PMATH 347 – RING THEORY LECTURES

R. WILLARD

Contents

20. Definition of a ring

Definition. A **ring** is an ordered triple $(R, +, \cdot)$ where

- R is a non-empty set;
- $+$ and $\cdot$ are binary operations on R;

which jointly satisfy the following conditions:

(i) $(R, +)$ is an abelian group;
(ii) $\cdot$ is associative;
(iii) There exists $1 \in R$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in R$.
(iv) (Distributive laws): for all $a, b, c \in R$,

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c)$$
$$a \cdot (b + c) = (a \cdot b) + (a \cdot c).$$

Notation/jargon.

- We denote $(R, +, \cdot)$ by R.
- The identity element of $(R, +)$ is denoted 0.
- The inverse of a in the group $(R, +)$ is denoted $-a$, and is called the *additive* inverse.
- We write $a - b$ for $a + (-b)$.
- The element 1 is called the *multiplicative identity*. It is (provably) unique.
- We usually write ab instead of $a \cdot b$.
- We say that R is *commutative* if it satisfies $ab = ba$ for all $a, b \in R$, and is *noncommutative* otherwise.

Example 20.1.

(1) $\mathbb{Z}$ (with usual addition and multiplication) is a commutative ring; it is the prototypical example of a commutative ring.
(2) $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are also commutative rings.
(3) For every $n \geq 2$, the set $M_n(\mathbb{R})$ of all $n \times n$ matrices over $\mathbb{R}$ (with matrix addition and multiplication) is a noncommutative ring. Similarly, $M_n(\mathbb{Z})$, $M_n(\mathbb{Q})$, $M_n(\mathbb{C})$.
(4) $\mathbb{Z}_n$ is a (finite) commutative ring for every $n \geq 2$.
(5) Let $C(\mathbb{R})$ be the set of all continuous, everywhere-defined functions $f : \mathbb{R} \to \mathbb{R}$ (a very big set). Define $f + g$ and $f \cdot g$ pointwise; that is,

$$\begin{aligned} (f + g)(x) &:= f(x) + g(x) \\ (f \cdot g)(x) &:= f(x) \cdot g(x). \end{aligned}$$

Then $(C(\mathbb{R}), +, \cdot)$ is a commutative ring. What is its zero element? Its identity element?
(6) Is $(C(\mathbb{R}), +, \circ)$ a ring? ($\circ$ means composition of functions.)
(7) Given any ring R, let $\mathcal{F}(R)$ denote the set of *all* functions $f : R \to R$. We can turn $\mathcal{F}(R)$ into a ring by defining $+$ and $\cdot$ pointwise as in $C(\mathbb{R})$ (using the

definitions of $+$ and $\cdot$ in R). It can be shown that $\mathcal{F}(R)$ with these operations is a ring. What is its zero element? Its identity element?

Warning: In general, you cannot assume that $\cdot$ satisfies left or right cancellation. For example in $\mathbb{Z}$ we have $0 \cdot x = 0 \cdot y$, but that does not imply $x = y$. In some rings, even if $a \neq 0$, one cannot assume that $a \cdot b = a \cdot c$ implies $b = c$.

Proposition 20.2. *Let R be a ring. Then*

(1) $0a = a0 = 0$ *for all* $a \in R$.
(2) $-a = (-1)a = a(-1)$ *for all* $a \in A$.
(3) $(-a)b = a(-b) = -(ab)$ *for all* $a, b \in R$.
(4) $(-a)(-b) = ab$.

Proof.

(1) $0+0 = 0$, so $0a = (0+0)a = (0a)+(0a)$. Hence $0a+0 = 0a+0a$, so cancelling (in the group $(R, +)$) gives $0a = 0$. Similar proof works for $a0 = 0$.
(2) $1 + (-1) = 0$. Hence $0 = 0a = (1 + (-1))a = (1a) + (-1)a = a + (-1)a$. Hence $a + (-a) = a + (-1)a$, so cancelling gives $(-1)a = -a$. Similar proof works for $a(-1) = -a$.
(3) $(-a)b = (a(-1))b = a((-1)b) = a(-b)$ by (2). Also, $(-a)b = ((-1)a)b = (-1)(ab) = -(ab)$ by (2).
(4) Exercise. □

Definition. Let R be a ring.

(1) An element $a \in R$ is a *unit* if there exists $b \in R$ satisfying $ab = ba = 1$. (We also say that a is *invertible*. b is called the *inverse* of a and is denoted a^{-1}; it is provably unique.)
(2) $R^\times$ denotes the set of units in R.

Remark. 2 is a unit in $\mathbb{Q}$ but is not a unit in $\mathbb{Z}$. $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\}$ while $\mathbb{Z}^\times = \{1, -1\}$.

In general, $(R^\times, \cdot)$ is a group; called the *group of units* of R.

Can $0 = 1$ in a ring? If $0 = 1$, then $a = a1 = a0 = 0$ for all $a \in R$, i.e., $R = \{0\}$. A 1-element ring is called *trivial*. Thus a ring is nontrivial iff it satisfies $0 \neq 1$.

Definition.

(1) A *division ring* is a ring D satisfying $0 \neq 1$ and $D^\times = D \setminus \{0\}$ (i.e., every nonzero element is a unit).
(2) A *field* is a commutative division ring.

Example 20.3.

(1) $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}_p$ (p prime) are fields.
(2) We will see an example of a noncommutative division ring in the next lecture.

21. Integral domains, subrings

Notation. Let R be a ring and $a \in R$.

(1) For $n > 1$ we let na denote $\underbrace{(a + a + \cdots + a)}_{n}$.
(2) For $n \geq 1$ we let $(-n)a$ denote $-(na)$. (Thus na is defined for all $n \in \mathbb{Z}$.)
(3) $\mathbb{Z}a = \{na : n \in Z\}$.

Note that $\mathbb{Z}a$ is the cyclic subgroup of $(R, +)$ generated by a.

Recall:

Definition.

(1) A *division ring* is a ring D satisfying $0 \neq 1$ and $D^\times = D \setminus \{0\}$ (i.e., every nonzero element is a unit).
(2) A *field* is a commutative division ring.

Is $M_2(\mathbb{R})$ a division ring?

Example 21.1. $\mathbb{H}$, the *ring of real Hamiltonion quaternions*, is the set of all expressions $a + bi + dj + dk$ where $a, b, c, d \in \mathbb{R}$ and i, j, k are primitive symbols.

(1) Addition is defined obviously:

$$(a + bi + cj + dk) + (a' + b'i + c'j + d'k) = (a + a') + (b + b')i + (c + c')j + (d + d')k$$

(2) Multiplication is first defined on the primitive symbols:

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

(3) Then multiplication is extended to expressions by assuming $ai = ia$, $aj = ja$, $ak = ka$ for all $a \in \mathbb{R}$, and assuming distributivity.

It can be shown that $\mathbb{H}$ is a ring. If $a + bi + cj + dk \neq 0$ then

$$(a + bi + cj + dk)^{-1} = \frac{a}{e} - \frac{b}{e}i - \frac{c}{e}j - \frac{d}{e}k$$

where $e = a^2 + b^2 + c^2 + d^2$, so $\mathbb{H}$ is a division ring. $\mathbb{H}$ is not a field (as $ij \neq ji$).

Definition. Let R be a ring. A *zero divisor* is an element $a \in R$ such that

(1) $a \neq 0$, and
(2) There exists $b \in R$ with $b \neq 0$ such that $ab = 0$ or $ba = 0$ (or both).

Example 21.2.

(1) $\mathbb{Z}$ has no zero divisors (since $a, b \neq 0$ imply $ab = ba \neq 0$).
(2) 2, 3, 4 are zero divisors in $\mathbb{Z}_6$, since $2 \cdot 3 = 4 \cdot 3 = 0$.
(3) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is a zero divisor in $M_2(\mathbb{R})$, since $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Proposition 21.3. *Suppose R is a ring and $a \in R$ with $a \neq 0$. If a is not a zero divisor, then we can "multiplicatively cancel by a." That is, for all $b, c \in R$,*

$$ab = ac \implies b = c$$
$$ba = ca \implies b = c.$$

Proof. Assume $ab = ac$. Then $a(b-c) = a(b+(-c)) = ab + a(-c) = (ab) + -(ac) = ab - ac = ab - ab = 0$. Since $a \neq 0$ and a is not a zero divisor, it must be that $b - c = 0$, i.e., $b = c$. The other implication is proved similarly. □

Lemma 21.4. *If R is a ring and $a \in R^\times$, then a is not a zero divisor. Hence we can always "multiplicatively cancel by units."*

Proof. Argue by contradiction. Assume $a \in R^\times$ and a is a zero divisor. Thus a^{-1} exists in R, and there exists $b \in R$ with $b \neq 0$ such that either $ab = 0$ or $ba = 0$. Suppose $ba = 0$; then $b = b1 = b(aa^{-1}) = (ba)a^{-1} = 0a^{-1} = 0$, contradiction. The equation $ab = 0$ also leads to a contradiction. □

Definition. A ring R is called an *integral domain* (or *domain*) if it is commutative, satisfies $0 \neq 1$, and has no zero divisors.

For example, $\mathbb{Z}$ is an integral domain.

Corollary 21.5. *Every field is an integral domain.*

Proof. Follows from the previous lemma. □

Definition. Suppose R is a ring. A *subring* of R is a subset $S \subseteq R$ such that

(1) S is a subgroup of $(R, +)$.
(2) S is closed under multiplication (i.e., $a, b \in S$ implies $ab \in S$).
(3) $1 \in S$.

Write $S \leq R$ to denote that S is a subring of R.

As was the case for groups, every subring of a ring is itself a ring (with operations inherited from the larger ring).

Example 21.6.

(1) $\mathbb{Z} \leq \mathbb{Q}$. $\mathbb{R} \leq \mathbb{C}$.
(2) $\mathbb{Z}_n \not\leq \mathbb{Z}$.
(3) Is $M_2(\mathbb{R})$ a subring of $M_3(\mathbb{R})$? (No.)
(4) Is $M_2(\mathbb{Z})$ a subring of $M_2(\mathbb{R})$? (Yes.)
(5) Recall $C(\mathbb{R})$ from Example 20.1, the ring of all *continuous* functions $\mathbb{R} \to \mathbb{R}$, and $\mathcal{F}(\mathbb{R})$, the ring of *all* functions $\mathbb{R} \to \mathbb{R}$.
   (a) $C(\mathbb{R}) \leq \mathcal{F}(\mathbb{R})$.
   (b) Define $P(\mathbb{R})$ to be the set of all *polynomial* functions in $C(\mathbb{R})$. Then $P(\mathbb{R}) \leq C(\mathbb{R})$.

22. Polynomial rings

Let R be a ring. Let x be a formal variable.

- A **polynomial in x over R** is an expression
$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$
where $n \geq 0$, $a_0, \ldots, a_n \in R$, and if $n > 0$ then $a_n \neq 0$.
- We also denote this by $\sum_{i=0}^{n} a_i x^i$.
- Note that if $n = 0$ then the expression is just a_0. When $n = 0$ and $a_0 = 0$ the expression is just 0. (This is the **zero polynomial**.)
- An expression $a_i x^i$ is called a **term** of the polynomial.
- The elements $a_0, a_1, \ldots, a_n$ of R are called the **coefficients** of the polynomial.
- The **degree** of the polynomial is n, except for the zero polynomial which has no degree.
- If the polynomial is not 0, then the **leading term** is $a_n x^n$, and the **leading coefficient** is a_n.
- By definition, two polynomials are equal iff they have the same degree and the same coefficients.

Definition. $R[x]$ denotes the set of all polynomials in x over R.

Key fact: every $p(x) \in R[x]$ can be viewed as a **formula** which defines a **function** $p : R \to R$. However, the polynomial is **not** the same thing as the function it defines.

- For example, let $R = \mathbb{Z}_2$. Put $p(x) = x^2 + x$ and $q(x) = x^3 + x$. (More precisely, $p(x) = 1x^2 + 1x + 0$ and $q(x) = 1x^3 + 0x^2 + 1x + 0$.) As polynomials, $p(x)$ and $q(x)$ are not equal. As functions $\mathbb{Z}_2 \to \mathbb{Z}_2$, they are identical (in fact, they are both the constant 0 function).

