A3 Rings and Modules

Richard Earl

Hilary Term 2019

Syllabus

Recap on rings (not necessarily commutative or with an identity) and examples: $\mathbb{Z}$, fields, polynomial rings (in more than one variable), matrix rings. Zero-divisors, integral domains. Units. The characteristic of a ring. Discussion of fields of fractions and their characterization (proofs non-examinable) [2]

Homomorphisms of rings. Quotient rings, ideals and the first isomorphism theorem and consequences, e.g. Chinese remainder theorem. Relation between ideals in R and R/I. Prime ideals and maximal ideals, relation to fields and integral domains. Examples of ideals. Application of quotients to constructing fields by adjunction of elements; examples to include $\mathbb{C} \cong \mathbb{R}[x]/\langle x^2+1\rangle$ and some finite fields. Degree of a field extension, the tower law. [4]

Euclidean Domains. Examples. Principal Ideal Domains. EDs are PIDs. Unique factorisation for PIDs. Gauss's Lemma and Eisenstein's Criterion for irreducibility. [3]

Modules: Definition and examples: vector spaces, abelian groups, vector spaces with an endomorphism. Submodules and quotient modules and direct sums. The first isomorphism theorem. [2]

Row and column operations on matrices over a ring. Equivalence of matrices. Smith Normal form of matrices over a Euclidean Domain. [1.5]

Free modules and presentations of finitely generated modules. Structure of finitely generated modules of a Euclidean domain. [2]

Application to rational canonical form and Jordan normal form for matrices, and structure of finitely generated Abelian groups. [1.5]

Recommended Texts

- M. E. Keating, First Course in Module Theory (Imperial Press, 1998) (Possibly out of print, but many libraries should have it. Covers much of the course.)
- Joseph Gallian, Contemporary Abstract Algebra (9th edition, CENGAGE 2016) (Excellent text covering material on groups, rings and fields).
- B. Hartley, T. O. Hawkes, Chapman and Hall, Rings, Modules and Linear Algebra. (Possibly out of print, but many libraries should have it. Relatively concise and covers all the material in the course).
- Neils Lauritzen, Concrete Abstract Algebra, CUP (2003) (Excellent on groups, rings and fields, and covers topics in the Number Theory course also. Does not cover material on modules).
- Michael Artin, Algebra (2nd ed. Pearson, (2010). (Excellent but highly abstract text covering everything in this course and much more besides).

EXAMPLES OF ALGEBRAIC STRUCTURES

RINGS

$\mathbb{Z}$ the integers under $+$ and $\times$.
$\mathbb{Z}_n$ – the integers, modulo n, under $+$ and $\times$.
$\mathbb{H}$ – the quaternions under $+$ and $\times$.
$R[x]$ – polynomials in x with coefficients in the ring R under $+$ and $\times$.
$R[[x]]$ – formal power series in x with coefficients in the ring R under $+$ and $\times$.
$R[A]$ – polynomials in a square matrix A with entries in a CRI R.
$R^S = \{f : S \to R\}$ – the set of maps from a set S to a ring R under pointwise $+$ and $\times$.
$C(\mathbb{R})$ – the ring of continuous functions $f : \mathbb{R} \to \mathbb{R}$ under pointwise $+$ and $\times$.
$C^1(\mathbb{R})$ – the ring of continuously differentiable $f : \mathbb{R} \to \mathbb{R}$ under pointwise $+$ and $\times$.
$\mathbb{Z}[i]$ – the Gaussian integers, i.e. the subring of complex numbers $a + bi$ where $a, b \in \mathbb{Z}$.
$M_n(F)$ – the $n \times n$ matrices with entries in the field F under matrix addition and mutliplication.
$\mathrm{End}(V)$ – endomorphisms (i.e. linear maps $V \to V$) of a vector space V under $+$ and $\circ$.
$\mathcal{P}(X)$ – the power set of X under Δ and $\cap$.
R/I – the quotient (or factor) ring of a ring R by an ideal I of R.

FIELDS

$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ – the rationals/reals/complex numbers under $+$ and $\times$.
$\mathbb{Z}_p$ – the integers modulo p (a prime) under $+$ and $\times$.
$\mathbb{Q}[i]$ – $\{q_1 + iq_2 : q_1, q_2 \in \mathbb{Q}\}$ under $+$ and $\times$.
$\mathbb{Q}[\sqrt{2}]$ – $\{q_1 + q_2\sqrt{2} : q_1, q_2 \in \mathbb{Q}\}$.
$F(x)$ – rational functions in x with coefficients in the field F.
$\mathbb{F}_q$ – the finite field with q elements; $q = p^n$ for a prime p and $n \geqslant 1$.

1. A RECAP ON RINGS

The material in this chapter that is covered in A0 Linear Algebra will be only briefly revisited.

Definition 1 *A **ring** $(R, +, \times)$ consists of a set R with $+$ and $\times$ being two binary operations on R*

$$\begin{aligned} + &: R \times R \to R, \qquad (a, b) \mapsto a + b; \\ \times &: R \times R \to R, \qquad (a, b) \mapsto a \times b, \end{aligned}$$

such that $(R, +)$ is an abelian group, that is:

- $(A1)$ *associativity: for all $a, b, c \in R$ we have $a + (b + c) = (a + b) + c$;*
- $(A2)$ *commutativity: for all $a, b \in R$ we have $a + b = b + a$;*
- $(A3)$ *zero element: there exists $0_R \in R$ such that for all $a \in R$ we have $a+0_R = a = 0_R+a$;*
- $(A4)$ *inverses: for any $a \in R$ there exists $-a \in R$ such that $a + (-a) = (-a) + a = 0$;*

and further that:

- $(M1)$ *for all $a, b, c \in R$ we have $a \times (b \times c) = (a \times b) \times c$;*
- (D) *for all $a, b, c \in R$ we have $a \times (b + c) = a \times b + a \times c$ and $(a + b) \times c = a \times c + b \times c$.*

Notation 2 *We will often suppress the $\times$ symbol and simply write ab for $a \times b$.*

Remark 3 *If the operations $+$ and $\times$ are clear then we will often refer simply to a ring R.*

Definition 4 *We say that a ring R is **commutative** if multiplication is commutative – that is if the following axiom holds*

- $(M2)$ *commutativity: for all $a, b \in R$ we have $ab = ba$.*

Definition 5 *We say that a ring R has a 1 or an **identity** if there is an element $1_R \in R$ such that*

- $(M3)$ *existence of an identity: for all $a \in R$ we have $a \times 1_R = a = 1_R \times a$;*
- (Z) *avoiding collapse: $1_R \neq 0_R$.*

If R has an identity then it is necessarily unique.

Basic rules of algebra following from the ring axioms are:

Proposition 6 *Let R be a ring with a 1 and $a, b \in R$.*
(a) If $a + b = a + c$ then $b = c$.
(b) $-(-a) = a$.
(c) $a0_R = 0_R = 0_R a$.
(d) $-(ab) = (-a)\, b = a\,(-b)$.
(e) $(-1_R)\, a = -a = a\,(-1_R)$.

