = u'$, and the sequence $q_1^{\ell_1}, \dots, q_u^{\ell_u}$ is the same as $p_1^{k_1}, \dots, p_u^{k_u}$ up to reordering and units.*

Remark. In the elementary divisor form, we also have $r = \operatorname{rank} M$. The list $p_1^{k_1}, \dots, p_u^{k_u}$ are called **elementary divisors**.

elementary divisors

Example. If $R = \mathbb{Z}$, then we get the elementary divisor classification of finitely generated abelian groups:

$$M \approx \mathbb{Z}^r \oplus \mathbb{Z}/(p_1^{k_1}) \oplus \cdots \oplus \mathbb{Z}/(p_u^{k_u}),$$

each p_i a prime number and each $k_i \geq 1$.

62. Existence of invariant factor form 1

W 26 Oct

We will proceed by proving in sequence: Existence of invariant factor form ⇒ Existence of elementary divisor form ⇒ Uniqueness of elementary divisor form ⇒ Uniqueness of invariant factor form.

The idea will be to observe note that any finitely generated module is isomorphic to a quotient M/N where M is a finite rank free module.

Proposition. *Let M be a free R-module of rank m, and $N \subseteq M$ a submodule. Then*

1. *N is a free R-module of some rank $n \leq m$, and*
2. *there exists*
   - *a free basis $x_1, \dots, x_m$ of M, and*
   - *elements $a_1, \dots, a_n \in R$ with $(a_1) \supseteq \cdots \supseteq (a_n) \supsetneq (0)$, such that*
   - *$y_1 = a_1 x_1, \dots, y_n = a_n x_n$ is a free basis of N.*

This gives the *existence* of invariant factor decomposition. If A is a finitely generated module, produce a surjective homomorphism $\phi \colon M \twoheadrightarrow A$ from some finite rank free module. Then $A \approx M/N$ with $N = \operatorname{Ker} \phi$. By the proposition, we get that

$$A \approx M/N \approx Rx_1/Ra_1x_1 \;\oplus\; \cdots \;\oplus\; Rx_n/Ra_nx_n \;\oplus\; Rx_{n+1} \;\oplus\; \cdots \;\oplus\; Rx_m.$$

We have $Rx_k/Ra_kx_k \approx R/(a_k)$ for $k = 1, \dots, n$, and $Rx_k \approx R/(0) = R$ if $k = n+1, \dots, m$. Note that it is possible that $(a_1) = \cdots = (a_p) = R$ for some $p > 0$, in which case $R/(a_k) \approx 0$ for $k \leq p$. If so, we can remove those factors and reindex, so that $t = m - p$.

63. Existence of invariant factor form 2

Now we prove the proposition.

Proof of proposition, Part 1. We are going to work by induction on rank of N. Note that the rank of N is in fact finite and $\leq m$: any R-independent subset of N is also an R-independent subset of M, so $\operatorname{rank} N \leq \operatorname{rank} M$.

If $\operatorname{rank} N = 0$ then N is torsion. But the only torsion element in the free module M is 0, so $N = 0$.

So we suppose $\operatorname{rank} N = n > 0$, whence $N \neq 0$.

We are going to find an R-module homomorphism $\phi_1 \colon M \to R$, an element $x_1 \in M$, and an element $a_1 \in R$, such that

(a) $a_1 x_1 \in N$ and $\phi_1(x_1) = 1$,
(b) $\phi_1(N) = (a_1) \neq (0)$, and
(c) If $\phi \in \operatorname{Hom}_R(M, R)$, then $(a_1) \subseteq \phi(N)$ implies $(a_1) = \phi(N)$.

Now we produce ϕ_1, x_1, a_1 and prove (a),(b),(c). Let

$$\Sigma_N = \{\, \phi(N) \mid \phi \in \operatorname{Hom}_R(M,R) \,\},$$

a set of ideals in R, which can be ordered by the subset relation. Note that if $N \neq 0$, then Σ_N contains at least one non-trivial ideal. Since M is free, there exists an isomorphism $M \xrightarrow{\sim} R^m$, with component functions $\pi_i \colon M \to R$. Thus any non-zero element $x \in M$ will have $\pi_i(x) \neq 0$ for some i, so there exists i such that $\pi_i(N) \neq (0)$.