Formally, we have a mapping $R[x] \to P(R)$ sending $p(x) \mapsto p$ (i.e., sending each polynomial to its corresponding polynomial function). This mapping is surjective but may fail to be injective, as we have just seen.

We use the usual shortcuts when writing polynomials. For example, we often do not bother to write the terms of the form $0x^i$. The abbreviated expressions are sometimes called **sparse polynomials**. For example, the expression $2x^3 + 3$ is technically a sparse polynomial which represents the "real" polynomial $2x^3 + 0x^2 + 0x + 3$. We also usually simplify a term of the form $1x^i$ or $(-1)x^i$ to just x^i or $-x^i$ respectively.

For convenience, we also define **sloppy polynomials over R** to be **all** expressions of the form $\sum_{i=0}^{n} a_i x^i$ $(a_0, \ldots, a_n \in R)$, allowing the last coefficient a_n to be 0. When discussing sloppy polynomials we talk about "the highest index n," rather than the "degree." Of course every sloppy polynomial determines a unique "real" polynomial by trimming (deleting) zero terms of highest index if needed. **Warning:** no one in the world except me (and now you) uses the term "sloppy polynomials."

Definition. Given a ring R, define $+$ and $\cdot$ on $R[x]$ "in the obvious way." That is, given $p(x), q(x) \in R[x]$:

(1) To define $p(x)+q(x)$:
Write $p(x)$ and $q(x)$ as sloppy polynomials with the same highest index and use
$$\left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} b_i x^i\right) = \sum_{i=0}^{n} (a_i + b_i) x^i.$$
(2) To define $p(x) \cdot q(x)$:
Write $p(x) = \sum_{i=0}^{m} a_i x^i$ and $q(x) = \sum_{j=0}^{n} b_j x^j$. Then

$$\begin{aligned}\left(\sum_{i=0}^{m} a_i x^i\right) \cdot \left(\sum_{j=0}^{n} b_j x^j\right) &= (a_0 + a_1 x + a_2 x^2 + \cdots) \cdot (b_0 + b_1 x + b_2 x^2 + \cdots) \\ &= (a_0 b_0) + (a_0 b_1 + a_1 b_0) x + (a_0 b_2 + a_1 b_1 + a_2 b_0) x^2 + \cdots \\ &= \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i b_j\right) x^k.\end{aligned}$$

The above formulas for addition and multiplication are very useful in proofs. In practice, though, they can be misleading, because the right-hand side expressions can be sloppy and so might need to be trimmed. For example:

(1) If $R = \mathbb{R}$ and $p(x) = 2x^2 + 3x + 1$ and $q(x) = (-2)x^2 + 4x + 3$, then
$$\begin{aligned}p(x) + q(x) &= (2x^2 + 3x + 1) + ((-2)x^2 + 4x + 3) \\ &= (2 + (-2)x^2 + (3+4)x + (1+3) \quad \text{(by the formula)} \\ &= 0x^2 + 7x + 4 \\ &= 7x + 4 \quad \text{(trimmed)}.\end{aligned}$$
(2) If $R = \mathbb{Z}_4$ and $p(x) = q(x) = 2x^2 + x + 1$, then
$$\begin{aligned}p(x) \cdot q(x) &= (2x^2 + 1x + 1) \cdot (2x^2 + 1x + 1) \\ &= (2{\cdot}2)x^4 + (2{\cdot}1 + 1{\cdot}2)x^3 + (2{\cdot}1 + 1{\cdot}1 + 1{\cdot}2)x^2 + (1{\cdot}1 + 1{\cdot}1)x + (1{\cdot}1) \\ &= 0x^4 + 0x^3 + 1x^2 + 2x + 1 \\ &= x^2 + 2x + 1 \qquad \text{(trimmed)}.\end{aligned}$$

Theorem 22.1. *$R[x]$ is a ring containing R as a subring.*

Proof sketch. A nightmare. To illustrate, I will prove that $\cdot$ is associative. Let $p(x) = \sum_i a_i x^i, q(x) = \sum_j b_j x^j$, and $r(x) = \sum_k c_k x^k$. Then $p(x) \cdot q(x) = \sum_s d_s x^s$ where
$$d_s = \sum_{i+j=s} a_i b_j,$$
so $(p(x) \cdot q(x)) \cdot r(x) = \sum_t e_t x^t$ where
$$e_t = \sum_{s+k=t} d_s c_k = \sum_{s+k=t} \left(\sum_{i+j=s} a_i b_j\right) c_k = \sum_{i+j+k=t} a_i b_j c_k.$$

It can similarly be proved that $p(x) \cdot (q(x) \cdot r(x))$ is represented by the same sloppy polynomial. □

The next theorem describes a property of the functions defined by polynomials.

Theorem 22.2. *Suppose $q(x), r(x) \in R[x]$ and let $p(x) = q(x) \cdot r(x)$. If R is commutative, then $p(c) = q(c) \cdot r(c)$ for all $c \in R$.*

Proof sketch. Write $q(x) = \sum_i a_i x^i$ and $r(x) = \sum_j b_j x^j$. Then

$$\begin{aligned} q(c) \cdot r(c) &= (a_0 + a_1 c + a_2 c^2 + \cdots) \cdot (b_0 + b_1 c + b_2 c^2 + \cdots) \\ &= a_0 b_0 + a_0(b_1 c) + a_0(b_2 c^2) + a_0(b_3 c^3) + \cdots \\ &\qquad + (a_1 c) b_0 + (a_1 c)(b_1 c) + (a_1 c)(b_2 c^2) + \cdots \\ &\qquad\qquad + (a_2 c^2) b_0 + (a_2 c^2)(b_1 c) + \cdots \end{aligned}$$

If R is commutative, then the terms can be rearranged to get

$$a_0 b_0 + (a_0 b_1 + a_1 b_0) c + (a_0 b_2 + a_1 b_1 + a_2 b_0) c^2 + \cdots \;=\; p(c). \qquad \square$$

We can generalize this as follows. Given a ring R, its **center** is the set $Z(R) = \{a \in R : ab = ba \text{ for all } b \in R\}$. **Fact:** $Z(R)$ is a subring of R (exercise).

Corollary 22.3. *Suppose $q(x), r(x) \in R[x]$ and let $p(x) = q(x) \cdot r(x)$. Then $p(c) = q(c) \cdot r(c)$ for all $c \in Z(R)$.*

23. Homomorphisms, ideals

Definition. Let R, S be rings. A function $\varphi : R \to S$ is a **homomorphism** (of rings) if

(1) $\varphi(a+b) = \varphi(a) + \varphi(b)$ for all $a, b \in R$.
(2) $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.
(3) $\varphi(1_R) = 1_S$.

Example 23.1.

(1) $\mathbb{Z} \to \mathbb{Z}_n$, $k \mapsto k \pmod n$
(2) If R is a ring and $c \in Z(R)$, then $\varphi_c : R[x] \to R$ given by $\varphi_c(p(x)) = p(c)$. Called "evaluation at c."

    We saw yesterday that φ_c preserves multiplication if $c \in Z(R)$. Preserving addition is easy (exercise). If $p(x)$ is the constant polynomial 1, then $p(c) = 1$ so $\varphi_c(1) = 1$.

Suppose $\varphi : R \to S$ is a ring homomorphism. Then it is automatically a group homomorphism $\varphi : (R, +) \to (S, +)$. Hence it has a kernel,

$$\ker(\varphi) = \{a \in R : \varphi(a) = 0_S\}.$$

Furthermore, φ is injective iff $\ker(\varphi) = \{0_R\}$.

Definition. As in the case of groups,

(1) An **isomorphism** is a bijective homomorphism.
(2) Write $R \cong S$ if there exists an isomorphism from R to S.

Definition. Let R be a ring and $I \subseteq R$.

(1) I is a **left ideal** of R if
    (a) I is a subgroup of $(R, +)$.
    (b) If $r \in R$ and $a \in I$, then $ra \in I$.
(2) Right deals are defined dually ($a \in I, r \in R \implies ar \in I$).
(3) I is an **ideal** if it is both a left and right ideal.

Proposition 23.2. *If I is an ideal of R and $1 \in I$, then $I = R$.*

Proof. For every every $r \in R$ we have $r \in R, 1 \in I \implies r1 = r \in I$, so $R \subseteq I$, so $R = I$. □

Proposition 23.3. *Let R, S be rings and $\varphi : R \to S$ a homomorphism.*

(1) $\mathrm{im}(\varphi)$ *is a subring of* S.
(2) $\ker(\varphi)$ *is an ideal of* R.

Proof. (1) is routine. Focus on (2). We already know that $\ker(\varphi)$ is a (normal) subgroup of $(R, +)$. Suppose $a \in \ker(\varphi)$ and $r \in R$. Then $\varphi(ra) = \varphi(r)\varphi(a) = \varphi(r) \cdot 0_S = 0_S$, proving $ra \in \ker(\varphi)$. A similar proof shows $ar \in \ker(\varphi)$. □

Suppose I is an ideal of R. Then I is a (normal) subgroup of the group $(R,+)$, so we may form the quotient group $(R,+)/I$. Its elements are the cosets of I, which we write additively as $a+I$. The group operation is the usual

$$(a+I)+(b+I)=\{c+d\,:\,c\in a+I \text{ and } d\in b+I\}$$

and is characterized by the rule $(a+I)+(b+I)=(a+b)+I$.

Claim. *The rule $(a+I)\cdot(b+I):=(ab)+I$ defines an operation $\cdot$ on R/I.*

Proof. We must show that the rule is well-defined. Suppose $a+I=a'+I$ and $b+I=b'+I$, so $a-a'\in I$ and $b-b'\in I$. We must show $(ab)+I=(a'b')+I$, and to do that it suffices to show $ab-a'b'\in I$. Well

$$\begin{aligned} ab-a'b' &= ab-a'b+a'b-a'b' \\ &= (a-a')b+a'(b-b'). \end{aligned}$$

Since I is an ideal and $a-a', b-b'\in I$, the above expression is in I as required. □

Warning: $(a+I)\cdot(b+I)$ does **not** equal the set $\{c\cdot d\,:\,c\in a+I \text{ and } d\in b+I\}$.

Claim. *If R is a ring and I is an ideal, then $(R/I,+,\cdot)$ is a ring.*

Proof sketch. We already know that $(R/I,+)$ is a group. It remains to show that $+$ is commutative in R/I, $\cdot$ is associative, $1+I$ is a multiplicative identity element, and the distributive laws hold in R/I. All of these facts can be quickly deduced from the corresponding facts in R. For example, to prove that $\cdot$ is associative in R/I, observe that for any $a,b,c\in R$,

$$\begin{aligned} (a+I)\cdot((b+I)\cdot(c+I)) &= (a+I)\cdot(bc+I) \\ &= a(bc)+I \qquad \text{(definition of } \cdot \text{ in } R/I) \\ &= (ab)c+I \qquad \text{as } a(bc)=(ab)c \text{ in } R \\ &= \cdots \\ &= ((a+I)\cdot(b+I))\cdot(c+I). \end{aligned}$$

□

We call $(R/I,+,\cdot)$ a **quotient ring** and denote it R/I.

Theorem 23.4 (First Isomorphism Theorem for rings)**.** *Suppose R,S are rings and $\varphi:R\to S$ is a surjective homomorphism. Then $R/\ker(\varphi)\cong S$.*

Proof sketch. Just like the proof for groups. □

24. Characteristic; Principal ideals

An important example of a subring of any ring R is $\mathbb{Z}1 = \{n1 : n \in \mathbb{Z}\}$. It is certainly a subgroup of $(R, +)$ and contains 1. Must check closure under products. Given $m1, n1 \in \mathbb{Z}1$, assume first that $m, n > 0$. Thus

$$m1 \cdot n1 = (\underbrace{1 + 1 + \cdots + 1}_{m}) \cdot (\underbrace{1 + 1 + \cdots + 1}_{n}).$$

Using distributivity repeatedly, one can prove that this equals $(mn)1$ so is in $\mathbb{Z}1$. (One also must check the cases where $m < 0$ or $n < 0$.) We call $\mathbb{Z}1$ the **prime subring** of R and denote it R_0. It is the smallest subring of R, contained in all other subrings.