Proof. Left as exercises. ■

Notation 7 *Most, but not all, of the rings we will study are commutative and also have an identity; we will use the acronym CRI as shorthand for "commutative ring with an identity".*

A ring, then, is a half-way house between a group and a field, in that the second operation $\times$ is not "fully-developed": it need not be commutative, there need not be an identity, and there need not be inverses. In a ring, multiplication is permissible but division may well not be.

Example 8 $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ *with $+$ and $\times$ are all CRIs. In each case $0_R = 0$ and $1_R = 1$.*

Example 9 $\mathbb{Z}_n$ *– the ring of integers modulo n – is a CRI with standard $+$ and $\times$ and $0_{\mathbb{Z}_n} = \bar{0}$ and $1_{\mathbb{Z}_n} = \bar{1}$.*

Example 10 *If R is a ring, then we can consider $R[x]$ the ring of polynomials with coefficients in R. If R is commutative, then so is $R[x]$ and if R has an identity then so does $R[x]$, namely the constant polynomial 1_R. Explicitly, given two polynomials p, q over R, say*

$$p(x) = \sum_i a_i x^i; \qquad q(x) = \sum_i b_i x^i$$

then the sum $p + q$ and product pq in $R[x]$ are defined by

$$p(x) + q(x) = \sum_i (a_i + b_i) x^i; \qquad p(x)q(x) = \sum_k \left(\sum_{i+j=k} a_i b_j \right) x^k.$$

We can more generally define polynomial rings over several variables, $R[x_1, \ldots, x_n]$. These rings can be inductively defined by

$$R[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}][x_n].$$

Example 11 *Note that the rule $\deg(fg) = \deg f + \deg g$ will not apply generally in rings. For example in $\mathbb{Z}_4[x]$ we have*

$$(\bar{2}x + 1)(\bar{2}x + 1) = \bar{4}x^2 + \bar{4}x + \bar{1} = \bar{1}$$

Also in $\mathbb{Z}_8[x]$ we see that the quadratic $x^2 - \bar{1}$ has four roots $\bar{1}, \bar{3}, \bar{5}, \bar{7}$ and has two distinct factorizations

$$x^2 - \bar{1} = (x - \bar{1})(x + \bar{1}) = (x - \bar{3})(x - \bar{5}).$$

Example 12 *Given a ring R and a square matrix A with entries in R, we can consider the ring $R[A]$ with addition and multiplication defined similarly to those in $R[x]$. Recall that if*

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n \qquad \text{then} \qquad p(A) = a_0 I + a_1 A + \cdots + a_n A^n.$$

So with $R = \mathbb{R}$ and

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad R[A] = \mathbb{R}I_2 \cong \mathbb{R}. \\ B &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad R[B] = \left\{ \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} : x, y \in \mathbb{R} \right\}. \\ C &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad R[C] = \left\{ \begin{pmatrix} x & y \\ -y & x \end{pmatrix} : x, y \in \mathbb{R} \right\} \cong \mathbb{C}. \end{aligned}$$

The isomorphism with $\mathbb{C}$ in the third example is shown in Exercise Sheet 1, #2. Note that $C^2 = -I_2$.

Example 13 *Given two rings R and S we can form their **direct sum** $R \oplus S$ which has $R \times S$ as its underlying set, and operations*

$$(r_1, s_1) + (r_2, s_2) = (r_1 +_R r_2, s_1 +_S s_2), \qquad (r_1, s_1) \times (r_2, s_2) = (r_1 \times_R r_2, s_1 \times_S s_2).$$

Example 14 *$C(\mathbb{R})$, the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ forms a CRI under pointwise addition and multiplication with $0_{C(\mathbb{R})}$ and $1_{C(\mathbb{R})}$ being the constant functions 0 and 1. With the same operations and identities, $C^1(\mathbb{R})$, the set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ likewise forms a CRI.*

Example 15 *The even integers form a ring $2\mathbb{Z}$ under $+$ and $\times$ which is commutative but which has no identity.*

Example 16 *The $n \times n$ real matrices $M_n(\mathbb{R})$ form a non-commutative ring with an identity I_n. We can similarly consider $M_n(\mathbb{Q})$, $M_n(\mathbb{C})$ or $M_n(F)$ where F is any field or indeed any ring.*

Example 17 *If V is a vector space then $\mathrm{End}(V)$, the set of linear maps $V \to V$, forms a ring under addition and composition. The zero and identity maps are the additive and multiplicative identities of $\mathrm{End}(V)$.*

Example 18 *The power set $\mathcal{P}(X)$ of a set X, that is, the set of subsets of X, forms a CRI under symmetric difference Δ and intersection $\cap$. That is, for $A, B \subseteq X$ we have*

$$A + B := A \,\Delta\, B = (A \backslash B) \cup (B \backslash A); \qquad A \times B := A \cap B.$$

We further have $0_{\mathcal{P}(X)} = \varnothing$ and $1_{\mathcal{P}(X)} = X$. The additive inverse of A is itself A. Notice that every element satisfies $A^2 = A$.

If $X = \{x_1, \ldots x_n\}$ then we can identify $\mathcal{P}(X)$ with $\mathbb{Z}_2^n$ by $S \leftrightarrow (e_1, \ldots, e_n)$ with $e_i = \bar{1}$ if and only if $x_i \in S$.

Example 19 *The following* ***sequence spaces*** *all form commutative rings under coordinate-wise addition and multiplication:*

- l^∞ *the space of bounded real sequences;*
- l^1 *the space of real sequences which are absolutely summable;*
- c *the space of convergent real sequences;*
- c_0 *the space of real sequences which converge to* 0*.*

Both l^∞ *and* c *have an identity in the constant sequence* $(1, 1, 1, \ldots)$*.*

Example 20 *The* ***Gaussian integers*** $\mathbb{Z}[i]$ *are the complex numbers* $\{a + bi : a, b \in \mathbb{Z}\}$ *and form a CRI.*

Example 21 *If* R *is a ring and* S *is a set, then we can form a ring* $R^S = \{f : S \to R\}$ *with the set of maps from* S *to* R *by taking* $+$ *and* $\times$ *to be pointwise addition and multiplication.* R^S *has an identity if and only if* R *has and is commutative if and only if* R *is.*

Example 22 *A* ***quaternion*** *is a "four-dimensional" number of the form*

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

where a, b, c, d *are real numbers. The ring of quaternions is denoted* $\mathbb{H}$ *after the mathematician William Rowan Hamilton (1805-1865) who discovered them in 1843. Two quaternions add component-wise as one would expect and multiply according to the rules*

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1.$$

From these rules one can further show that

$$\mathbf{ij} = \mathbf{k} = -\mathbf{ji}, \qquad \mathbf{jk} = \mathbf{i} = -\mathbf{kj}, \qquad \mathbf{ki} = \mathbf{j} = -\mathbf{ik},$$

so we see that $\mathbb{H}$ *is non-commutative. The elements*

$$\{1, -1, \mathbf{i}, -\mathbf{i}, \mathbf{j}, -\mathbf{j}, \mathbf{k}, -\mathbf{k}\}$$

form a group which is denoted Q_8*.*

Definition 23 *Let* R *be a ring. A subset* $S \subseteq R$ *is a* ***subring*** *of* R *if the operations* $+$ *and* $\times$ *restrict to make a ring of* S*. That is,* S *is a subring of* R *if:*