We choose ϕ_1 so that $\phi_1(N)$ is maximal in the poset Σ_N, and a_1 so that $\phi_1(N) = (a_1)$, and $(a_1) \neq (0)$ since (0) cannot be maximal. We have proved (b). Maximality of ϕ_1 is precisely (c).

Pick $y_1 \in N$ such that $\phi_1(y_1) = a_1$. I will show that for any $\phi \in \operatorname{Hom}_R(M,R)$, we have that $\phi(y_1) \in (a_1)$. Since R is a PID, there exists

$$(d) = (a_1, \phi(y_1)) = (\phi_1(y_1), \phi(y_1)), \qquad d \in R.$$

Thus there exists

$$d = c_1\phi_1(y_1) + c_2\phi(y_1), \qquad c_1, c_2 \in R.$$

Let $\phi' := c_1\phi_1 + c_2\phi \in \operatorname{Hom}_R(M,R)$, whence $d = \phi'(y_1)$. But $\phi'(N) \supseteq (d) \supseteq \phi(N)$, so maximality of ϕ implies $\phi'(N) = \phi(N)$, so $(d) = (a_1)$ and thus $\phi(y_1) \in (a_1)$.

Apply this to the projection homomorphisms $\pi_1, \dots, \pi_m \colon M \to R$ associated to some choice of free basis $e_1, \dots, e_m$ of M. Then $\pi_i(y_1) = c_i a_i$ for some $c_i \in R$. Set $x_1 := c_1e_1 + \cdots + c_me_m$, whence $a_1x_1 = y_1$. Since $a_1\phi(x_1) = \phi_1(y_1) = a_1$, we must have $\phi(x_1) = 1$ since $a_1 \neq 0$. We have proved (a). □

64. Existence of invariant factor form 3

Proof of proposition, part 2. Having produced ϕ_1, x_1, a_1 satisfying (a),(b),(c), we prove the following.

(d) $M = Rx_1 \oplus M'$, where $M' = \operatorname{Ker}(\phi)$.
(e) $N = Ra_1x_1 \oplus N'$, where $N' = N \cap \operatorname{Ker}(\phi)$.
(f) $\operatorname{rank} N' = n - 1$.
(g) for all $\phi \in \operatorname{Hom}_R(M,R)$ we have $\phi(N') \subseteq (a_1)$.

Claim (d) is clear: $Rx_1 \cap \operatorname{Ker}(\phi_1) = 0$ since $\phi(rx_1) = r$, while if $x \in M$, then $x - \phi_1(x)x_1 \in \operatorname{Ker}(\phi_1)$, using (a). Similarly for (e): $Ra_1x_1 \cap \operatorname{Ker}(\phi_1) = 0$, and if $y \in N$, then $\phi_1(y) = ca_1$ for some $c \in R$ using (b), so $y - \phi_1(y)x_1 = y - c(a_1x_1) \in \operatorname{Ker}(\phi_1)$. To prove (f): since $\phi_1(a_1x_1) = a_1 \neq 0$ by (b), we have $\operatorname{rank}(Ra_1x_1) = 1$, so $\operatorname{rank} N = 1 + \operatorname{rank} N'$.

Finally, to prove (g): Since we have a direct sum decomposition $M = Rx_1 \oplus M'$, given a homomorphism $\phi \colon M \to R$, we can define a new homomorphism $\phi' \colon M \to R$ by

$$\phi'(cx_1 + m') := c + \phi(m'), \qquad c \in R, \quad m' \in M'.$$

Since $N = Ra_1x_1 \oplus N'$, we compute compute

$$\phi'(N) = (a_1) + \phi'(N') \supseteq (a_1), \qquad \phi'(N') = \phi(N').$$

By (c) this first statement implies $\phi'(N) = (a_1)$, and therefore

$$\phi(N') = \phi'(N') \subseteq \phi'(N) = (a_1),$$

as desired.