Let R be a ring and define $\varphi : \mathbb{Z} \to R_0$ by $\varphi(n) = n1$. The same calculations that showed that R_0 is a subring of R also show that φ is a ring homomorphism. Obviously φ is surjective. So we can apply the First Isomorphism Theorem to get

$$\mathbb{Z}/\ker(\varphi) \cong R_0.$$

So we will know R_0 up to isomorphism as soon as we know $\ker(\varphi)$.

Since $\ker(\varphi)$ is a subgroup of $(\mathbb{Z}, +)$, it must equal $n\mathbb{Z}$ for some $n \geq 0$. If $n = 0$ then $\ker(\varphi) = \{0\}$ and so φ is injective. Thus in this case φ is an isomorphism and $R_0 \cong \mathbb{Z}$. If instead $n > 0$ then we get $R_0 \cong \mathbb{Z}/n\mathbb{Z}$.

Definition. Let R be a ring. The **characteristic** of R is the integer n in the previous discussion.

Definition. Let R be a ring and $a \in R$.

(1) $Ra = \{ra : r \in R\}$.
(2) $aR = \{ar : r \in R\}$.
(3) (a) denotes the smallest ideal of R containing a. (More precisely, (a) is the intersection of all ideals containing a.)

We call (a) the **principal ideal generated by** a.

Lemma 24.1. *Suppose R is a ring and $a \in R$.*

(1) *Ra is a left ideal. It is the smallest left ideal of R containing a.*
(2) *Similarly, aR is the smallest right ideal of R containing a.*
(3) $Ra \cup aR \subseteq (a)$.

Proof. (1) Obviously $Ra \neq \varnothing$. Suppose $ra, sa \in Ra$. Then $ra + sa = (r + s)a \in Ra$ and $-(ra) = (-r)a \in Ra$, so Ra is a subgroup of $(R, +)$. Clearly if $ra \in Ra$ and $s \in R$ then $s(ra) = (sr)a \in Ra$, proving Ra is a left ideal. Clearly $a = 1a$ so $a \in Ra$. Now suppose that I is any left ideal of R containing a. Since $a \in I$ and I is a left ideal, we get $ra \in I$ for all $r \in R$; hence $Ra \subseteq I$. This proves Ra is contained in every left ideal containing a, so is the smallest such left ideal.

(2) is proved similarly. (3) Since (a) is an ideal and $a \in (a)$, it follows that $ra, ar \in (a)$ for all $r \in R$. This proves $Ra \cup aR \subseteq (a)$. □

Note: If R is commutative, then $Ra = aR$ and Ra is an ideal of R containing a. Since (a) is by definition the smallest ideal containing a, we get $(a) \subseteq Ra$. We already know that $Ra \subseteq (a)$. Hence $\boxed{(a) = Ra = aR \text{ if } R \text{ is commutative}}$.

Example 24.2. Consider the ring $\mathbb{R}[x]$. Let $I = (x^2+1)$, the principal ideal generated by $x^2 + 1$. Thus

$$I = \{(x^2 + 1)q(x) : q(x) \in \mathbb{R}[x]\} = \{f(x) \in \mathbb{R}[x] : x^2 + 1 \text{ is a factor of } f(x)\}.$$

(In the expression $(x^2+1)q(x)$, (x^2+1) does NOT denote the ideal I; the parentheses are just being used to surround the factor of $x^2 + 1$. It will be your job to recognize when parentheses are being used as brackets and when they are being used to name a principal ideal.)

I is an ideal, so we can form the quotient ring $\mathbb{R}[x]/I$. What is this quotient ring isomorphic to? Take an arbitrary element, i.e. a coset $f(x)+I$. Divide $f(x)$ by x^2+1 to get a quotient $q(x)$ and remainder $r(x) = a + bx$. Then

$$f(x) = (x^2 + 1)q(x) + (a + bx),$$

Hence

$$\begin{aligned} f(x) + I &= [(x^2 + 1)q(x) + I] + [(a + bx) + I] \\ &= I + [(a + bx) + I] \qquad \text{because } x^2 + 1 \in I \\ &= (a + bx) + I \qquad \text{because } I \text{ is the zero element of } \mathbb{R}[x]/I. \end{aligned}$$

In other words, every coset of I can be expressed as $(a + bx) + I$ for some $a, b \in \mathbb{R}$. Hence

$$\mathbb{R}[x]/I = \{(a + bx) + I : a, b \in \mathbb{R}\}.$$

Let's explore how addition and multiplication work in $\mathbb{R}[x]/I$. Let $(a+bx)+I, (c+dx)+I$ be two elements of $R[x]/I$. Their sum is easily computed.

$$\begin{aligned} [(a + bx) + I] + [(c + dx) + I] &= [(a + bx) + (c + dx)] + I \\ &= [(a + c) + (b + d)x] + I. \end{aligned}$$

Similarly,

$$\begin{aligned} [(a + bx) + I] \cdot [(c + dx) + I] &= [(a + bx) \cdot (c + dx)] + I \\ &= [(ac) + (ad + bc)x + (bd)x^2] + I. \end{aligned}$$

We can simplify this last expression as follows. Since $x^2+1 \in I$ we get $x^2+I = -1+I$, so $(bd)x^2 + I = -bd + I$, so

$$\begin{aligned} [(a + bx) + I] \cdot [(c + dx) + I] &= [(ac) + (ad + bc)x - bd] + I \\ &= [(ac - bd) + (ad + bc)x] + I. \end{aligned}$$

This resembles multiplication in $\mathbb{C}$. We might conjecture that $\mathbb{R}[x]/I \cong \mathbb{C}$. To prove this conjecture, define $\varphi_i : \mathbb{R}[x] \to \mathbb{C}$ by $\varphi_i(p(x)) = p(i)$. φ_i is a homomorphism (see Example 23.1(2).) Obviously φ_i is surjective, since for any complex number $a + ib$ we have $a + ib = \varphi_i(a + bx)$. I claim that $\ker(\varphi_i) = I$. Indeed, if $f(x) \in \ker(\varphi_i)$, i.e., $f(i) = 0$, then both $i, -i$ are roots of $f(x)$ so $x^2 + 1$ is a factor of $f(x)$, meaning $f(x) \in (x^2 + 1) = I$. This proves $\ker(\varphi_i) \subseteq I$. For the converse inclusion, note that $\ker(\varphi_i)$ is an ideal which contains $x^2 + 1$ (obviously), so $(x^2 + 1) \subseteq \ker(\varphi_i)$ (since $(x^2 + 1)$ is contained in every ideal containing $x^2 + 1$). This proves $\ker(\varphi_i) = I$.

Now apply the First Isomorphism Theorem to get $\mathbb{R}[x]/I \cong \mathbb{C}$.

25. Maximal ideals

The ideals of a ring R are ordered by inclusion and hence form a partially ordered set (or *poset*). We can schematically draw this poset with R at the top, $\{0\}$ at the bottom, and other ideals in between.

Lemma 25.1. *Suppose I, J are ideals of R.*

(1) *$I \cap J$ is an ideal; it is the largest ideal of R contained in both I and J.*
(2) *$I + J := \{a + b : a \in I \text{ and } b \in J\}$ is the smallest ideal of R containing both I and J.*

Proof. (2) I, J are both (normal) subgroups of $(R, +)$, so $I+J$ is also a subgroup and it contains I and J. Suppose $a+b \in I+J$ and $r \in R$. Then $r(a+b) = ra+rb \in I+J$ and similarly $(a+b)r = ar+br \in I+J$, so $I+J$ is an ideal. We've already noted $I, J \subseteq I+J$. Suppose K is any other ideal with $I, J \subseteq K$. Then for all $a+b \in I+J$ we have $a, b \in K$ so $a+b \in K$, proving $I+J \subseteq K$. □

Definition. Let R be a ring.

(1) An ideal I is **proper** if $I \neq R$. (Equivalently, if $1 \notin I$.)
(2) If I, J are ideals, then J **properly contains** I if $I \subseteq J$ and $I \neq J$.
(3) I is a **maximal ideal** if (i) it is a proper ideal, and (ii) the only ideal properly containing it is R.

Proposition 25.2. *Suppose R is a commutative ring and I is an ideal. R/I is a field iff I is a maximal ideal.*

Proof. Throughout the proof, if $a \in R$ then $\overline{a}$ denotes $a + I \in R/I$. In particular, $\overline{0} = 0 + I$ is the zero of R/I and $\overline{1} = 1 + I$ is the multiplicative identity of R/I.

($\Rightarrow$) Assume R/I is a field. Then $\overline{0} \neq \overline{1}$, meaning $I \neq 1+I$, so $1 \notin I$, so I is proper. Suppose J is an ideal properly containing I. Pick $a \in J \backslash I$. Thus $a+I \neq I$, i.e., $\overline{a} \neq \overline{0}$. As R/I is a field, there exists $\overline{b} \in R/I$ such that $\overline{a} \cdot \overline{b} = \overline{1}$, i.e., $(a+I)(b+I) = 1+I$, so $1 = ab + c$ for some $c \in I$. As $a, c \in J$ and J is an ideal, we get $1 \in J$ so $J = R$. This proves I is maximal.

($\Leftarrow$) Suppose I is maximal. We run through the defining properties of being a field.

(1) R/I is commutative (because R is).
(2) $1 \notin I$ because I is proper, so $I \neq 1 + I$, so $\overline{0} \neq \overline{1}$.
(3) Let $\overline{a} \in R/I$ with $\overline{a} \neq \overline{0}$. (Thus $a \notin I$.) We must show that $\overline{a}$ has a multiplicative inverse in R/I. Recall that $(a) = Ra$. By hypothesis, $a \notin I$, but clearly $a \in (a) + I$, so $(a) + I$ properly contains I, so $(a) + I = R$. In particular, $1 \in (a) + I$, so there exists $r \in R$ and $c \in I$ such that $1 = ar + c$. Hence $1 + I = ar + I = (a + I)(r + I)$, meaning $\overline{1} = \overline{a} \cdot \overline{r}$, so $\overline{r}$ is a multiplicative inverse to $\overline{a}$. □

We would like a result similar to Proposition 25.2 but characterizing ideals I such that R/I is an integral domain. Assume R is commutative and I is an ideal; what

properties of I determine whether R/I is an integral domain? R/I is already commutative (because R is). Clearly we need I to be proper (so $\overline{0} \neq \overline{1}$). To achieve the condition of having no zero divisors, we need $\overline{a} \cdot \overline{b} = \overline{0}$ to imply $\overline{a} = 0$ or $\overline{b} = \overline{0}$.

- $\overline{a} \cdot \overline{b} = \overline{0}$ means $(a + I)(b + I) = I$, *i.e.*, $ab \in I$.
- $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$ means $a + I = I$ or $b + I = I$, i.e., $a \in I$ or $b \in I$.

Thus we need: $ab \in I$ implies $a \in I$ or $b \in I$.

Definition. Suppose R is a commutative ring. An ideal I of R is a *prime ideal* if it is proper and $ab \in I$ implies $a \in I$ or $b \in I$.

Proposition 25.3. *Suppose R is a commutative ring and I is an ideal. R/I is an integral domain iff I is a prime ideal.*

Proof. Proved in the earlier discussion. □

Corollary 25.4. *Every maximal ideal of a commutative ring is a prime ideal.*

Proof. Let I be a maximal ideal of the commutative ring R. Then R/I is a field by Proposition 25.2. Hence R/I is an integral domain (since every field is an integral domain). Thus I is a prime ideal by Proposition 25.3. □

The converse is not true. (Example: in $\mathbb{Z}, \{0\}$ is a prime ideal but is clearly not a maximal ideal.)

26. Zorn's Lemma

Proposition 26.1. *Let R be a ring. Every proper ideal of R is contained in a maximal ideal of R.*

Proof. Here is the idea of the proof. Let I be a proper ideal of R. define $I_0 = I$. If I_0 is maximal then we're done. Otherwise, there exists a proper ideal I_1 properly containing I_0. If I_1 is maximal, we're done, and if not, then there exists a proper ideal I_2 properly containing I_1. In this way we either reach a maximal ideal or we construct an infinite sequence of proper ideals:

$$I = I_0 \subset I_1 \subset I_2 \subset \cdots \subset I_n \subset \cdots$$

Define $I_\infty = \bigcup_{n=0}^{\infty} I_n$. We claim that I_∞ is a proper ideal of R.