(a) $0_R \in S$*;*
(b) whenever $r_1, r_2 \in S$ *then* $r_1 \pm r_2 \in S$*.*
(c) whenever $r_1, r_2 \in S$ *then* $r_1 r_2 \in S$*.*

Remark 24 *Note that the axioms* $A1, A2, M1, D$ *automatically apply in* S *as they apply to all elements of* R*, including those of* S*.*

Example 25 *(a)* $\mathbb{Z}[x]$ *is a subring of* $\mathbb{Q}[x]$.
(b) $\mathbb{Z}[i]$ *is a subring of* $\mathbb{C}$.
(c) $\mathbb{R}[x]$ *is a subring of* $C(\mathbb{R})$.
(d) $2\mathbb{Z}$ *is a subring of* $\mathbb{Z}$.
(e) $\mathbb{C} = \{a + b\mathbf{i} : a, b \in \mathbb{R}\}$ *is a subring of* $\mathbb{H}$.
(e) $\mathbb{Z}_n$ *is* ***not*** *a subring of* $\mathbb{Z}$ *as it is not a subset of* $\mathbb{Z}$.

Example 26 $\{0_R\}$ *and* R *are always subrings of* R. *Given a subset* X *of* R *the subring generated by* X *is the smallest subring containing* X. *It is the intersection of all subrings which contain* X.

Definition 27 *In a ring* R, *we say that a non-zero element* $a \in R$ *is a* ***zero-divisor*** *if there exist non-zero* $b, c \in R$ *such that* $ab = 0 = ca$. *In a commutative ring, if* a *is a zero-divisor, then we can assume* $b = c$.

Definition 28 *An* ***integral domain*** R *is a CRI which has no zero-divisors.*

Definition 29 *In a ring* R, *with an identity* 1_R, *we say that* $a \in R$ *is a* ***unit*** *if there exists* $b \in R$ *such that* $ab = 1_R = ba$. *If such* b *exists then it is unique and we will denote it* a^{-1}.

Example 30 *(a)* $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ *are integral domains. In* $\mathbb{Z}$ *the units are only* 1 *and* -1 *whilst every non-zero element is a unit in* $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.
(b) $\mathbb{R}[x]$ *is an integral domain and the units are the non-zero constant polynomials.*
(c) In $M_n(\mathbb{R})$, *the non-zero singular matrices are the zero-divisors and the invertible matrices are the units.*
(d) $\mathbb{Z}[i]$ *is an integral domain and the units are* $1, -1, i, -i$ *only.*
(e) In $\mathcal{P}(X)$, *every non-empty proper set is a zero-divisor and the only unit is* X.
(f) In $\mathbb{H}$, *there are no zero divisors (but, not being commutative,* $\mathbb{H}$ *is not an integral domain) and every non-zero element is a unit.*

Proposition 31 ***(Cancellation Law)*** *Let* R *be an integral domain and* $a, b, c \in R$ *with* $a \neq 0$. *If* $ab = ac$ *then* $b = c$.

Proof. As $ab = ac$ then $a(b - c) = 0_R$. As R is an integral domain then a is not a zero-divisor and so $b - c = 0_R$. ■

Proposition 32 *Let* R *be a ring with a* 1. *Then* R^*, *the set of units in* R, *forms a group under multiplication.*

Proof. Suppose now that u, v are units in R. Then

$$\begin{aligned}(uv)(v^{-1}u^{-1}) &= u(vv^{-1})u^{-1} = uu^{-1} = 1;\\ (v^{-1}u^{-1})(uv) &= v^{-1}(u^{-1}u)v = v^{-1}v = 1,\end{aligned}$$

so that uv is a unit. So multiplication is a binary operation on R^* and is associative by $M1$. Further 1_R is the identity in R^*. Also

$$(u^{-1})u = 1 = u(u^{-1})$$

so that u^{-1} is a unit and the inverse of u in R^*. ■

Example 33 *(a)* $\mathbb{Z}^* = \{1, -1\}$.
(b) $(M_n(F))^* = GL(n, F)$ *for any field* F.
(c) $(F[x])^* = F^*$ *when identified with the non-zero constant polynomials.*
(d) $\mathbb{Z}_8^* = \{\bar{1}, \bar{3}, \bar{5}, \bar{7}\}$ *and is isomorphic to* $C_2 \times C_2$.

Proposition 34 *Let* R *be a ring with an identity.*
(a) An element cannot be both a zero-divisor and a unit.
(b) If R *is finite and commutative then every non-zero element is a zero-divisor or a unit.*
(c) More generally, a non-zero element may be neither a zero-divisor or a unit.

Proof. (a) Suppose that $a \in R$ and that b, c, d are non-zero and such that $ab = 0_R = ca$, and $ad = 1_R = da$. Then

$$b = 1_R b = (da)b = d(ab) = d0_R = 0_R$$

which is a contradiction.
(b) For $x \neq 0$ the map $r \mapsto xr$ is either 1-1 or not. If it is 1-1 then it is onto as R is finite, and hence there exists r such that $xr = 1_R$ and we have that x is a unit. If the map is not 1-1 then there exist $r_1 \neq r_2$ such that $xr_1 = xr_2$. Hence $x(r_1 - r_2) = 0$ and we see that x is a zero-divisor.
(c) In $\mathbb{Z}$ we see 2 is neither a unit nor a zero-divisor. ■

In the linear algebra courses, we defined a field to be a triple $(R, +, \times)$ satisfying the field axioms $A1, A2, A3, A4, M1, M2, M3, M4, Z, D$. An integral domain satisfies each of these except (M4): existence of multiplicative inverses. So we can reformulate the field axioms as:

Definition 35 *A* ***field*** *is an integral domain in which every non-zero element is a unit.*

Theorem 36 *A finite integral domain is a field.*

Proof. Let R be a finite integral domain and $a \in R$ with $a \neq 0$. Consider the map $m_a : R \to R$ given by $r \mapsto ar$. By the cancellation law, m_a is 1-1 and hence is onto as R is finite. In particular, there is $r \in R$ such that $ar = 1_R$. Hence a is a unit and R is a field. ■

We know that $\mathbb{Z}_p$, where p is prime, is a finite field. There are in fact examples of other finite fields. Below is an example of a finite field with 4 elements.

Example 37 *The field of 4 elements can be written as follows. Let* $F_4 = \{0, 1, 2, 3\}$ *where* 0 *is the additive identity and* 1 *is the multiplicative identity. It must be the case that, as an additive group,* $F \cong C_2 \times C_2$ *for if* $F \cong C_4$ *we would find we had zero-divisors. And it must be the case that* $F^* \cong C_3$ *as this is the only group of order 3. The addition and multiplication tables for* F *are then*

$+$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

$\times$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

.

Definition 38 *Given any ring R we define its characteristic* $\mathrm{char}R$ *as follows:*

- *If such a positive integer exists, we define the* $\mathrm{char}R$ *to be the smallest positive integer n such that $nr = 0$ for all $r \in R$.*
- *If no such n exists, then we define* $\mathrm{char}R = 0$.