Now we show (1) that N is free, by induction on $n = \operatorname{rank} N$. If $\operatorname{rank} N > 0$, then the above argument gives us a decomposition $N = Ra_1x_1 \oplus N'$, with $\operatorname{rank} Ra_1x_1 = 1$ and $\operatorname{rank} N' = n - 1$. Induction on rank gives that N' is free, and hence N is free. Because $N \subseteq M$ we already know $\operatorname{rank} N \leq \operatorname{rank} M$.

Now we prove (2), by induction on $n = \operatorname{rank} N$. If $\operatorname{rank} N > 0$, then the above argument gives us submodules $N' \subseteq M' \subseteq M$. By (1) already proved we see that M' is free, and we know $\operatorname{rank} N' = n - 1$. Thus by induction we have

- a free basis $x_2, \dots, x_m$ of M', and
- elements $a_2, \dots, a_n \in R$ with $(a_2) \supseteq \cdots \supseteq (a_n) \supsetneq (0)$, such that
- $y_2 = a_2 x_2, \dots, y_n = a_n x_n$ is a free basis of N'.

Combining this with $x_1 \in M$ and $a_1 \in R$ as in (d) and (e), we obtain a free basis $x_1, \dots, x_m$ of M, and a free basis $a_1 x_1, \dots, a_n x_n$ of N. It remains to show that $(a_1) \supseteq (a_2)$.

Define $\phi\colon M \to R$ so that $\phi(x_2) = 1$ and $\phi(x_i) = 0$ if $i \neq 2$. Then

$$\phi(N') = \phi(Ra_2x_2 + \cdots + Ra_nx_n) = (a_2).$$

By (g), we conclude that $(a_2) = \phi(N') \subseteq (a_1)$, as desired.

□

65. Existence of elementary divisor form

Earlier, we proved a general "Chinese remainder theorem" for commutative rings:

> If $A_1, \dots, A_n \subseteq R$ are ideals which are pairwise comaximal ($A_i + A_j = R$ if $i \neq j$), and $A = A_1 \cdots A_n$ is the product ideal, then there is an isomorphism of rings
>
> $$R/A \xrightarrow{\sim} R/A_1 \times \cdots \times R/A_n,$$
>
> defined by $x + A \mapsto (x + A_1, \dots, x + a_n)$.

A feature of this is that this isomorphism is *also* an isomorphism of R-modules.

Now suppose R is a PID (and thus a UFD). Given $a \in R \smallsetminus \{0\}$ which is not a unit, it has a prime factorization

$$a = p_1^{k_1} \cdots p_r^{k_r}, \qquad p_i \text{ primes, distinct up to units}, \quad k_i \geq 1.$$

For $i \neq j$, we see that $p_i^{k_i}, p_j^{k_j}$ are relatively prime, so $(p_i^{k_i}) + (p_j^{k_j}) = (p_i^{k_i}, p_j^{k_j}) = R$ is the unit ideal. Thus we can apply the CRT and obtain an isomorphism of rings and of R-modules:

$$R/(a) \approx R/(p_1^{k_1}) \oplus \cdots \oplus R/(p_r^{k_r}).$$

This is a primary divisor decomposition of the torsion cyclic module $R/(a)$.

For a general finitely generated module M we have an invariant factor decomposition:

$$M \approx R^r \oplus R/(a_1) \oplus \cdots \oplus R/(a_s), \qquad R \supsetneq (a_1) \subseteq \cdots \supseteq (a_s) \supsetneq (0).$$

We can replace each $R/(a_k)$ with a primary divisor decomposition. This gives a primary divisor decomposition for M.

66. Modules and quotient rings

Let R be a commutative ring R, and M an R-module. Suppose $I \subseteq R$ is an ideal such that **F 28 Oct**
$I \subseteq \operatorname{Ann}(M)$. That is, $IM = 0$, or more concretely, $am = 0$ for all $a \in I$ and $m \in M$.

Then M admits the structure of an R/I-module, defined so that

$$(r + I)m := rm.$$

It is clear this is well-defined, exactly because $rm = 0$ if $r \in R$.

Furthermore, if $M \approx N$ as R-modules, and if $IM = 0$, then also $IN = 0$ and the isomorphism is also an isomorphism of R/I-modules.