- Suppose $a, b \in I_\infty$. Then there exists n with $a, b \in I_n$. so $a + b, -a \in I_n \subseteq I_\infty$.
- Suppose $a \in I_\infty$ and $r \in R$. Then $a \in I_n$ for some n. Hence $ra \in I_n \subseteq I_\infty$.
- Thus I_∞ is an ideal. To show it is proper, suppose instead that $I_\infty = R$. Then $1 \in I_\infty$. Hence $1 \in I_n$ for some n. But then I_n isn't proper, contradiction. Hence I_∞ is proper.

We continue the argument. If I_∞ is maximal we're done. Otherwise, there exists a proper ideal $I_{\infty+1}$ properly containing I_∞. Continue: either a maximal ideal $I_{\infty+n}$ is found, or we get another infinite sequence of proper ideals:

$$I_\infty \subset I_{\infty+1} \subset I_{\infty+2} \subset \cdots \subset I_{\infty+n} \subset \cdots$$

Define $I_{\infty+\infty} = \bigcup_{n=0}^{\infty} I_{\infty+n}$. Again this is a proper ideal.

The intuition is that this cannot go on forever. To prove it, we must clarify what we mean by "forever." This is the job of set theory; for example, countable sequences (no matter how many times applied) do not capture "forever." We can "sweep this under the carpet" by a trick from set theory.

Definition. A **chain of proper ideals** is set S of proper ideals with the property that for all $I, J \in S$, either $I \subseteq J$ or $J \subseteq I$. (Note: S can be uncountable.)

By a similar argument as above, if S is a chain of proper ideals, then $\bigcup_{I \in S} I$ is still a proper ideal. In particular, for every chain of proper ideals there exists a proper ideal (namely, the union of the chain) containing all the elements of the chain.

Now let $\mathcal{I}(R)$ be the set of all proper ideals of R. The relation $\subseteq$ is a partial ordering of $\mathcal{I}(R)$ (reflective, antisymmetric, transitive). We have proved that every chain in $(\mathcal{I}(R), \subseteq)$ has an upper bound in $\mathcal{I}(R)$.

Lemma 26.2 (Zorn's Lemma)**.** *Suppose $(A, \leq)$ is a set equipped with a partial order. If every chain in $(A, \leq)$ has an upper bound in A, then every element of A lies below a maximal element of A.*

(A maximal element is an element $a \in A$ such that $a \leq b \in A$ implies $b = a$.)

If we apply Zorn's Lemma to $(A, \leq) = (\mathcal{I}(R), \subseteq)$ we finish the proof of Proposition 26.1. □

Commentary. The proof of Zorn's Lemma is a souped-up version of the intuition presented above. It constructs a "transfinite" chain

$$a = a_0 < a_1 < a_2 < \cdots < a_\infty < a_{\infty+1} \cdots < a_{\infty+\infty} < \cdots$$

of elements of A. However, "constructs" is not quite right. At stage α, we have an element a_α which is not maximal. To "construct" $a_{\alpha+1}$, we need to **choose** one element (from potentially many) which properly extends a_α. There may be no natural way to do this (even though we know some such element must exist). Some mathematicians and philosophers have objected to "constructions" that require infinitely many ad hoc choices. The Axiom of Choice (in set theory) asserts that constructions of this kind are OK, so Zorn's Lemma is correct (unless the Axiom of Choice is false ...).

27. Rings of fractions

Suppose R is an integral domain and $D \subseteq R$ is a subset of R satisfying

(1) $1 \in D$.
(2) $0 \notin D$.
(3) D is closed under multiplication (i.e., $a, b \in D$ implies $ab \in D$).

(For example, the set $D = R\backslash\{0\}$ satisfies these properties.)

I will show that the standard construction of $\mathbb{Q}$ (as fractions n/d where $n, d \in \mathbb{Z}$ with $d \neq 0$) can be carried out to construct an integral domain of "fractions" r/d where $r \in R$ and $d \in D$.

Let $\mathcal{F} = R \times D = \{(r, d) : r \in R, d \in D\}$. Define a relation $\sim$ on $\mathcal{F}$ by

$$(r, d) \sim (s, e) \qquad \text{iff} \qquad re = sd.$$

Claim. *$\sim$ is an equivalence relation on $\mathcal{F}$.*

Proof. It is easily shown to be reflexive and symmetric. For transitivity, suppose $(r, d) \sim (s, e)$ and $(s, e) \sim (t, f)$. Thus $re = sd$ and $sf = te$. Hence

$$ref = sdf = sfd = ted,$$

so $(rf - td)e = 0$. As R is an integral domain, we can deduce $rf - td = 0$ or $e = 0$. However, $e \in D$ so $e \neq 0$ by (2). Hence $rf - td = 0$, so $(r, d) \sim (t, f)$, proving $\sim$ is transitive. □

For $(r, d) \in \mathcal{F}$ let r/d denote the equivalence class of $\sim$ containing (r, d). That is,

$$r/d = \{(s, e) \in \mathcal{F} : (r, d) \sim (s, e)\}.$$

Define F to be the set of these equivalence classes:

$$F = \{r/d : (r, d) \in \mathcal{F}\}.$$

Note that $r/d = s/e$ <u>means</u> $(r, d) \sim (s, e)$.

By (1), $r/1 \in F$ for all $r \in R$. We call $r/1$ the **image** of r. Note that distinct elements of R have distinct images in F, since $r/1 = s/1$ implies $(r, 1) \sim (s, 1)$, i.e., $r1 = s1$, i.e., $r = s$.

Next, we define $+$ and $\cdot$ on F in the "grade school" way:

$$\begin{aligned} r/d + s/e &:= (re + sd)/de \\ (r/d) \cdot (s/e) &:= rs/de. \end{aligned}$$

Note that $d, e \in D$ implies $de \in D$ by (3), so the right-hand sides make sense.

Theorem 27.1.

(1) *$+$ and $\cdot$ are well-defined.*
(2) *$(F, +, \cdot)$ is an integral domain.*
(3) *$\{r/1 : r \in R\}$ is a subring of F isomorphic to R.*

Commentary. The assertion that "+ is well-defined" means the following: for all $r_1, r_2, s_1, s_2 \in R$ and all $d_1, d_2, e_1, e_2 \in D$, if $r_1/d_1 = r_2/d_2$ and $s_1/e_1 = s_2/e_2$, then $(r_1e_1 + s_1d_1)/d_1e_1 = (r_2e_2 + s_2d_2)/d_2e_2$; equivalently, if $(r_1, d_1) \sim (r_2, d_2)$ and $(s_1, e_1) \sim (s_2, e_2)$, then $((r_1e_1 + s_1d_1), d_1e_1) \sim ((r_2e_2 + s_2d_2), d_2e_2)$. The proof of this claim, and everything else claimed in this theorem, is left as an excellent exercise. (In particular, $0/1$ will be the zero element and $1/1$ will be the identity element of F.)

In practice, we identify each element $r \in R$ with its image $r/1 \in F$. This makes R a virtual subring of F.

Definition. The ring F constructed above is called the **ring of fractions of R over D**, and is denoted $D^{-1}R$.

Claim. *If $D = R\backslash\{0\}$, then F is a field.*

Proof. We already know that $D^{-1}R$ is an integral domain, so it remains to show that every nonzero element has a multiplicative inverse. Suppose $r/d \in D^{-1}R$ is nonzero, i.e., $r/d \neq 0/1$. This means $r/d \neq 0/1$, i.e., $(r, d) \not\sim (0, 1)$, i.e., $r1 \neq 0d$, so $r \neq 0$. So $r \in D$. So $d/r \in F$, and clearly $(r/d) \cdot (d/r) = (rd, rd) = 1/1$. Hence r/d is invertible with inverse d/r. □

Example 27.2. Let $R = \mathbb{R}[x]$ and $D = R \setminus \{0\}$. Then $D^{-1}R$ is a field containing $\mathbb{R}[x]$ (virtually) as a subring, and every element of D^{-1}/R can be expressed as a fraction $p(x)/q(x)$ for some $p(x), q(x) \in \mathbb{R}[x]$ with $q(x) \neq 0$. This field is denoted $\mathbb{R}(x)$ and is called the field of **rational functions over** $\mathbb{R}$, but note that the elements of $\mathbb{R}(x)$ are **not** functions; they are equivalence classes of a relation $\sim$ defined on the set $\mathcal{F}$ of pairs $(p(x), q(x))$.

Example 27.3. Let $R = \mathbb{Z}$ and $D = \{d \in \mathbb{Z} : 3 \nmid d\}$. D satisfies assumptions (1)−(3), so the above construction gives an integral domain $D^{-1}\mathbb{Z}$ properly containing $\mathbb{Z}$, in which every integer in D becomes a unit. More precisely,

$$D^{-1}\mathbb{Z} = \{n/d : n, d \in \mathbb{Z}, d \not\equiv 0 \pmod 3\}.$$

Note that $D^{-1}\mathbb{Z}$ is not a field, since e.g. the element 3 is not invertible.

Example 27.4. More generally, suppose R is an integral domain and I is a prime ideal of R. Let $D = R \setminus I$, i.e., the complement of I. Then D satisfies assumptions (1)−(3) (exercise), so $D^{-1}R$ is defined. It is called the **localization of R at the prime ideal I**.

28. Chinese Remainder Theorem

Consider the ring $\mathbb{Z}$. Fix $m \geq 1$. Let $I = (m)$, and consider the quotient ring $\mathbb{Z}/I$. Note that for all $a, b \in \mathbb{Z}$,

$$\begin{aligned} a + I = b + I &\iff b - a \in I \\ &\iff b - a \in (m) \\ &\iff m | b - a \\ &\iff a \equiv b \pmod{m}. \end{aligned}$$

This motivates the next definition.

Definition. If R is a ring, I is an ideal, and $a, b \in R$, then we write

$$a \equiv b \pmod{I}$$

to mean $a + I = b + I$ (equivalently, $b - a \in I$).

Consider again the ring $\mathbb{Z}$. Suppose $m, n \in \mathbb{Z}$ are coprime, i.e., $\gcd(m, n) = 1$. We know from MATH 135 or 145 that there exist $r, s \in \mathbb{Z}$ with $rm + sn = 1$.

Now let $I = (m)$ and $J = (n)$. By the above, we have $rm \in I$ and $sn \in J$, so $1 = rm + sn \in I + J$. Since $I + J$ is an ideal, this proves $I + J = \mathbb{Z}$. This motivates:

Definition. Let R be a ring. Two ideals I, J are **coprime** if $I + J = R$.

Theorem 28.1 (Chinese Remainder Theorem)**.** [1] *Suppose R is a ring and I, J are coprime ideals. Then for all $a, b \in R$ there exists $c \in R$ such that*

$$\begin{aligned} c &\equiv a \pmod{I}, \quad \textit{and} \\ c &\equiv b \pmod{J}. \end{aligned}$$

Proof. Because $I + J = R$, there exist $e \in I$ and $f \in J$ with $1 = e + f$. Let $c = af + be$. Observe that

$$\begin{aligned} e &\equiv 0 \pmod{I} \quad \text{as } e \in I \\ f &\equiv 1 \pmod{I} \quad \text{as } 1 - f = e \in I \end{aligned}$$

Hence

$$c = af + be \equiv a1 + b0 \pmod{I}$$

i.e., $c \equiv a \pmod{I}$. A similar proof shows $c \equiv b \pmod{J}$. □

[1] This terminology is standard in European languages, at least. The reference is to the *Sūnzǐ Suànjīng*, a mathematical book which is believed to have been written some time in 386–589 C.E. in China. This book contains a problem which shows the author was aware of the version of this theorem for the integers. The first known statement of an algorithm for this theorem (for the integers) is due to an Indian mathematician from the 6th century.

Definition. Suppose $R = (R, \oplus, \odot)$ and $S = (S, \boxplus, \boxdot)$ are rings. Their **direct product** is $(R \times S, +, \cdot)$ where $+$ and $\cdot$ are defined coordinatewise:

$$\begin{aligned}(r_1, s_1) + (r_2, s_2) &:= (r_1 \oplus r_2, s_1 \boxplus s_2)\\ (r_1, s_1) \cdot (r_2, s_2) &:= (r_1 \odot r_2, s_1 \boxdot s_2).\end{aligned}$$

It is a ring. Moreover, $(R \times S, +)$ is the direct product of the abelian groups $(R, \oplus)$ and $(S, \boxplus)$. The zero element of $R \times S$ is $(0_R, 0_S)$. The identity element of $R \times S$ is $(1_R, 1_S)$.