In the event that R has an identity, then $\mathrm{char}R$ *is the additive order of 1_R when this is finite, and 0 when that order is infinite.*

Example 39 *(a)* $\mathrm{char}\mathbb{Z}_n = n$.
(b) $\mathrm{char}\mathbb{Z} = \mathrm{char}\mathbb{Q} = \mathrm{char}\mathbb{R} = \mathrm{char}\mathbb{C} = 0$.
(c) F_4 has characteristic 2.
(d) For the ring R of 4 elements, with $(R, +) \cong C_4$ and multiplication defined by $ab = 0$ for all a, b, then $\mathrm{char}R = 4$.

Proposition 40 *If R is an integral domain then* $\mathrm{char}R$ *is prime or zero.*

Proof. If the additive order n of 1_R is finite then we have

$$\underbrace{1_R + \cdots + 1_R}_{n \text{ times}} = 0_R,$$

and if $n = uv$, where neither u nor v were 1, we'd have

$$\underbrace{1_R + \cdots + 1_R}_{u \text{ times}} \times \underbrace{1_R + \cdots + 1_R}_{v \text{ times}} = 0_R.$$

As R is an ID then we'd have either that

$$\underbrace{1_R + \cdots + 1_R}_{u \text{ times}} = 0_R \qquad \text{or} \qquad \underbrace{1_R + \cdots + 1_R}_{v \text{ times}} = 0_R,$$

both of which would contradict the minimality of n. ■

Proposition 41 *Let F be a field.*
(a) If $\mathrm{char}F = p$, *a prime, then there is a smallest subfield of F isomorphic to $\mathbb{Z}_p$.*
(b) If $\mathrm{char}F = 0$, *then there is a smallest subfield of F isomorphic to $\mathbb{Q}$.*
This subfield is called the ***prime subfield****.*

Proof. In both cases, as 0_F and 1_F both need to be in any subfield, then the smallest subfield of F is the one generated by 1_F. In the case that $\mathrm{char}F = p$ then the map

$$\phi : \mathbb{Z}_p \to F \qquad \text{given by} \qquad \phi(\bar{k}) = \underbrace{1_F + \cdots + 1_F}_{k \text{ times}}$$

is well-defined and an isomorphism onto its image. In the case that $\mathrm{char}F = 0$ then the map

$$\phi : \mathbb{Q} \to F \qquad \text{given by} \qquad \phi\left(\frac{m}{n}\right) = \left(\underbrace{1_F + \cdots + 1_F}_{m \text{ times}}\right)\left(\underbrace{1_F + \cdots + 1_F}_{n \text{ times}}\right)^{-1}$$

is well-defined and an isomorphism onto its image. ■

Proposition 42 *Let R be an integral domain. Then we can create a* ***field of fractions*** *F for R as follows. We define an equivalence relation on R^2 by*

$$(m,n) \sim (M,N) \qquad \Longleftrightarrow \qquad mN = Mn,$$

and define $F = R^2/\sim$ to be the set of $\sim$-equivalence classes with the operations

$$[(m_1,n_1)] + [(m_2,n_2)] = [(m_1n_2 + m_2n_1, n_1n_2)], \qquad [(m_1,n_1)] \times [(m_2,n_2)] = [(m_1m_2, n_1n_2)].$$

Then

(a) F is a well-defined field with $0_F = [(0_R, 1_R)]$ and $1_F = [(1_R, 1_R)]$.

(b) F contains a copy of R as an isomorphic subring, namely the elements $R \equiv \{[(r, 1_R)] : r \in R\}$.

(c) F is uniquely characterized by the "universal property" that any 1-1 homomorphism $\phi : R \to K$ to a field K, can be uniquely extended to a 1-1 homorphism $\tilde{\phi} : F \to K$.

Remark 43 *The above formulae may be a little less surprising when we realize in $\mathbb{Q}$ that*

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1n_2 + m_2n_1}{n_1n_2} \qquad \text{and} \qquad \frac{m_1}{n_1} \times \frac{m_2}{n_2} = \frac{m_1m_2}{n_1n_2}.$$

We can also see for non-zero n_1, n_2 that

$$m_1n_2 = m_2n_1 \qquad \Longleftrightarrow \qquad \frac{m_1}{n_1} = \frac{m_2}{n_2}.$$

Thus the above construction would yield $\mathbb{Q}$ as the field of fractions of $\mathbb{Z}$ and would identify m/n with the equivalence class $[(m,n)]$.

Proof. (Non-examinable) We will not prove the details here, but rather just give a list of all the facts that need checking, none of which are particularly difficult.

(i) $\sim$ is an equivalence relation.

(ii) $+$ and $\times$ are well-defined binary operations on the set of equivalence classes.

(iii) the field axioms hold amongst the equivalence classes with $0_F = [(0_R, 1_R)]$ and $1_F = [(1_R, 1_R)]$.

(iv) $r \mapsto [(r, 1_R)]$ is a 1-1 homomorphism from R to F.

(v) defining $\phi\,[(m,n)] = \phi(m)\phi(n)^{-1}$ is a well-defined 1-1 homomorphism from F to K which extends ϕ.

(vi) For any other field $\tilde{F}$ with this universal property, extending inclusion $\iota : R \to F$ yields an isomorphosm $\tilde{\iota} : \tilde{F} \to F$. ■

Example 44 *(a) The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$.*

(b) The field of fractions of $\mathbb{Z}[x]$ is $\mathbb{Q}(x)$.

(c) The field of fractions of $\mathbb{Z}[\sqrt{2}]$ is $\mathbb{Q}[\sqrt{2}]$.

(d) The field of fractions of $\mathbb{Z}[\pi]$ is $\mathbb{Q}(\pi)$.

2. THE ISOMORPHISM THEOREMS

Definition 45 *Let R be a ring. A non-empty subset $I \subseteq R$ is said to be a* ***ideal*** *of R if*

$$\text{whenever } i_1, i_2 \in I \text{ then } i_1 \pm i_2 \in I; \qquad \text{whenever } i \in I, r \in R \text{ then } ri \in I \text{ and } ir \in I.$$

This is then written $I \lhd R$. In particular, ideals are subrings though the converse is not generally true. [N.B. any text which insists that rings have 1s will then have a convention that ideals are not generally subrings.]

Definition 46 *Let R be a ring and $a \in R$. Then the* ***principal ideal*** *$\langle a \rangle$ generated by a is the smallest ideal to contain a. So*

$$\langle a \rangle = \left\{ \sum_i r_i a s_i : r_i, s_i \in R \right\}.$$

$\langle a \rangle$ is also commonly written as (a). An ideal I of R is said to be ***principal*** *if there exists $a \in R$ such that $I = \langle a \rangle$.*

Example 47 *$\{0_R\}$ and R are always ideals of R.*

Example 48 *Let $R = C(\mathbb{R})$ and $S = \mathbb{R}[x]$. Then S is a subring of R but is not an ideal of R, as $x \in S$, $e^x \in R$ but $xe^x \notin S$.*

Example 49 *Let $R = C(\mathbb{R})$ and $I = \{f \in R : f(0) = 0\}$. Then I is an ideal of R.*

Example 50 *Let R be a ring with a 1_R and $a_1, \ldots, a_k \in R$. Then the* ***ideal generated by*** *$\{a_1, \ldots, a_k\}$ is*

$$I = \langle a_1, \ldots, a_k \rangle = \{r_1 a_1 s_1 + \cdots + r_k a_k s_k : r_i, s_i \in R\}$$

is the smallest ideal of R containing $a_1, \ldots, a_k$.