Example. Let $R = \mathbb{Z}$ and M a $\mathbb{Z}$-module (=abelian group), such that $(p)M = 0$ for some prime number p. (Note: we usually write $pM = (p)M$ when we have a principal ideal.) In other words, every element of the abelian group M has order either 1 or p.

Then M is naturally a module over $\mathbb{F}_p = \mathbb{Z}/p$, i.e., it is an $\mathbb{F}_p$-vector space. This has a numerical invariant, which is its dimension $\dim_{\mathbb{Z}/p} M$. By the above remarks, this is an isomorphism invariant of abelian groups M for which $pM = 0$. That is, if $M \approx N$ and $pM = 0$, then $\dim_{\mathbb{Z}/p} M = \dim_{\mathbb{Z}/p} N$.

Submodules IM and quotient modules M/IM. Let $I \subseteq R$ be an ideal in a commutative ring R. Given any R-module I, we can form

$$IM := \{\, a_1m_1 + \cdots a_km_k \mid a_i \in I,\ m_i \in M,\ k \geq 0 \,\}.$$

This is a submodule of M.

We thus have a quotient module M/IM. Note that the ideal I annihilates M/IM by construction: $a(x + IM) = ax + IM \in IM$ if $a \in I$. Therefore the R-module structure on M/IM descends to an R/I-module structure.

Proposition.

1. *If $\phi\colon M \to N$ is an isomorphism of R-modules, then ϕ restricts to an isomorphism $IM \to IN$ of submodules. It further induces an isomorphism $M/IM \to N/IN$ on quotient modules, which is an isomorphism of R/I-modules.*
2. *If $M = M_1 \oplus \cdots \oplus M_n$ is an internal direct sum decomposition of an R-module, then $IM = IM_1 \oplus \cdots \oplus IM_n$, and thus $M/IM \approx M/IM_1 \oplus \cdots \oplus M/IM_n$ as R/I-modules.*
3. *If M is a finitely generated R-module, then and M/IM is a finitely generated as an R-module, and as an R/I-module.*
4. *If M is a finitely generated R-module, and $I \subseteq R$ is a finitely generated ideal, then IM is a finitely generated R-module.*

Proof. Exercise. □

67. Uniqueness of decompositions

Let R be a PID. For any prime p in R and $k \geq 1$, I will constuct an isomorphism invariant $\beta_{p^k}(M) \in \mathbb{Z}_{\geq 0}$ of finitely generated R-modules. Then I will show that, for any elementary divisor decomposition of M, the number $\beta_{p^k}(M) =$ the number of elementary divisors in that decomposition which are equal to p^k up-to-units. Since the definition of $\beta_{p^k}(M)$ makes no reference to the decomposition, this will immediately imply uniqueness of the lists of elementary divisors.

We can combine this with the invariant $\operatorname{rank}(M) \in \mathbb{Z}_{\geq 0}$, which is also an isomorphism invariant and counts the number of free factors in any elementary divisor decomposition. Together these demonstrate the uniqueness of elementary divisor decompositions.

Let R be a PID, and $p \in R$ a prime, and consider a finitely generated module M. Note that $p^{k+1}M = p(p^kM) \subseteq p^kM$. Thus we obatin a chain of submodules

$$M = p^0M \supseteq p^1M \supseteq p^2M \supseteq \cdots,$$

each of which is also finitely generated. We therefore get quotients

$$M/pM, \qquad pM/p^2M, \qquad p^2M/p^3M, \quad \ldots,$$

each of which is a finitely generated R/p-module. We can also write these as $p^{k-1}M/p^kM = N/pN$, where $N = p^{k-1}M$.

Note that since p is irreducible, R/p is a field. For $k \geq 1$ define

$$\alpha_{p^k}(M) := \dim_{R/p} p^{k-1}M/p^kM.$$

Proposition.

1. *The function α_{p^k} is an isomorphism invariant of finitely generated R-modules.*
2. *If $M \approx M_1 \oplus \cdots \oplus M_n$, then $\alpha_{p^k}(M) = \alpha_{p^k}(M_1) + \cdots + \alpha_{p^k}(M_n)$.*
3. *If $M \approx R/(a)$ for some $a \in R$, then*

$$\alpha_{p^k}(M) = \begin{cases} 1 & \text{if } p^k \mid a, \\ 0 & \text{if } p^k \nmid a. \end{cases}$$

In particular, when $a = 0$ this says that $\alpha_{p^k}(R) = 1$.