In lecture 8 - Direct Products - I explained the test for recognizing direct products of groups: if $H, K \triangleleft G$ and $H \cap K = \{1\}$ and $HK = G$, then $G \cong H \times K$. In lecture 12 - 2nd and 3rd Isomorphism Theorems - I showed that, under the same hypotheses, we have $G/H \cong K$ and $G/K \cong H$. Hence

$$G \cong G/H \times G/K.$$

This last fact has a version that works for rings.

Corollary 28.2. *Suppose R is a ring and I, J are coprime ideals.*

(1) $R/(I \cap J) \cong R/I \times R/J$.
(2) *If $I \cap J = \{0\}$ then $R \cong R/I \times R/J$.*

Proof. (2) follows from (1) since $R \cong R/\{0\}$ To prove (1), define $\varphi : R \to R/I \times R/J$ by

$$\varphi(r) = (r + I, r + J).$$

The idea is to show that φ is a surjective ring homomorphism and apply the 1st Isomorphism Theorem. I won't check that φ is a homomorphism (but it is a good exercise in understanding definitions).

I will prove that φ is surjective. Suppose $(a + I, b + J)$ is an arbitrary element of $R/I \times R/J$. By the Chinese Remainder Theorem, there exists $c \in R$ with

$$\begin{aligned}c &\equiv a \pmod{I}, \quad \text{and}\\ c &\equiv b \pmod{J}.\end{aligned}$$

Thus

$$\varphi(c) = (c + I, c + J) = (a + I, b + J).$$

Finally, we compute the $\ker(\varphi)$. If $r \in R$, then

$$\begin{aligned}r \in \ker(\varphi) &\iff \varphi(r) = 0_{R/I \times R/J} = (0_{R/I}, 0_{R/J})\\ &\iff (r + I, r + J) = (I, J)\\ &\iff r \in I \text{ and } r \in J\\ &\iff r \in I \cap J.\end{aligned}$$

Hence $\ker(\varphi) = I \cap J$. □

Example 28.3. Let $R = \mathbb{Z}$ and $I = (m)$ and $J = (n)$ where $\gcd(m, n) = 1$. Then

$$\begin{aligned} I \cap J &= \{a \in \mathbb{Z} : m|a \text{ and } n|a\} \\ &= \{a \in \mathbb{Z} : mn|a\} \quad (\text{because } \gcd(m, n) = 1) \\ &= (mn). \end{aligned}$$

Thus $\mathbb{Z}/I \cong \mathbb{Z}_m$, $\mathbb{Z}/J \cong \mathbb{Z}_n$, and $\mathbb{Z}/(I \cap J) \cong \mathbb{Z}_{mn}$, so the CRT gives

$$\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n.$$

29. PIDs

Proposition 29.1. *Every ideal of $\mathbb{Z}$ is principal.*

Proof. Suppose I is an ideal of $\mathbb{Z}$. If $I = \{0\}$ then $I = (0)$. Otherwise, pick $a \in I$ with $a \neq 0$ and $|a|$ minimum. Clearly $(a) \subseteq I$. To prove $\supseteq$, assume $b \in I$. Divide b by a to get quotient q and remainder r, so

$$b = aq + r, \quad 0 \leq r < |a|.$$

$a, b \in I$ implies $r = b - aq \in I$. Hence $r = 0$, so $b = aq$, proving $b \in (a)$. □

Definition. A ring R is a **Principal Ideal Domain** (or **PID**) if

(1) R is an integral domain (commutative, $0 \neq 1$, no zero divisors).
(2) Every ideal of R is principal.

Example 29.2. The following are examples of PIDs.

(1) $\mathbb{Z}$.
(2) Any field (Because a field F only has two ideals: $\{0\} = (0)$ and $F = (1)$.)
(3) $\mathbb{R}[x]$.

    Proof. It is an integral domain. Let I be an ideal. If $I = \{0\}$ then $I = (0)$. Otherwise pick $f(x) \in I$ with $f(x) \neq 0$ and with $\deg(f(x))$ minimum. Clearly $(f(x)) \subseteq I$. For $\supseteq$, assume $g(x) \in I$. Divide $g(x)$ by $f(x)$ to get quotient $q(x)$ and remainder $r(x)$ (in $\mathbb{R}[x]$), so

    $$g(x) = f(x)q(x) + r(x), \qquad r(x) = 0 \text{ or } \deg(r(x)) < \deg(f(x)).$$

    $f(x)$, $g(x) \in I$ implies $r(x) = g(x) - f(x)q(x) \in I$. Hence $r(x) = 0$, so $g(x) = f(x)q(x)$, proving $g(x) \in (f(x))$. □

(4) More generally, $F[x]$ where F is a field. (Same argument, using the division algorithm in $F[x]$.)
(5) Even more generally, any integral domain for which we have a "division algorithm" which, given any $a, b \in R$ with $a \neq 0$, produces a quotient/remainder pair $q, r \in R$ satisfying
    - $b = aq + r$.
    - r is "strictly simpler" than a.

    There are several ways to formulate this. A standard way leads to the definition of *Euclidean domains*. $\mathbb{Z}[x]$ and polynomial rings $F[x]$ (where F is a field) are examples of Euclidean domains. We won't study Euclidean domains in this course, but you might want to explore them on your own.

30. Primes and irreducibles

The remainder of this course focusses on the properties of factorizations in integral domains.

Recall that if R is a ring then $R^\times$ denotes the set of **units** (invertible elements) of R. Also recall that in a commutative ring R we say a **divides** b and write $a|b$ if $b = ar$ for some $r \in R$.

Lemma 30.1. *In a commutative ring R, an element u is a unit iff $u|1$.*

Proof. $u|1$ iff $1 = uv$ for some $v \in R$, iff $v = u^{-1}$, i.e., $u \in R^\times$. □

Corollary 30.2. *In a commutative ring R, u is a unit iff $(u) = (1)$.*

Proof. $(u) = (1)$ iff $(1) \subseteq (u)$ (since the opposite inclusion is always true, as $(1) = R$). $(1) \subseteq (u)$ iff $u|1$ (by an assignment problem?). □

Definition. We say that a and b are **associates** and write $a \sim b$ if $a = ub$ for some unit $u \in R^\times$.

Example 30.3.

(1) In $\mathbb{Z}$, $a \sim b$ iff $a = \pm b$.
(2) In $\mathbb{R}[x]$, $2x + 3 \sim x + \frac{3}{2}$ since $2x + 3 = 2(x + \frac{3}{2})$ and 2 is invertible in $\mathbb{R}[x]$.

Lemma 30.4. *In an integral domain R, $a \sim b$ iff $a|b$ and $b|a$.*

Proof. ($\implies$). Assume $a \sim b$, so $a = ub$ with $u \in R^\times$. Then obviously $b|a$. And $u^{-1}a = b$ with $u^{-1} \in R$, by assumption, so $a|b$.
($\impliedby$). Assume $a|b$ and $b|a$. This means $b = ar$ and $a = bs$ for some $r, s \in R$. Hence $a = bs = (ar)s = a(rs)$, so $a(1 - rs) = 0$. As we are in an integral domain, we can deduce $a = 0$ or $1 - rs = 0$.

Case 1. $a = 0$
Then $b = ar$ implies $b = 0$, so we can write e.g. $a = 1b$. 1 is a unit so $a \sim b$.

Case 2. $1 - rs = 0$
Then $rs = sr = 1$, so s is a unit, so $a = sb$ gives $a \sim b$.

Thus $a \sim b$ in either case, proving ($\impliedby$). □

Corollary 30.5. *In an integral domain R, $a \sim b$ iff $(a) = (b)$.*

Proof. $(a) = (b)$ iff $(a) \subseteq (b)$ and $(b) \subseteq (a)$. By an assignment problem (?), this is equivalent to $b|a$ and $a|b$. □

Definition. Let R be an integral domain. Assume $a \in R$ with $a \neq 0$ and $a \notin R^\times$.

(1) A **nontrivial factorization** of a is an equation $a = bc$ where $b, c \in R$ and neither b nor c is a unit.
(2) a is **reducible** if it has a nontrivial factorization in R.
(3) Otherwise a is **irreducible** (equivalently, $a = bc$ implies b or c is a unit).

(4) We say that a is a **prime** if for all $b, c \in R$, if $a|bc$ then $a|b$ or $a|c$.

Note that these definitions are always **relative to** R. For example,

- 3 is both prime and irreducible in $\mathbb{Z}$.
- 3 is reducible in $\mathbb{Z}[\sqrt{3}]$, because 3 is not a unit and $3 = (\sqrt{3})(\sqrt{3})$ is a nontrivial factorization.
- 3 is neither reducible nor irreducible in $\mathbb{R}$, because it is a unit there.

Proposition 30.6. *In an integral domain, every prime is irreducible.*

Proof. Suppose p is prime and $p = bc$. We can write $p1 = bc$, so $p|bc$, so by definition of being a prime, $p|b$ or $p|c$.

Case 1: $p|c$.

We also have $c|p$ (from $p = bc$). So $p \sim c$, say $p = uc$ with $u \in R^\times$. Obviously $c \neq 0$ (as $p \neq 0$), so $bc = uc$ implies $b = u$ so $b \in R^\times$.

Case 2: $p|b$.

Then a similar argument shows $p \sim b$ and $c \in R^\times$. Since either Case 1 or 2 holds, we've shown that if $p = bc$ then b or c is a unit. So p has no nontrivial factorization in R, meaning it is irreducible. □

The converse is not always true, as the next example shows.

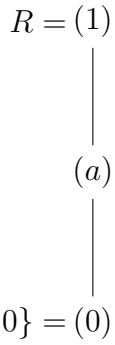
Example 30.7. Let R be the set of all complex numbers of the form $a + bi\sqrt{5}$ where $a, b \in \mathbb{Z}$. R is a subring of $\mathbb{C}$ and so is an integral domain. It is possible to show that $R^\times = \{1, -1\}$ and that 3 is irreducible in R, i.e., cannot be factored nontrivially. Let $c = 2 + i\sqrt{5}$ and $d = 2 - i\sqrt{5}$. So $c, d \in R$ and $cd = 4 + 5 = 9$, so clearly $3|cd$. But 3 divides neither c nor d in R (since $\frac{2}{3} \pm \frac{1}{3}i\sqrt{5} \notin R$). Thus 3 is **not** a prime in R.

31. Complete Factorizations

Recap: in integral domain R, suppose $a \neq 0$ and $a \notin R^\times$.

- A factorization $a = bc$ is **trivial** if b or c is a unit, and is **nontrivial** otherwise.

We can picture nontrivial factorization in the partially ordered set of ideals of R.

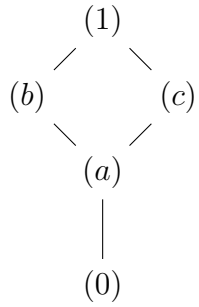


The hypothesis translates as follows:

- $a \neq 0 \iff (a) \neq (0)$
- $a \notin R^\times \iff (a) \neq (1)$

Now suppose $a = bc$. Focus on (b).

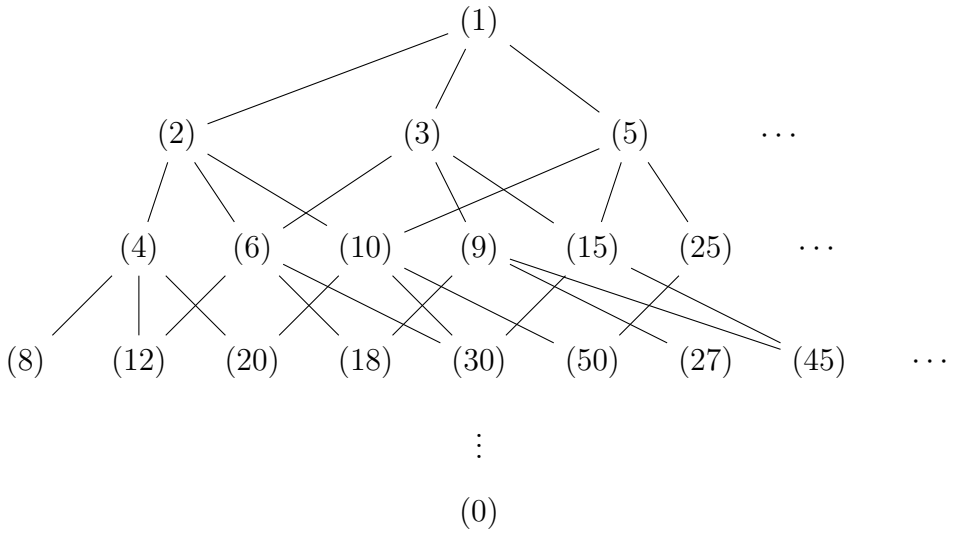
- If this factorization is trivial, then b or c is a unit.
  - If b is a unit, then $(b) = (1)$
  - If c is a unit, then $b \sim a$, so $(b) = (a)$
- If the factorization is nontrivial, then neither b nor c is a unit. Because b is not a unit, $(b) \neq (1)$. Because c is not a unit, $b \nsim a$, so $(b) \neq (a)$. Of course $(b) \subseteq (1)$. Finally, $b|a$ so $(a) \subseteq (b)$. Hence $(a) \subset (b) \subset (1)$. (Similar remarks hold for (c).)