Example 51 *Let $R = \mathbb{Z}[x]$ and $I = \langle 2, x \rangle$. Then I is an example of an ideal which is not principal. Note*

$$I = \langle 2, x \rangle = \{p(x) \in \mathbb{Z}[x] : p(x) \text{ has an even constant term}\}.$$

If I were principal then, for some $f(x)$, we would have $I = \langle f(x) \rangle = f(x)\mathbb{Z}[x]$ and then $f(x)$ would divide both 2 and x. The only such polynomials are $f(x) = \pm 1$ but $\langle \pm 1 \rangle = \mathbb{Z}[x] \neq I$.

Proposition 52 *The ideals of $\mathbb{Z}$ are $n\mathbb{Z} = \langle n \rangle$ where $n \in \mathbb{Z}$. So every ideal of $\mathbb{Z}$ is principal.*

Proof. For any $n \in \mathbb{Z}$, we have $n\mathbb{Z} = \langle n \rangle \lhd \mathbb{Z}$. Conversely, suppose that $I \lhd \mathbb{Z}$. Then, in particular, I is a subgroup of $\mathbb{Z}$ and $I = n\mathbb{Z}$ for some $n \in \mathbb{Z}$. ■

Proposition 53 *Let R be a CRI. Then R is a field if and only if the only ideals of R are $\{0_R\}$ and R.*

Proof. Suppose that R is a field and $I \lhd R$. If $I \neq \{0_R\}$ then there exists non-zero $a \in I$. As R is a field, then a is a unit and so there exists $b \in R$ such that $ba = 1_R$. Then, for any $r \in R$,

$$(rb)\, a = r \in I$$

and we see $I = R$. Conversely, suppose that the only ideals are $\{0_R\}$ and R and that $a \neq 0$. Then $1_R \in R = \langle a \rangle$ and there exists $r \in R$ such that $ar = 1_R$. So a is a unit and R is a field. ■

Remark 54 *As with the case of normal subgroups for groups, ideals are the natural subsets a ring might be "modulo"-ed by. If we are to introduce new rules into the algebra of a ring R and set $i, j \sim 0_R$ (where $\sim$ is an equivalence relation) then for reasonable algebra in $R/\sim$, we must expect to have*

$$i \pm j \sim 0_R \qquad \textit{and} \qquad ri \sim 0_R \quad \textit{and} \quad ir \sim 0_R \quad \textit{for any } r \in R.$$

Theorem 55 *Let R be a ring and let $I \lhd R$. The **coset** of $r \in R$ is*

$$r + I = \{r + i : i \in I\}.$$

The operations

$$\begin{aligned} (r_1 + I) \oplus (r_2 + I) &= (r_1 + r_2) + I; \\ (r_1 + I) \otimes (r_2 + I) &= (r_1 r_2) + I; \end{aligned}$$

*lead to well-defined binary operations on the set R/I of cosets of I with $(R/I, \oplus, \otimes)$ being known as the **quotient ring**. If R is commutative then R/I is commutative, and if R has an identity then so does R/I with $1_{R/I} = 1_R + I$.*

Notation 56 *The notation $r + I$ will become cumbersome and so we will write $\bar{r} = r + I$. This very much ties in with the notion that R/I is "$R \bmod I$" and that $r + I$ contains all those s such that $r \equiv s \bmod I$.*

Proof. Suppose that $I \lhd R$. Note for $r_1, r_2 \in R$ that

$$\overline{r_1} = \overline{r_2} \qquad \Longleftrightarrow \qquad r_1 - r_2 \in I.$$

So suppose that $\overline{r_1} = \overline{r_2}$ and $\overline{s_1} = \overline{s_2}$. Then

$$(r_1 + s_1) - (r_2 + s_2) = (r_1 - r_2) + (s_1 - s_2) \in I$$

showing that

$$\overline{r_1} \oplus \overline{s_1} = \overline{r_1 + s_1} = \overline{r_2 + s_2} = \overline{r_2} \oplus \overline{s_2}.$$

That is $\oplus$ is well-defined. Similarly

$$r_1 s_1 - r_2 s_2 = r_1 (s_1 - s_2) + s_2 (r_1 - r_2) \in I$$

showing that

$$\overline{r_1} \otimes \overline{s_1} = \overline{r_1 s_1} = \overline{r_2 s_2} = \overline{r_2} \otimes \overline{s_2}$$

and that $\otimes$ is well-defined on the set of cosets R/I.

$(R/I, \oplus, \otimes)$ meets the axioms of a ring because these properties are inherited from the fact that $(R, +, \times)$ is a ring. R/I has a 1 and/or is commutative as and when R has a 1 and/or is commutative. In order to have $0 \neq 1$ in R/I then I needs to be a proper ideal.

$$\begin{aligned}
(A1) &: (\bar{r} \oplus \bar{s}) \oplus \bar{t} = \overline{(r+s)+t} = \overline{r+(s+t)} = \bar{r} \oplus (\bar{s} \oplus \bar{t}).\\
(A2) &: \bar{r} \oplus \bar{s} = \overline{r+s} = \overline{s+r} = \bar{s} \oplus \bar{r}\\
(A3) &: 0_{R/I} = \overline{0_R} = I.\\
(A4) &: -\bar{r} = \overline{-r}\\
(M1) &: (\bar{r} \otimes \bar{s}) \otimes \bar{t} = \overline{(rs)t} = \overline{r(st)} = \bar{r} \otimes (\bar{s} \otimes \bar{t}).\\
(D) &: \bar{r} \otimes (\bar{s} \oplus \bar{t}) = \overline{r(s+t)} = \overline{rs+rt} = (\bar{r} \otimes \bar{s}) \oplus (\bar{r} \otimes \bar{t}).\\
(M2) &: \bar{r} \otimes \bar{s} = \overline{rs} = \overline{sr} = \bar{s} \otimes \bar{r}\\
(M3) &: 1_{R/I} = \overline{1_R} = 1_R + I.
\end{aligned}$$

■

Example 57 *Let $R = \mathbb{Z}$, $I = \langle n \rangle = n\mathbb{Z}$. As a ring, we can naturally identify $\mathbb{Z}/n\mathbb{Z}$ with $\mathbb{Z}_n$.*

Example 58 *Let $R = \mathbb{Q}[x]$ and $I = \langle x^2 \rangle$. Then R/I is not an integral domain as $\bar{x}$ is a zero-divisor (because $\bar{x} \neq \bar{0}$ and $\bar{x}\bar{x} = \bar{0}$). To better understand the ring R/I note by the division algorithm, that every polynomial $p(x)$ can uniquely be written*

$$p(x) = a + bx + x^2 p(x) \quad \text{for some } a, b \in \mathbb{Q}, f(x) \in \mathbb{Q}[x].$$

So every coset in R/I can uniquely be represented as $a + b\bar{x}$ where $a, b \in \mathbb{Q}$. These cosets than add and multiply as

$$\begin{aligned}
(a + b\bar{x}) + (c + d\bar{x}) &= (a+c) + (b+d)\bar{x};\\
(a + b\bar{x})\,(c + d\bar{x}) &= ac + (bc + ad)\bar{x},
\end{aligned}$$

as $\bar{x}^2 = \bar{0}$. From this we can see that the only zero-divisors in R/I are of the form $b\bar{x}$ where $b \neq 0$.