Proof. Statement (1) is a consequence of part (1) of the previous proposition, and the fact that if $M \approx N$, then $p^{k-1}M \approx p^{k-1}N$, and therefore $p^{k-1}M/p(p^{k-1}M) \approx p^{k-1}N/p(p^{k-1}N)$. Statement (2) is a immediate from part (2) of the previous proposition.

For statement (3), let $M = R/(a)$ and $N = p^{k-1}M$, so that $\alpha_{p^k}(M) = \dim_{R/p}(N/pN)$. First I'll describe N.

Consider any $b \in R$ with $b \neq 0$, and suppose $N = bM = b(R/(a))$. We have

$$N = b(R/(a)) = (a, b)/(a),$$

where $(a, b) \subseteq R$ is the submodule (also ideal) generated by a and b. The point is that $\overline{r} \in b(R/(a))$ iff $\overline{r} = bx + Ra$ for some $x \in R$, iff $r = bx + ay$ for some $x, y \in R$.

In a PID, we have that $(a, b) = (d)$ where d is a GCD of a and b. Then we can write $a = da'$ for some $a \in R$, and we hhave an isomorphism of R-module.

$$\mathbb{R}/(a') \xrightarrow{\sim} (d)/(a), \qquad r \mapsto rd.$$

The point is that multiplication by d gives a surjective function $R \to (d)$, and the preimage of (a) under this function is (a'). Thus $N \approx R/(a')$ where $a' = a/\gcd(a, b)$.

Now suppose $b = p^{k-1}$, and consider $M = R/(a)$ where $a = p^t m$ with $t \geq 0$ and $p \nmid m$. We want to compute

$$\dim_{R/p}(p^{k-1}M/p^k M) = \dim_{R/p}(N/pN)$$

where $N = p^{k-1}M$. Writing $N = R/(a')$, we have

$$\dim_{R/p}(N/pN) = \begin{cases} 1 & \text{if } p \mid a', \\ 0 & \text{if } p \nmid a'. \end{cases}$$

By the above discussion, we have

$$a' = \frac{a}{\gcd(a, p^{k-1})} = \frac{p^t m}{\gcd(p^t m, p^{k-1})} = \frac{p^t m}{p^{\max(t,k-1)}}.$$

Thus $p \mid a'$ iff $t \geq k$ iff $p^k \mid a$, which is what we wanted. □

As a consequence, if

$$M \approx R^r \oplus R/(a_1) \oplus \cdots \oplus R/(a_m), \qquad a_k \in R \setminus \{0\},$$

we have

$$\alpha_{p^k}(M) = r + \text{number of } j \in \{1, \dots, m\} \text{ such that } p^k \mid a_j.$$

Now define

$$\beta_{p^k}(M) = \alpha_{p^k}(M) - \alpha_{p^{k+1}}(M).$$

Then for the above M, we have

$$\beta_{p^k}(M) = \text{number of } j \in \{1, \dots, m\} \text{ such that } p^k \mid a_j \text{ and } p^{k+1} \nmid a_j.$$

By construction, α_{p^k} and thus β_{p^k} are isomorphism invariants, and we have shown that, for any elementary divisor decomposition

$$M \approx R^r \oplus R/(p_1^{k_1}) \oplus \cdots \oplus R/(p_u^{k_u}),$$

we have that $\beta_{p^k}(M) =$ the number of elementary divisors in $p_1^{k_1}, \dots, p_u^{k_u}$ which are the same as p^k up-to-units. This proves uniqueness of elementary divisor decompostions.