It follows that the factorization is nontrivial iff $(a) \subset (b) \subset (1)$. This proves:

Proposition 31.1. *Suppose R is an integral domain and $a \in R$. Then a is irreducible iff $(a) \neq (0)$, $(a) \neq (1)$, and there is no principal ideal (b) properly between (a) and (1).*

Example 31.2. Draw a picture of the principal ideals in $\mathbb{Z}$.



Notice:

- $6 = 2 \cdot 3$ translates to (2) and (3) above (6), below (1).
- $4 = 2 \cdot 2$ translates to just (2) above (4), below (1).
- Irreducibles (= primes) are just below the top (they are maximal ideals).
- Can also "see" $(12) + (30)$; it must be the smallest ideal containing both (12) and (30). We see that it is (6). Consistent with $6 = \gcd(12, 30)$.

Definition. Suppose R is an integral domain, $a \in R$, $a \neq 0$, and $a \notin R^\times$. A **complete factorization** of a is an equation

$$a = p_1 p_2 \cdots p_n$$

where $n \geq 1$, $p_1, p_2, \ldots, p_n \in R$, and each p_i is irreducible.

Naive algorithm to find a complete factorization. Given $a \in R$ with $a \neq 0$ and $a \notin R^\times$:

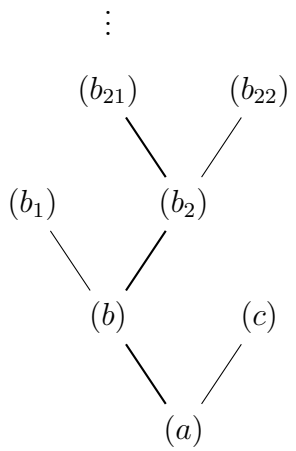
- If a is irreducible, then done.
- Else pick a nontrivial factorization $a = bc$.
- Recursively find complete factorizations for b and c:

$$b = p_1 p_2 \cdots p_n \quad \text{and} \quad c = q_1 q_2 \cdots q_m$$

- Then $a = p_1 p_2 \cdots p_n q_1 q_2 \cdots q_m$ is a complete factorization of a.

There is one potential problem with this algorithm? What is the problem? (Answer: it may never terminate)

For example, b might have a nontrivial factorization $b = b_1 b_2$. Then b_2 might have a nontrivial factorization $b_2 = b_{21} b_{22}$. And so on forever.



The bad thing (failure to terminate) can **only** happen if there is an **infinite** strictly increasing chain $(a) \subset (b) \subset (b_2) \subset (b_{21}) \subset \cdots$ of principal ideals. This proves:

Proposition 31.3. *Suppose R is an integral domain and R does **not** have an infinite strictly increasing chain of principal ideals. Then every $a \in R$ with $a \neq 0$, $a \notin R^\times$ has a complete factorization.*

32. Unique Factorization

Proposition 31.3 addresses the existence of complete factorizations. Next we study uniqueness.

Example 32.1. In $\mathbb{Z}$, 6 has four complete factorizations:

$$6 = (2)(3)$$
$$6 = (3)(2)$$
$$6 = (-2)(-3)$$
$$6 = (-3)(-2)$$

These are "essentially the same" factorization.

Definition. Let R be an integral domain and $a \in R$ with $a \neq 0$, $a \notin R^\times$.

(1) Two complete factorizations of a,

$$a = p_1 p_2 \cdots p_n \quad \text{and} \quad a = q_1 q_2 \cdots q_m$$

are **essentially the same** provided:
   (a) $m = n$, and
   (b) After a suitable re-ordering of the q_i's we have $p_i \sim q_i$ for all $i = 1, \ldots, n$.

(2) We say that **complete factorizations in R are unique, when they exist**, and we write "R has UCF," provided for any $a \in R$ with $a \neq 0$ and $a \notin R^\times$, if a has a complete factorization, then any two complete factorizations of a are essentially the same.

Example 32.2. Recall the integral domain $R = \{a+bi\sqrt{5} : a, b \in \mathbb{Z}\}$ from a previous lecture. The element 9 has two essentially different complete factorizations:

$$9 = 3 \cdot 3 \quad \text{and} \quad 9 = (2 + i\sqrt{5})(2 - i\sqrt{5}).$$

Hence R does not have UCF.

Recall (Section 30) that an element $p \in R$ of an integral domain is a **prime** if for all $a, b \in R$, $p|ab$ imples $p|a$ or $p|b$. Suppose p is prime and $p|a_1 a_2 \cdots a_n = a_1(a_2 \cdots a_n)$. Then $p|a_1$ or $p|a_2 \cdots a_n$; in the latter case $p|a_2$ or $p|a_3 \cdots a_n$, etc. Hence

Lemma 32.3. *In an integral domain, if p is a prime and $p|a_1 a_2 \cdots a_n$, then $p|a_i$ for some i.*

Corollary 32.4. *Suppose R is an integral domain, $p \in R$ is a prime, and $a = q_1 \ldots q_m$ is a complete factorization of $a \in R$. Then $p|a$ iff $p \sim q_i$ for some i.*

Proof. Obviously if $p \sim q$, then $p|q_i$ so $p|a$. Conversely, suppose $p|a$. Then $p|q_i$ for some i, by the Lemma. Thus $q_i = pu$ for some $u \in R$. q_i is irreducible, so p or u must be a unit. p is not a unit (it is prime), so u is a unit. Hence $p \sim q_i$. □

Proposition 32.5. *Suppose R is an integral domain in which every irreducible element is prime. Then R has UCF.*

Proof. We repeat the proof from MATH 135/145. Suppose $a \in R$, $a \neq 0$, $a \notin R^\times$, and

$$a = p_1 p_2 p_3 \cdots p_n \quad \text{and} \quad a = q_1 q_2 q_3 \cdots q_m.$$

where each p_i, q_j is irreducible. By assumption, each p_i is a prime. Clearly $p_1 | a$, so $p_1 | q_1 q_2 \cdots q_m$. As p_1 is prime, the Corollary gives $p_1 \sim q_i$ for some i. We can re-order the q's so that $p_1 \sim q_1$. Then $p_1 = u_1 q_1$ for some $u_1 \in R^\times$. Thus

$$(u_1 q_1) p_2 p_3 \cdots p_n = q_1 q_2 q_3 \cdots q_m.$$

Cancelling q_1 (OK since we're in an integral domain) gives

$$u_1 p_2 \cdots p_n = q_2 \cdots q_m.$$

p_2 divides the left side, so divides the right side. Hence $p_2 \sim q_j$ for some $j = 2, \ldots, m$. Again we can re-order the remaining q's and assume $p_2 \sim q_2$, say $p_2 = u_2 q_2$. Then

$$u_1 (u_2 q_2) p_3 \cdots p_n = q_2 q_3 \cdots q_m.$$

Cancelling q_2 gives

$$(u_1 u_2) p_3 \ldots p_n = q_3 \cdots q_m.$$

Continuing in this way, we pair up each p_i with one of the remaining q's, until we run out of p's or q's. If we run out of q's before running out of p's, i.e. $m < n$, then after m steps we will have

$$(u_1 u_2 \cdots u_m) p_{m+1} \cdots p_n = 1.$$

But then $p_n | 1$, which is impossible as p_n is not a unit. A similar contradiction arises if we run out of p's before running out of q's, i.e., $n < m$. Hence $m = n$ and the p's and q's are perfectly matched in associate pairs, meaning the two factorizations are essentially the same. $\square$

33. UFDs

Summary. Suppose R is an integral domain.

(1) If R does **not** have an infinite strictly increasing chain of principal ideals, then complete factorization always exits. (Proposition 31.3)
(2) If every irreducible in R is a prime, then complete factorizations are unique (when they exist). (Proposition 32.5)

Definition. An integral domain R is a **Unique Factorization Domain** (UFD) if (1) R does not have an infinite strictly increasing chain of principal ideals, and (2) every irreducible in R is a prime.

Example 33.1. $\mathbb{Z}$ is a UFD.

- We've already seen the poset of principal ideals of $\mathbb{Z}$; it has no upward chain of infinite length.
- The set of irreducibles in $\mathbb{Z}$ is $\{2, -2, 3, -3, 5, -5, \ldots\}$ and every element p in this set satisfies $p|ab \implies p|a$ or $p|b$, so is a prime.

UFDs are the integral domains in which factorization works "just like in $\mathbb{Z}$." It will become clear what I mean by this in the following discussion. Let R be an integral domain in which every $a \in R$ with $a \neq 0$, $a \notin F^\times$ has a complete factorization. Let R^{ir} denote the set of irreducible elements of R. Consider the "associates" equivalence relation $\sim$ on R^{ir}, and choose a set $\mathcal{P}$ of representatives of the equivalence classes; that is, choose $\mathcal{P} \subseteq R^{\text{ir}}$ so that $\mathcal{P}$ contains exactly one element from each equivalence class of $\sim$ in R^{ir}. Now given $a \in R$ with $a \neq 0$ and $a \notin R^\times$, and suppose that

$$a = q_1 q_2 \cdots q_n$$

is a complete factorization of a. The elements $q_1, \ldots, q_n$ are in R^{ir}, and it is possible that some of them belong to the same equivalence class of $\sim$. Choose $p_1, \ldots, p_k \in \mathcal{P}$ to be the *distinct* elements of $\mathcal{P}$ associated to some q_i in the above complete factorization. Thus there is a surjective function $\sigma : \{1, \ldots, n\} \to \{1, \ldots, k\}$ such that $q_i \sim p_{\sigma(i)}$ for each i; hence there exist units $u_1, \ldots, u_n \in R^\times$ such that $q_i = u_i p_{\sigma(i)}$ for each i. Hence

$$\begin{aligned} a &= (u_1 p_{\sigma(1)})(u_2 p_{\sigma(2)}) \cdots (u_n p_{\sigma(n)}) \\ &= (u_1 u_2 \cdots u_n) p_{\sigma(i)} p_{\sigma(2)} \cdots p_{\sigma(n)} \\ &= u p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k} \end{aligned}$$

where $u = u_1 u_2 \cdots u_n$ and each t_ℓ is the number of times q_i is associated to p_ℓ. We will call the equation

$$a = u p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k}$$

describing a as a product of a unit and powers of distinct elements of $\mathcal{P}$ a **standard factorization using** $\mathcal{P}$.

Example 33.2. $\mathbb{Z}^{\text{ir}}$ is the set $\{2, -2, 3, -3, 5, -5, \ldots\}$. The equivalence classes of $\sim$ in $\mathbb{Z}^{\text{ir}}$ are the sets of the form $\{p, -p\}$ where p is a prime number. Thus we can

choose $\mathcal{P} = \{2, 3, 5, \ldots\}$. Then every integer $a \in \mathbb{Z}$ with $n \neq 0$ and $n \notin \{1, -1\}$ can be given a standard factorization using $\mathcal{P}$, i.e.,

$$a = up_1^{t_1} \cdots p_k^{t_k}$$

where $u \in \{1, -1\}$ and $p_1, \ldots, p_k$ are distinct prime numbers.