Example 59 *Write down the addition and multiplication tables of R_1 and R_2 where*

$$R_1 = \frac{\mathbb{Z}_2[x]}{\langle x^2 \rangle}, \qquad R_2 = \frac{\mathbb{Z}_2[x]}{\langle x^2 + 1 \rangle}.$$

Show that the rings are isomorphic.

Solution. By the division algorithm, there are four cosets of $\langle x^2 \rangle$ in $\mathbb{Z}_2[x]$ and of $\langle x^2 + 1 \rangle$ in $\mathbb{Z}_2[x]$, and we can represent these cosets in both rings with the elements

$$\{\bar{0}, \bar{1}, \bar{x}, \bar{x} + \bar{1}\}.$$

The addition and mutliplication tables of R_1 are then

$+$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{1}$	$\bar{1}$	$\bar{0}$	$\bar{x}+\bar{1}$	$\bar{x}$
$\bar{x}$	$\bar{x}$	$\bar{x}+\bar{1}$	$\bar{0}$	$\bar{1}$
$\bar{x}+\bar{1}$	$\bar{x}+\bar{1}$	$\bar{x}$	$\bar{1}$	$\bar{0}$

,

$\times$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{x}$	$\bar{0}$	$\bar{x}$	$\bar{0}$	$\bar{x}$
$\bar{x}+\bar{1}$	$\bar{0}$	$\bar{x}+\bar{1}$	$\bar{x}$	$\bar{1}$

,

and those for R_2 are

$+$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{1}$	$\bar{1}$	$\bar{0}$	$\bar{x}+\bar{1}$	$\bar{x}$
$\bar{x}$	$\bar{x}$	$\bar{x}+\bar{1}$	$\bar{0}$	$\bar{1}$
$\bar{x}+\bar{1}$	$\bar{x}+\bar{1}$	$\bar{x}$	$\bar{1}$	$\bar{0}$

,

$\times$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{x}$	$\bar{x}+\bar{1}$
$\bar{x}$	$\bar{0}$	$\bar{x}$	$\bar{1}$	$\bar{x}+\bar{1}$
$\bar{x}+\bar{1}$	$\bar{0}$	$\bar{x}+\bar{1}$	$\bar{x}+\bar{1}$	$\bar{0}$

.

We can see that the first set of tables match the second set of tables when we swap every $\bar{x}$ with $\bar{x}+\bar{1}$ and vice-versa. This is perhaps less of a surprise when we note

$$x^2+1=(x+1)^2 \quad \text{in } \mathbb{Z}_2[x].$$

■

Definition 60 *Let R and S be rings. A **(ring) homomorhism** $\phi\colon R\to S$ is a map satisfying*

$$\phi(r_1+_R r_2)=\phi(r_1)+_S\phi(r_2), \qquad \phi(r_1*_R r_2)=\phi(r_1)*_S\phi(r_2) \qquad \text{for all } r_1,r_2\in R.$$

*If further ϕ is bijective then ϕ is called a **(ring) isomorphism.***

Proposition 61 *Let R and S be rings and $\phi\colon R\to S$ a homomorphism. Let $r\in R$ and n be an integer. Then:*

(a) $\phi(0_R)=0_S$.
(b) if S is an integral domain and R has an identity, then either $\phi\equiv 0$ or $\phi(1_R)=1_S$.
(c) $\phi(nr)=n\phi(r)$.
(d) if $n\geqslant 1$ then $\phi(r^n)=\phi(r)^n$.

Proof. Left as exercises. ■

Example 62 *Let R be a CRI and $a\in R$. Then the map $\phi\colon R[x]\to R$ given by $\phi(p(x))=p(a)$ is a homomorphism because – by definition – we have*

$$(\phi+\psi)(a)=\phi(a)+\psi(a), \qquad (\phi\psi)(a)=\phi(a)\psi(a).$$

Example 63 *In a similar vein, let A be an $n\times n$ real matrix. Then the map $\phi:\mathbb{R}[x]\to\mathbb{R}[A]$, defined by $p(x)\mapsto p(A)$, is a homomorphism.*

Example 64 *Let $n \geqslant 2$ be an integer. Then $\phi\colon \mathbb{Z} \to \mathbb{Z}_n$ given by $x \mapsto \bar{x}$ is a ring homomorphism as*

$$\overline{x+y} = \bar{x} + \bar{y}, \qquad \overline{xy} = \bar{x}\bar{y}.$$

More generally if $I \lhd R$ then the quotient maps $\pi : R \to R/I$ given by $\pi(r) = \bar{r}$ is a homomorphism.

Example 65 *Let $n \geqslant 2$. The only ring homomorphism $\phi\colon \mathbb{Z}_n \to \mathbb{Z}$ is the zero map because $0 = \phi(\bar{0}) = \phi(\bar{n}) = n\phi(\bar{1})$ and so $\phi(\bar{1}) = 0$ and $\phi \equiv 0$.*

Example 66 *The only ring homomorphisms $\phi\colon \mathbb{Z} \to \mathbb{Z}$ are $x \mapsto 0$ for all x and $x \mapsto x$.*

Example 67 *Find all the ring homomorphisms $\phi\colon \mathbb{Q} \to \mathbb{Q}$.*

Solution. If $\phi(1) = 1$ then $\phi(m) = m$ for any integer m. Likewise, for $n > 0$, we have $n\phi(m/n) = \phi(m) = m$ and so $\phi(m/n) = m/n$. Hence the only ring homomorphisms are

$$\phi(x) = 0 \qquad \text{and} \qquad \phi(x) = x.$$

■

Example 68 *A ring homomorphism $\phi\colon \mathbb{Z}_n \to \mathbb{Z}_n$ is of the form $\phi(\bar{x}) = \bar{a}\bar{x}$ where $\bar{a}^2 = \bar{a}$. Hence the ring homomorphisms $\phi\colon \mathbb{Z}_6 \to \mathbb{Z}_6$ are*

$$\phi(\bar{x}) = \bar{0}, \qquad \phi(\bar{x}) = \bar{x}, \qquad \phi(\bar{x}) = \bar{3}\bar{x}, \qquad \phi(\bar{x}) = \bar{4}\bar{x},$$

and the only ring homomorphisms $\phi\colon \mathbb{Z}_p \to \mathbb{Z}_p$ (where p is a prime) are

$$\phi(\bar{x}) = \bar{0}, \qquad \phi(\bar{x}) = \bar{x}.$$

Solution. Certainly any homomorphism $\phi\colon \mathbb{Z}_n \to \mathbb{Z}_n$ is of the form $\phi(\bar{x}) = \bar{a}\bar{x}$ where $\bar{a} = \phi(\bar{1})$. For such maps we immediately have

$$\phi(\overline{x+y}) = \phi(\bar{x}) + \phi(\bar{y}).$$

Futher we need

$$\bar{a} = \phi(\bar{1}) = \phi(\bar{1})\phi(\bar{1}) = \bar{a}^2.$$

Finally if $\bar{a}^2 = \bar{a}$ then we have

$$\phi(\bar{x}\bar{y}) = \bar{a}\bar{x}\bar{y} = (\bar{a}\bar{x})\,(\bar{a}\bar{y}) = \phi(\bar{x})\phi(\bar{y}).$$