Now suppose

$$M \approx R^r \oplus R/(a_1) \oplus \cdots \oplus R/(a_s), \qquad R \supsetneq (a_1) \supseteq \cdots \supseteq (a_s) \supsetneq (0)$$

is an invariant factor decomposition. By an exercise I will leave to you, it turns out that

$$s = \max\{\, \alpha_p(M) \mid p \in R \text{ prime element} \,\} - r, \qquad r = \operatorname{rank}(M),$$

and

$$(a_j) = (p_1^{c_{1,j}} \cdots p_d^{c_{r,j}}),$$

where $p_1, \dots, p_d$ are all primes (distinc up-to-units) such that $\alpha_p(M) > r$, and

$$c_{i,j} = \text{the integer } k \text{ such that } \alpha_{p_i^{k+1}}(M) < r + s + 1 - j \leq \alpha_{p_i^k}(M).$$

Thus the primes p_i and numbers r, s, and $c_{i,j}$ are determined by the values of $\alpha_{p^k}(M)$ and $\operatorname{rank}(M)$. This proves uniqueness of invariant factor decompositions.

68. Canonical form for linear maps

Recall that given a **linear operator**, i.e., a pair $(V, T\colon V \to V)$, where V is an F-vector space and T is an F-linear operator on V, we can give V the structure of an $R = F[x]$-module, so that

linear operator

$$fv := f(T)v, \qquad f \in F[x], \quad v \in V.$$

I'm going to write V_T for this $F[x]$-module.

Conversely, every $F[x]$-module M is of the form V_T for some (V, T), where V the underlying F-vector space of the module M (so $V = M$ as an abelian group), and T is defined by $T(v) := xv$. So $F[x]$-modules are really the same as F-linear operators.

There is a dictionary for going back and forth between properties of the linear operator (V, T) and properties of the $F[x]$-module V_T.

- Submodules of V_T correspond to *T-invariant subspaces*, i.e., vector subspaces $W \subseteq V$ such that $T(W) \subseteq W$.
- Homomorphisms $\phi\colon V_T \to W_U$ of $F[x]$-modules correspond to linear maps which *interwine* U and V, i.e., linear maps $\phi\colon V \to W$ such that $\phi \circ T = U \circ \phi$.
- V_T and V_U are isomorphic as $F[x]$-modules iff the linear operators T and U are *similar*, i.e., if there exists a linear isomorphism $\phi\colon V \to V$ such that $U = \phi \circ T \circ \phi^{-1}$.
- Given (V, T) the space V is finite dimensional over F if and only if V_T is finitely generated and torsion as an $F[x]$-module.

  To see this, note that if V_T contains a non-torsion element v, then it contains a free cyclic submodule $Rv \subseteq V_T$, and $\dim_F Rv$ is infinite. Conversely, if V_T is finitely generated and torsion, then the classification theorem gives an isomorphism $V_T \approx F[x]/(f_1) \oplus \cdots F[x]/(f_m)$ for some polynomials f_k. If some $f_k = 0$ then $\dim_F V_T$ is infinte, while if all $f_k \neq 0$ then $\dim_F V_T = \deg(f_1) + \cdots + \deg(f_m)$.

Given (V, T) finite dimensional, consider the annihilator ideal $\operatorname{Ann}(V_T) = (f) \subseteq F[x]$. By the classification theorem we can write $V_T \approx \bigoplus_{k=1}^m R/(f_k)$ for some nonzero f_k, and therefore $0 \neq f_1 \cdots f_m \in \operatorname{Ann}(V_T)$, whence $f \neq 0$. We usually assume f is monic, in which case we call f the **minimal polynomial** of T.

minimal polynomial

Proposition. *Consider (V, T) with V finite dimensional, and $f =$ the minimal polynomial of T. For any $c \in V$ TFAE.*

1. *There exists $v \in V$ with $v \neq 0$ such that $Tv = cv$. That is, c is an eigenvalue of T.*
2. *$f(c) = 0$.*

Proof. By the division algorithm we have $f = (x - c)g + r = g \cdot (x - c) + r$ with $g \in F[x]$ and $r \in F$.

(1) $\Longrightarrow$ (2). Given $v \neq 0$ with $Tv = cv$, we have $(x - c)v = 0$, and so

$$0 = f(T)v = g(T)(T - cI)v + rv = rv,$$

whence $r = 0$.