Continuing the general discussion, suppose that now in addition we assume that R is a UFD. Let $a \in R$ with $a \neq 0$ and $a \notin R^\times$, and let $\mathcal{P}$ be a set of representatives for the equivalence classes of $\sim$ in R^{ir}. In this situation, a has *exactly one* standard factorization using $\mathcal{P}$, up to a reordering of the irreducibles $p_1, \ldots, p_k$ from $\mathcal{P}$. That is, *if*

$$a \;=\; up_1^{s_1} \cdots p_k^{s_k} \;=\; vq_1^{t_1} \cdots q_\ell^{t_\ell}$$

where $p_1, \ldots, p_k$ are distinct elements of $\mathcal{P}$, $q_1, \ldots, q_\ell$ are distinct elements from $\mathcal{P}$, the exponents $s_1, \ldots, s_k, t_1, \ldots, t_\ell$ are all positive integers, and $u, v \in R^\times$, *then*

- $k = \ell$ and $u = v$.
- There exists a permutation $\sigma \in S_k$ such that $(q_i, t_i) = (p_{\sigma(i)}, s_{\sigma(i)})$ for each i.

This can be proved by an argument similar to the proof of Proposition 32.5. The details are left as an exercise.

Lemma 33.3. *Suppose R is a UFD, $\mathcal{P}$ is a set of representatives of the equivalence classes of $\sim$ in R^{ir}, and $a \in R$ with $a \neq 0$, $a \notin R^\times$. Let the standard factorization of a using $\mathcal{P}$ be*

$$a = up_1^{t_1} \cdots p_k^{t_k}.$$

Then for any $b \in R$, $b|a$ iff b can be written

$$b = vp_1^{s_1} \cdots p_k^{s_k}$$

for some $v \in R^\times$ and some integers $s_1, \ldots, s_k$ satisfying $0 \leq s_i \leq t_i$ for each i.

Proof sketch. ($\Leftarrow$) If b can be written $b = vp_1^{s_1} \cdots p_k^{s_k}$ with each $s_i \leq t_i$, then $a = bc$ where $c = uv^{-1}p_1^{t_1 - s_1} \cdots p_k^{t_k - s_k} \in R$, so $b|a$.

($\Rightarrow$) This is proved similarly to the proof of Proposition 32.5 and is left as an exercise. $\square$

34. PIDs are UFDs

Recall (Section 25) that an ideal I of a commutative ring R is a **prime ideal** if $I \neq R$ and for all $a, b \in R$, if $ab \in I$ then $a \in I$ or $b \in I$. In particular, every maximal ideal is a prime ideal (see last lecture)

Lemma 34.1. *Let R be an integral domain and $p \in R$ with $p \neq 0$. (p) is a prime ideal iff p is a prime.*

Proof. ($\Rightarrow$) Assume (p) is a prime ideal. We already know that $p \neq 0$. p cannot be a unit, since if it were, then we should have $(p) = (1)$, contradicting the assumption that $(p) \neq R$. Finally, assume $a, b \in R$ and $p|ab$. Then $ab \in (p)$. Since (p) is prime, we get $a \in (p)$ or $b \in (p)$, i.e., $p|a$ or $p|b$. Thus p is prime.

($\Leftarrow$) Proved similarly (exercise). □

Recall (Section 29) that a **Principal Ideal Domain** (PID) is an integral domain in which every ideal is principal.

Proposition 34.2. *Suppose R is a PID and $p \in R$ with $p \neq 0$. The following are equivalent:*

1. *p is irreducible.*
2. *p is a prime.*
3. *(p) is a maximal ideal.*

Proof. (3) $\Rightarrow$ (2). Assume that (p) is a maximal ideal. Then (p) is a prime ideal by Corollary 25.4. Hence p is prime by Lemma 34.1.

(2) $\Rightarrow$ (1). Every prime is irreducible by Proposition 30.6.

(1) $\Rightarrow$ (3). Assume p is irreducible. Then $(p) \neq (1)$ and there is no principal ideal properly between (p) and (1). But R is a PID, so this means there is no ideal properly between (p) and (1). That means (p) is a maximal ideal. □

Here is an easy corollary that will be important in PMATH 348.

Corollary 34.3. *Suppose R is a PID and p is an irreducible element in R. Then $R/(p)$ is a field.*

Proof. (p) is a maximal ideal by Proposition 34.2, so $R/(p)$ is a field by Proposition 25.2. □

Here is the main theorem of this section.

Theorem 34.4. *Every PID is a UFD.*

Proof. Let R be a PID. Proposition 34.2 shows that every irreducible element of R is a prime. It remains to show that R has no infinite strictly increasing chain of principal ideals. Suppose, to the contrary, that $(a_1) \subset (a_2) \subset \cdots \subset (a_n) \subset \cdots$ is an infinite strictly increasing chain of principal ideals. Let $I = \cup_{n=1}^{\infty}(a_n)$. Recall that I is an ideal.

Because R is a PID, I is principal, say $I = (c)$. Then $c \in \cup_{n=1}^{\infty}(a_n)$, so $c \in (a_n)$ for some n. But then $(c) \subseteq (a_n)$, contradiction. □

Corollary 34.5. *If F is a field, then $F[x]$ is a UFD.*

Proof. $F[x]$ is a PID (because it has a division algorithm - see Example 29.2). □

Example 34.6. $\mathbb{Z}[x]$ is not a PID. So we cannot use the above theorem to deduce that $\mathbb{Z}[x]$ is a UFD. Similarly, the ring of polynomials $F[x, y]$ in two variables is not a PID, even if F is a field, so we cannot use the above theorem to prove that such polynomial rings are UFDs. In the final sections we will see results that imply $\mathbb{Z}[x]$ and $F[x_1, \ldots, x_n]$ are UFDs.

35. GCDs

Lemma 33.3 implies that "greatest common divisors" behave in UFDs just as they do in $\mathbb{Z}$. First, we define greatest common divisors similarly to how they are defined in MATH 135/145.

Definition. Let R be an integral domain and $a, b, d \in R$. We say that d is **a greatest common divisor** of a and b if:

(1) d is a common divisor: $d|a$ and $d|b$.
(2) d is divisible by every common divisor: for all $c \in R$, if $c|a$ and $c|b$, then $c|d$.

Greatest common divisors, when they exist, are unique up to associates. That is, if d, e are greatest common divisors of a and b, then $e|d$ and $d|e$ so $d \sim e$.

Greatest common divisors of a finite list of elements $a_1, \ldots, a_n$ are defined similarly: they are common divisors which are divisible by all common divisors.

Lemma 35.1. *Suppose R is a UFD. For every finite list $a_1, \ldots, a_n \in R$, if at least one of the a_i's is nonzero, then the list has a greatest common divisor.*

Proof. Re-order the list so that the first k entries $a_1, \ldots, a_k$ are nonzero and the rest (if any) all equal 0. Note that every element of R is a divisor of 0, so a greatest common divisor of $a_1, \ldots, a_k$ will also be a greatest common divisor of $a_1, \ldots, a_k, 0, \ldots, 0$. Thus we can ignore the zero entries and focus on $a_1, \ldots, a_k$.

If one of the a_i's is a unit, then the common divisors of $a_1, \ldots, a_k$ are just the units and so 1 is a greatest common divisor.

Assume none of the a_i's is a unit. Thus each a_i has a standard factorization using a set $\mathcal{P}$ of representatives of the equivalence classes of $\sim$ in R^{ir}. Let $p_1, \ldots, p_r$ be the distinct elements of $\mathcal{P}$ occurring in these standard factorizations. Then we can write

$$\begin{aligned} a_1 &= u_1 p_1^{t_{1,1}} \cdots p_r^{t_{1,r}} \\ a_2 &= u_2 p_1^{t_{2,1}} \cdots p_r^{t_{2,r}} \\ &\vdots \\ a_k &= u_k p_1^{t_{k,1}} \cdots p_r^{t_{k,r}} \end{aligned}$$

where $u_1, \ldots, u_k$ are units and the exponents $t_{i,j}$ are nonnegative integers (some of them can equal 0). In other words, the equations above are standard factorizations using $\mathcal{P}$, possibly padded by powers of the form p_j^0.

For each $j = 1, \ldots, r$ define $m_j = \min(t_{1,j}, t_{2,j} \ldots, t_{k,j})$, i.e., the minimum exponent of p_j in the standard factorizations. Now let

$$d = p_1^{m_1} \cdots p_r^{m_r}.$$

Lemma 33.3 can be used to show that the common divisors of $a_1, \ldots, a_k$ are precisely the elements of the form $v p_1^{e_1} \cdots p_r^{e_r}$ where v is a unit and the exponents satisfy $0 \leq e_j \leq m_j$. These are exactly the divisors of d, so d is a greatest common divisor of $a_1, \ldots, a_k$. □

Definition. Suppose R is an integral domain and $a_1, \ldots, a_n \in R$. We say that $a_1, \ldots, a_n$ are **relatively prime** if the only common divisors of $a_1, \ldots, a_n$ are the units in $R^\times$; equivalently, if 1 is a greatest common divisor of $a_1, \ldots, a_n$.

Lemma 35.2. *Suppose R is a UFD and $a_1, \ldots, a_n \in R$ with at least one $a_i \neq 0$. Let $d \in R$ be a greatest common divisor of $a_1, \ldots, a_n$ (which we know exists by Lemma 35.1). Define $a_1', \ldots, a_n' \in R$ by $a_i' := a_i/d$ (i.e., a_i' is the unique solution x to $a_i = dx$). Then $a_1', \ldots, a_n'$ are relatively prime.*

In other words, if R is a UFD then you can always divide a finite list of elements by their "greatest common divisor" to get a new list that is relatively prime.

Proof. Suppose c is a common factor of $a_1', \ldots, a_n'$ and c is not a unit. Then c has a complete factorization (since R is a UFD), so there exists an irreducible element p with $p|c$. Then p is also a common divisor of $a_1', \ldots, a_n'$. Thus pd is a common divisor of $a_1, \ldots, a_n$. Since d is a greatest common divisor of $a_1, \ldots, a_n$ we must have $pd|d$, which implies $p|1$, which implies p is a unit, contradiction. □

36. Gauss' Lemma

Lemma 36.1. *Suppose R is an integral domain and $p \in R$ is a prime in R. Then p is prime in $R[x]$.*

Proof. Assume $f(x), g(x) \in R[x]$ and $p|f(x)g(x)$. Write

$$f(x) = a_0 + a_1x + \ldots + a_mx^m$$
$$g(x) = b_0 + b_1x + \ldots + b_nx^n.$$

Thus

$$f(x)g(x) = c_0 + c_1x + \ldots + c_{m+n}x^{m+n} \quad \text{where } c_k = \sum_{i+j=k} a_ib_j.$$

Since $p|f(x)g(x)$, we have $p|c_k$ for all k.

Suppose neither $f(x)$ nor $g(x)$ is divisible by p. Thus at least one coeffcient of $f(x)$ and one of $g(x)$ are not divisible by p. Let r and s be the first such that $p \nmid a_r$ and $p \nmid b_s$. Let $k = r + s$ and look at c_k:

$$c_k = (a_0b_k + \ldots + a_{r-1}b_{s+1}) + a_rb_s + (a_{r+1}b_{s-1} + \ldots + a_kb_0).$$

By the choice of r and s, p divides each a_i for $i < r$ and b_j for each $j < s$. Since p also divides c_k, we get $p|a_rb_s$. As p is prime, we get $p|a_r$ or $p|b_s$, contradicting our choice of r, s. □

Lemma 36.2. *Suppose R is a UFD, $f(x), g(x) \in R[x]$, and $u \in R$, $u \neq 0$. If $u|f(x)g(x)$, then there exists a factorization $u = cd$ of u in R such that $c|f(x)$ and $d|g(x)$.*

Proof. If u is a unit (i.e., $u \in R^\times$), then we use $u = u1$. Clearly $u|f(x)$ (since $u|1$) and $1|g(x)$.

Assume u is not a unit. Because R is a UFD, u has a complete factorization

$$u = p_1p_2 \ldots p_n, \quad \text{each } p_i \text{ irreducible.}$$

Again because R is a UFD, each p_i is prime in R and so is a prime in $R[x]$ by Lemma 36.1.