We can check that $\bar{0}, \bar{1}, \bar{3}, \bar{4}$ are the only roots of $x\,(x-1) = 0$ in $\mathbb{Z}_6$. In $\mathbb{Z}_p$, which is a field when p is prime, there are no zero-divisors and so 0 and 1 are the only roots of this quadratic.
■

Example 69 *Show that $\mathbb{R}$ is not (ring) isomorphic to $\mathbb{C}$.*

Solution. Suppose that $\phi\colon \mathbb{C} \to \mathbb{R}$ is a ring isomorphism. Then $\phi(1) = 1$ and so

$$\phi(i)^2 = \phi(i^2) = \phi(-1) = -1$$

but there is no element $x \in \mathbb{R}$ which satisfies $x^2 = -1$. ■

Definition 70 *Let $\phi\colon R \to S$ be a homomorphism between rings. Then the **kernel** of ϕ, written $\ker\phi$, is*

$$\ker\phi = \{r \in R : \phi(r) = 0_S\} \subseteq R.$$

*The **image** of ϕ, written $\operatorname{Im}\phi$, is*

$$\operatorname{Im}\phi = \{\phi(r) : r \in R\} \subseteq S.$$

Proposition 71 *Let $\phi\colon R \to S$ be a homomorphism between rings. Then:*
(a) $\ker\phi$ is an ideal of R.
(b) $\operatorname{Im}\phi$ is a subring of S.

Proof. (a) Suppose that $i_1, i_2 \in \ker\phi$ and $r \in R$. Then

$$\phi(i_1 \pm i_2) = \phi(i_1) \pm \phi(i_2) = 0_S \pm 0_S = 0_S;$$
$$\phi(ri_1) = \phi(ri_1) = \phi(r)\phi(i_1) = \phi(r)0_S = 0_S,$$

showing that $i_1 \pm i_2 \in \ker\phi$ and that $ri_1 \in \ker\phi$. Similarly $i_1 r \in \ker\phi$.

(b) Suppose that $s_1, s_2 \in \operatorname{Im}\phi$. Then there exist $r_i \in R$ such that $\phi(r_i) = s_i$ and so

$$s_1 \pm s_2 = \phi(r_1) \pm \phi(r_2) = \phi(r_1 \pm r_2) \in \operatorname{Im}\phi;$$
$$s_1 s_2 = \phi(r_1)\phi(r_2) = \phi(r_1 r_2) \in \operatorname{Im}\phi.$$

■

Example 72 *Let R be a CRI and $a \in R$. The kernel and image of the map $\phi\colon R[x] \to R$ given by $\phi(p(x)) = p(a)$ are*

$$\ker\phi = \langle x - a\rangle; \qquad \operatorname{Im}\phi = R.$$

Note for the former that we cannot presume the division algorithm applies (which would be the case if R were a field) but we can work with identities such as

$$x^n = (x^{n-1} + ax^{n-2} + \cdots + a^{n-2}x + a^{n-1})(x - a) + a^n.$$

Example 73 *The kernel and image of the map $\phi\colon \mathbb{Z} \to \mathbb{Z}_n$ given by $x \mapsto \bar{x}$ are*

$$\ker\phi = n\mathbb{Z}; \qquad \operatorname{Im}\phi = \mathbb{Z}_n.$$

More generally $\pi\colon R \to R/I$ given by $\pi(r) = \bar{r}$ has kernel I and image R/I.

Example 74 *The kernel and image of the map* $\phi\colon \mathbb{R}[x] \to \mathbb{C}$ *given by* $p(x) \mapsto p(i)$ *are*

$$\ker\phi = \langle x^2+1\rangle; \qquad \operatorname{Im}\phi = \mathbb{C}.$$

This is because any real polynomial which has i *as a root also has* $-i$ *as a root (conjugate pairs).*

Example 75 *The kernels and images of the ring homomorphisms* $\phi\colon \mathbb{Z}_6 \to \mathbb{Z}_6$ *are*

ϕ	$\phi(\bar{x}) = \bar{0}$	$\phi(\bar{x}) = \bar{x}$	$\phi(\bar{x}) = \bar{3}\bar{x}$	$\phi(\bar{x}) = \bar{4}\bar{x}$
$\ker\phi$	$\mathbb{Z}_6$	$\{\bar{0}\}$	$\{\bar{0},\bar{2},\bar{4}\}$	$\{\bar{0},\bar{3}\}$
$\operatorname{Im}\phi$	$\{\bar{0}\}$	$\mathbb{Z}_6$	$\{\bar{0},\bar{3}\}$	$\{\bar{0},\bar{2},\bar{4}\}$

Example 76 *Let*

$$X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad Y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \textit{and} \quad Z = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that $X^2 = 0_2$, $Y^2 = I_2$, $Z^2 = -I_2$. *The homomorphisms*

$$\begin{aligned} \alpha &: \mathbb{R}[x] \to \mathbb{R}[X] \quad \textit{given by} \quad \alpha(p(x)) = p(X); \\ \beta &: \mathbb{R}[x] \to \mathbb{R}[Y] \quad \textit{given by} \quad \beta(p(x)) = p(Y); \\ \gamma &: \mathbb{R}[x] \to \mathbb{R}[Z] \quad \textit{given by} \quad \gamma(p(x)) = p(Z), \end{aligned}$$

can be shown to have kernels

$$\ker\alpha = \langle x^2\rangle, \qquad \ker\beta = \langle x^2-1\rangle, \qquad \ker\gamma = \langle x^2+1\rangle.$$

As with the case of groups we then have an equivalent version of the isomorphism theorem.

Theorem 77 ***(First Isomorphism Theorem for Rings)*** *Let* $\phi : R \to S$ *be a homomorphism between CRIs. Then:*

(a) $\ker\phi \lhd R$.

(b) $\operatorname{Im}\phi \leqslant S$.

(c) The map

$$\bar{\phi}\colon \frac{R}{\ker\phi} \to \operatorname{Im}\phi \quad \textit{given by} \quad \bar{\phi}(r+\ker\phi) = \phi(r)$$

is a (ring) isomorphism.