(2) $\Longrightarrow$ (1). If $f(c) = 0$ then $r = 0$, so $f = (x - c)g$. Since $g \notin (f) = \operatorname{Ann}(V_T)$, there exists $w \in V$ such that $v := g(T)w \neq 0$. Therefore

$$0 = f(T)w = (T - cI)g(T)w = (T - cI)v,$$

so $v \neq 0$ and $Tv = cv$. □

Given any $F[x]$-module decomposition

$$V_T \approx M_1 \oplus \cdots \oplus M_n = F[x]/(f_1) \oplus \cdots \oplus F[x]/(f_m),$$

we can give a block matrix representation of T of the form

$$\left(\begin{array}{c|c|c|c} B_1 & & & \\ \hline & B_2 & & \\ \hline & & \ddots & \\ \hline & & & B_m \end{array}\right)$$

by choosing an F-basis $e_1, \ldots, e_n$ of V, so that the first batch of basis elements are in M_1, the second batch in M_2, and so on. I'll describe some choices for cyclic modules.

Given $V_T = F[x]/(f)$ with $f = x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$ a monic polynomial over F, we can use the basis

$$e_1 = \overline{1}, \quad e_2 = \overline{x}, \quad \ldots, \quad e_i = \overline{x}^{i-1}, \quad \ldots, \quad e_k = \overline{x}^{k-1}.$$

Then the matrix describing the operator T in this basis is the $k \times k$ **companion matrix**

companion matrix

$$C_f = \begin{pmatrix} 0 & 0 & \ldots & \ldots & 0 & -b_0 \\ 1 & 0 & \ldots & \ldots & 0 & -b_1 \\ 0 & 1 & \ldots & \ldots & 0 & -b_2 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & 1 & -b_{k-1} \end{pmatrix}$$

A matrix is in **rational canonical form** if it is a diagonal block matrix whose non-trivial blocks are companion matrices $C_{f_1}, \ldots, C_{f_m}$ for non-constant monic polynomials f_k such that $f_1 \mid f_2 \mid \cdots \mid f_m$.

rational canonical form

Theorem (Rational canonical form)**.** *Given an operator (V, T) on a finite dimensional vector space, there exists a basis with respect to which the matrix A of T is in rational canonical form. Furthermore, the rational canonical form matrix is unique.*

Example. If $F = \mathbb{R}$ and T is the operator on $V = \mathbb{R}^2$ given by left multiplication by $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, then the rational canonical form is $A = \begin{bmatrix} 0 & -1 \\ 1 & 2\cos\theta \end{bmatrix}$.

Note that the characteristic polynomial of the companion matrix is

$$\det(xI - C_f) = f(x),$$

and thus if $V_T \approx \bigoplus_{k=1}^m F[x]/(f_k)$ with f_k monic, then the characteristic polynomial of T is

$$\det(xI - T) = f_1(x) \cdots f_m(x).$$

If f is the minimal polynomial of T then $f_1 \cdots f_m \in \operatorname{Ann}(V_T) = (f)$.

Corollary (Cayley-Hamilton)**.** *For any linear operator T on a finite dimensional space, its minimal polynomial divides its characteristic polynomial.*

If $V_T = F[x]/(x-c)^k$, then in terms of the basis

$$e_1 = (\overline{x} - c)^{k-1}, \quad e_2 = (\overline{x} - c)^{k-2}, \quad \ldots, \quad e_{k-1} = \overline{x} - c, \quad e_k = 1,$$

the matrix describing T is the $k \times k$ **Jordan matrix**

Jordan matrix

$$J_k(c) := \begin{pmatrix} c & 1 & 0 & \dots & \dots & 0 & 0 \\ 0 & c & 1 & \dots & \dots & 0 & 0 \\ 0 & 0 & c & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 1 & 0 \\ 0 & 0 & 0 & \dots & \dots & c & 1 \\ 0 & 0 & 0 & \dots & \dots & 0 & c \end{pmatrix}$$

Thus for an operator T whose characteristic (or minimal) polynomial is a product of linear factors in $F[x]$, the elementary divisors of T will all have the form $(x - c_i)^{k_i}$ with $c_i \in F$ and $k_i \geq 1$, in which case there exists a basis such that T is represented in **Jordan canonical form**, i.e., as a diagonal block matrix whose blocks are Jordan matrices, and which is unique up to reordering the Jordan blocks.

Jordan canonical form

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL
Email address: rezk@math.uiuc.edu