We have $p_1|f(x)g(x)$, so p_1 divides $f(x)$ or $g(x)$. Say $p_1|f(x)$. Let $f_1(x) \in R[x]$ be the result of dividing $f(x)$ by p_1; then

$$p_1p_2 \cdots p_n|(p_1f_1(x))g(x).$$

Canceling p_1, we get

$$p_2 \cdots p_n|f_1(x)g(x)$$

Repeating the argument, p_2 must divide $f_1(x)$ or $g(x)$. Continuing in this way, we can "factor out" each p_i. If c is the product of the p_i's we remove from $f(x)$ and d is the product of the p_i's we remove from $g(x)$, then $cd = u$, $c|f(x)$, and $d|g(x)$. □

In the next Proposition, think of R being $\mathbb{Z}$ and F being $\mathbb{Q}$.

Proposition 36.3 (Gauss' Lemma)**.** *Suppose R is UFD and F is its field of fractions $\{n/d : n, d \in R,\ d \neq 0\}$. Let $p(x) \in R[x]$ by a polynomial of degree ≥ 1.*

Every nontrivial factorization of $p(x)$ in $F[x]$ can be essentially realized in $R[x]$, in the following sense: if $p(x) = A(x)B(x)$ is a nontrivial factorization of $p(x)$ in $F[x]$, then there exists $t \in F^\times$ such that $tA(x) \in R[x]$ and $t^{-1}B(x) \in R[x]$.

The point is that if $a(x) := tA(x)$ and $b(x) := t^{-1}B(x)$ then $p(x) = a(x)b(x)$ is a factorization of $p(x)$ in $R[x]$ whose factors have the same degrees as the degrees of the factors $A(x)$ and $B(x)$ in the original factorization.

Example 36.4. Let $R = \mathbb{Z}, F = \mathbb{Q}$, and $p(x) = 2x^2 + 7x = 3$. A nontrivial factorization of $p(x)$ in $\mathbb{Q}[x]$ is

$$p(x) = (x + \frac{1}{2})(2x + 6).$$

We can multiply the first factor by 2 and the second factor by $\frac{1}{2}$ to get an equivalent factorization

$$p(x) = (2x + 1)(x + 3),$$

which is a factorization in $\mathbb{Z}[x]$

Proof of Gauss' Lemma. Write each coefficient of $A(x)$ as a fraction n_i/d_i with $n_i, d_i \in R$. Let r be the product of all the denominators in these expressions and let $f(x) = rA(x)$. Then $f(x) \in R[x]$ (we have "cleared the denominators"). Similarly, write the coefficients of $B(x)$ as fractions, let s be the product of the denominators in $B(x)$, and define $g(x) := sB(x) \in R[x]$. Finally let $u = rs$ and note that $u \in R$ and

$$up(x) = (rs)A(x)B(x) = f(x)g(x).$$

Because $f(x), g(x), p(x) \in R[x]$, we can use the previous Lemma to obtain a factorization $u = cd$ of u in R such that $c|f(x)$ and $d|g(x)$. Thus $f(x) = ca(x)$ and $g(x) = db(x)$ with $a(x), b(x) \in R[x]$. Note that $cd = rs$, so $r/c = d/s$ in F. Let $t = r/c \in F$. Then

$$tA(x) = (r/c)A(x) = (1/c)rA(x) = (1/c)f(x) = a(x) \in R[x]$$
$$t^{-1}B(x) = (s/d)B(x) = (1/d)sB(x) = (1/d)g(x) = b(x) \in R[x]$$

as required. □

37. Primitive polynomials over a UFD

The following result is particularly useful in Galois theory (PMATH 348).

Corollary 37.1. *Suppose $f(x) \in \mathbb{Z}[x], deg(f(x)) \geq 1$, and $f(x)$ is irreducible in $\mathbb{Z}[x]$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. If $f(x)$ has a nontrivial factorization in $\mathbb{Q}[x]$, then $f(x)$ has a nontrivial factorization in $\mathbb{Z}[x]$ by Gauss' Lemma. □

The converse is false. For example, $6x + 8$ is irreducible in $\mathbb{Q}[x]$ (every polynomial of degree 1 is irreducible), but it is reducible in $\mathbb{Z}[x]$ since $6x + 8 = 2(3x + 4)$ is a nontrivial factorization in $\mathbb{Z}[x]$ (neither 2 nor $3x + 4$ is a unit).

Definition. Suppose R is an integral domain and $f(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. We say that $f(x)$ is **primitive** in $R[x]$ if its coefficients $a_0, a_1, \ldots, a_n$ are relatively prime in R.

Corollary 37.2. *Suppose R is a UFD and F is its field of fractions. Let $f(x) \in R[x]$ with $deg(f) \geq 1$. The following are equivalent:*

1. *$f(x)$ is irreducible in $R[x]$.*
2. *$f(x)$ is primitive in $R[x]$ and irreducible in $F[x]$.*

Proof sketch. (1) ⇒ (2) uses Gauss' Lemma to prove irreducibility in $F[x]$.

(2) ⇒ (1). Assume $f(x)$ is primitive in $R[x]$ and irreducible in $F[x]$, but is reducible in $R[x]$. Then the nontrivial factorization of $f(x)$ in $R[x]$ must be of the form $f(x) = dg(x)$ (if both factors had degrees ≥ 1 then it would be a nontrivial factorization in $F[x]$). Thus $d \notin R^\times$ and $d|f(x)$ in $R[x]$, so d divides each coefficient of $f(x)$, contradicting that $f(x)$ is primitive. □

Corollary 37.3. *Suppose R is a UFD. Every nonzero polynomial $f(x) \in R[x]$ can be factored $f(x) = dg(x)$ were $d \in R$, $g(x) \in R[x]$, and $g(x)$ is primitive.*

Example. In $\mathbb{Z}[x], 6x + 8 = 2(3x + 4)$ with $3x + 4$ primitive.

Proof of Corollary 37.3. Write $f(x) = a_0 + a_1x + \ldots + a_nx^n$. By the 1st Lemma from previous lectures, there exists $d \in R$ which is a common divisor of $a_0, \ldots, a_n$ and, if $a_i = da_i'$ for $i = 0, \ldots, n$ then $a_0', \ldots, a_n'$ are relatively prime. Obviously $f(x) = d(a_0' + a_1'x + \ldots + a_n'x^n) = dg(x)$ and $g(x)$ is primitive. □

Crucial Lemma. *Suppose R is a UFD, $c, d \in R$ are non zero, and $f(x), g(x) \in R[x]$ are primitive. If $(cf) \subset (dg)$ then*

1. *$(c) \subseteq (d)$,*
2. *$\deg(f) \geq \deg(g)$, and*
3. *Either $(c) \subset (d)$ or $\deg(f) > \deg(g)$.*

Remark. (cf) and (dg) are ideals in $R[x]$. (c) and (d) denote the ideals in R.

Proof of the Crucial lemma. Assume $(cf) \subset (dg)$. Hence $dg(x)|cf(x)$, so

$$d|cf(x) \quad \text{and} \quad g(x)|cf(x).$$

The second obviously implies $\deg(f) \geq \deg(g)$, proving (2). Because $d|cf(x)$, yesterday's second lemma says that d has a factorization $d = ab$ such that $a|c$ and $b|f(x)$. Because $f(x)$ is primitive, b must be a unit. Hence (using $d = ab$) we get that d and a are associates, i.e., $d \sim a$, which implies $d|a$. As $d|a$ and $a|c$, we get $d|c$ and hence $(c) \subseteq (d)$. This proves (1).

To prove (3), assume that (3) fails, i.e., $(c) = (d)$ and $\deg(f) = \deg(g)$.

- From $(c) = (d)$, we can write $\boxed{d = cu}$ for some unit $u \in R^\times$
- From $dg(x)|cf(x)$ and the fact that f, g have the same degree, we get $\boxed{cf(x) = e(dg(x))}$ for some $e \in R$.
- Hence $cf(x) = e(cu)g(x)$, so cancelling c we get $\boxed{f(x) = eug(x)}$.
- Hence $e|f(x)$. But $f(x)$ is primitive. Hence e is a unit.
- Hence (from the 2nd bullet) $\boxed{cf(x) \sim dg(x)}$.

But that would imply $(cf) = (dg)$, contradicting our assumption. $\square$

38. The Big Theorem

Theorem 38.1. *If R is a UFD, then so is $R[x]$.*

Proof. Assuming R is a UFD, we must show that

(1) $R[x]$ has no infinite strictly increasing chain of principal ideals, and
(2) Every irreducible polynomial in $R[x]$ is prime.

(1) Assume that $(f_1) \subset (f_2) \subset \ldots \subset (f_n) \subset \ldots$ is an infinite strictly increasing sequence of principal ideals in $R[x]$.

By Corollary 37.3, we can factor each $f_n(x) = c_n g_n(x)$ where $c_n \in R$ and $g_n(x)$ is primitive. Thus

$$(c_1 g_1) \subset (c_2 g_2) \subset \ldots \subset (c_n g_n) \subset \ldots$$

By the Crucial Lemma, we have

$$(c_1) \subseteq (c_2) \subseteq \ldots \subseteq (c_n) \subseteq \ldots$$

and

$$\deg(g_1) \geq \deg(g_2) \geq \ldots \geq \deg(g_n) \geq \ldots$$

and for every i,

$$(c_i) \subset (c_{i+1}) \quad \text{or} \quad \deg(g_i) > \deg(g_{i+1}).$$

The second option cannot happen infinitely often, since degrees are non-negative integers. Hence beyond some point we always have the first option, meaning

$$(c_N) \subset (c_{N+1}) \subset \ldots \subset (c_{N+k}) \subset \ldots$$

But that means R has an infinite strictly increasing chain of principal ideals, contradicting that R is a UFD. This proves (1).

(2) Assume that $p(x)$ is an irreducible polynomial in $R[x]$ and $a(x), b(x) \in R[x]$ with $p(x)|a(x)b(x)$. I must show that $p(x)|a(x)$ or $p(x)|b(x)$.

By Corollary 37.2, we know that $p(x)$ is primitive and irreducible in $F[x]$, where F is the field of fractions of R.

We also know that $F[x]$ is a UFD (because F is a field, so $F[x]$ is a PID). Hence every irreducible in $F[x]$ is a prime in $F[x]$. Hence $p(x)$ is a prime in $F[x]$.

Since $a(x), b(x) \in F[x]$ and $p(x)|a(x)b(x)$, it follows that $p(x)|a(x)$ in $F[x]$ or $p(x)|b(x)$ in $F[x]$. Assume for simplicity that $p(x)|a(x)$ in $F[x]$. This means there exists $g(x) \in F[x]$ such that $a(x) = p(x)g(x)$.

Our goal is to prove $g(x) \in R[x]$, which will imply $p(x)|a(x)$ in $R[x]$. For now, however, we do not know that $g(x) \in R[x]$.

The coefficients of $g(x)$ are fractions. Let d be the product of all the denominators and let $g_1(x) = dg(x)$. Then $d \in R$ and $g_1(x) \in R[x]$ (this is "clearing denominators"). Multiplying the equation $a(x) = p(x)g(x)$ gives

$$da(x) = p(x)g_1(x)$$

where everything is now in $R[x]$ (or R).

Thus $d|p(x)g_1(x)$ in $R[x]$. By Lemma 36.2, d has a factorization $d = uv$ with $u, v \in R$, such that $u|p(x)$ and $v|g_1(x)$. But $p(x)$ is primitive, so u must be a unit, which implies $d|g_1(x)$ (in $R[x]$). Since $g_1(x) = dg(x)$, this means $g(x) \in R[x]$. □

Corollary 38.2. *$\mathbb{Z}[x]$ is a UFD*

Proof. Because $\mathbb{Z}$ is a UFD. □

Corollary 38.3. *If R is a UFD (for example, $\mathbb{Z}$ or any field), then the ring $R[x, y]$ of polynomials over R in two variables is a UFD.*

Proof. Every polynomial in two variables, say $3x^2y + 5xy - 2xy^2 + 4x - y + 2$, can be written as a polynomial in one variable (y) whose coefficients are elements of $R[x]$. For example,

$$3x^2y + 5xy - 2xy^2 + 4x - y + 2 = (-2x)y^2 + (3x^2 + 5x - 1)y + (4x + 2).$$

Hence $R[x, y] = (R[x])[y]$. Since R is a UFD, so is $R[x]$, and hence so is $(R[x])[y]$ by two applications of the Theorem. □

Obviously we can repeat this to show that $R[x_1, \ldots, x_n]$ is a UFD for any n.