Proof. (a) and (b) were proven earlier. To show (c) we note (in a similar fashion to groups) that

$$\begin{aligned} \bar{\phi}(r_1+\ker\phi) = \bar{\phi}(r_2+\ker\phi) &\iff \phi(r_1) = \phi(r_2) \\ &\iff \phi(r_1-r_2) = 0_S \\ &\iff r_1 - r_2 \in \ker\phi \\ &\iff r_1 + \ker\phi = r_2 + \ker\phi. \end{aligned}$$

The above implications right-to-left show that $\bar{\phi}$ is well-defined and those left-to-right that $\bar{\phi}$ is injective. We also note $\bar{\phi}$ clearly maps onto $\operatorname{Im}\phi$. Finally we note

$$\begin{aligned}\bar{\phi}\left((r_1+\ker\phi)+(r_2+\ker\phi)\right) &= \bar{\phi}\left((r_1+r_2)+\ker\phi\right)\\ &= \phi\left(r_1+r_2\right)\\ &= \phi(r_1)+\phi(r_2)\\ &= \bar{\phi}\left(r_1+\ker\phi\right)+\bar{\phi}\left(r_2+\ker\phi\right)\end{aligned}$$

and

$$\begin{aligned}\bar{\phi}\left((r_1+\ker\phi)\left(r_2+\ker\phi\right)\right) &= \bar{\phi}\left((r_1r_2)+\ker\phi\right)\\ &= \phi\left(r_1r_2\right)\\ &= \phi(r_1)\phi(r_2)\\ &= \bar{\phi}\left(r_1+\ker\phi\right)\bar{\phi}\left(r_2+\ker\phi\right).\end{aligned}$$

■

If we apply the isomorphism theorem to the following homomorphisms we find:

Example 78 *Let $a\in\mathbb{R}$ and $\phi\colon \mathbb{R}[x]\to\mathbb{R}$ where $p(x)\mapsto p(a)$. The isomorphism theorem then says that*

$$\frac{\mathbb{R}[x]}{\langle x-a\rangle}\cong\mathbb{R}.$$

Example 79 *The isomorphism theorem applied to the map $\phi\colon \mathbb{Z}\to\mathbb{Z}_n$ where $n\mapsto\bar{n}$ shows that*

$$\frac{\mathbb{Z}}{n\mathbb{Z}}\cong\mathbb{Z}_n.$$

Example 80 *The isomorphism theorem applied to the map $\phi\colon \mathbb{R}[x]\to\mathbb{C}$ where $p(x)\mapsto p(i)$ shows that*

$$\frac{\mathbb{R}[x]}{\langle x^2+1\rangle}\cong\mathbb{C}.$$

Example 81 *The isomorphism theorem applied to the map $\phi\colon \mathbb{Q}[x]\to\mathbb{Q}[\sqrt{2}]$ where $p(x)\mapsto p(\sqrt{2})$ shows that*

$$\frac{\mathbb{Q}[x]}{\langle x^2-2\rangle}\cong\mathbb{Q}[\sqrt{2}].$$

Example 82 *The isomorphism theorem applied to the map $\phi\colon \mathbb{Z}[x]\to\mathbb{Q}$ where $p(x)\mapsto p\left(\frac{1}{2}\right)$ shows that*

$$\frac{\mathbb{Z}[x]}{\langle 2x-1\rangle}\cong\left\{\frac{m}{2^k}:m\in\mathbb{Z},\,k\geqslant 0\right\}=\mathbb{Z}\left[1/2\right].$$

Example 83 *Let*

$$X=\begin{pmatrix}0&1\\0&0\end{pmatrix}\quad\textit{and}\quad Y=\begin{pmatrix}0&1\\1&0\end{pmatrix}\quad\textit{and}\quad Z=\begin{pmatrix}0&1\\-1&0\end{pmatrix},$$

(as in Example 76) so that

$$\begin{aligned}\mathbb{R}[X] &= \{aI + bX : a, b \in \mathbb{R}\} = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} : a, b \in \mathbb{R} \right\}. \\ \mathbb{R}[Y] &= \{aI + bY : a, b \in \mathbb{R}\} = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}. \\ \mathbb{R}[Z] &= \{aI + bZ : a, b \in \mathbb{R}\} = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.\end{aligned}$$

The isomorphism theorem applied to the maps

$$\alpha(p(x)) = p(X), \qquad \beta(p(x)) = p(Y), \qquad \gamma(p(x)) = p(Z),$$

shows that

$$\left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} : a, b \in \mathbb{R} \right\} \cong \frac{\mathbb{R}[x]}{\langle x^2 \rangle},$$

that

$$\left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\} \cong \frac{\mathbb{R}[x]}{\langle x^2 - 1 \rangle}$$

which further we know to be isomorphic to $\mathbb{R}^2$ *(as in Sheet 1, Question 7), and that*

$$\left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{R} \right\} \cong \frac{\mathbb{R}[x]}{\langle x^2 + 1 \rangle} \cong \mathbb{C},$$

(as in Sheet 1, Question 2).

We conclude this discussion with an important application of the isomorphism theorem, namely the Chinese remainder theorem. Specific instances of its use date back as far as the 3rd century, and a general alogorithm for its solution in the integers to the 6th century. Historically it applies to problems such as the following.

Example 84 *Say that an integer satisfies* $x \equiv 3 \bmod 7$ *and* $x \equiv 7 \bmod 13$*, what is* $x \bmod 91$*?*

Solution. A slightly ad hoc approach is to note that if $x \equiv r \bmod 91$, where $0 \leqslant r < 91$, then $x = r + 91k$ and so

$$r \bmod 13 = x \bmod 13 = 7 \quad \Longrightarrow \quad r = 7, 20, 33, 46, 59, 72, 85.$$

But these numbers r satisfy

$$7, 20, 33, 46, 59, 72, 85 \equiv 0, 6, 5, 4, 3, 2, 1 \bmod 7$$

so that we see $r = 59$ is the only one satisfying $r \equiv 3 \equiv \bmod 7$. So we see $x \equiv 59 \bmod 91$. ■

More generally the Chinese remainder theorem is a means of decomposing a quotient ring $R/(I \cap J)$ where I and J are *coprime* ideals.

Definition 85 *Given ideals I and J of a ring R, then we can define the following ideals in terms of them.*

(a) Their ***sum*** $I+J$ *equals*

$$I+J=\{i+j : i\in I,\, j\in J\}.$$

We say that I and J are ***coprime*** *if $I+J=R$.*

(b) Their ***intersection*** $I\cap J$ *equals*

$$I\cap J=\{r : r\in I \quad and \quad r\in J\}.$$

(c) Their ***product*** IJ *equals*

$$IJ=\left\{\sum_{k=1}^{n} i_k j_k : i_k\in I,\, j_k\in I\right\}.$$

Example 86 *If m and n are integers, note that*

$$\begin{aligned}\langle m\rangle+\langle n\rangle &= \langle h\rangle \qquad where\ \ h=\mathrm{hcf}(m,n).\\ \langle m\rangle\cap\langle n\rangle &= \langle l\rangle \qquad where\ \ l=\mathrm{lcm}(m,n).\\ \langle m\rangle\langle n\rangle &= \langle mn\rangle.\end{aligned}$$

Note that $\langle m\rangle$ and $\langle n\rangle$ are coprime ideals if and only if m and n are coprime integers.

Theorem 87 ***(Chinese Remainder Theorem)*** *Let I and J be coprime ideals of a ring R with an identity 1_R. Then the map*

$$\phi\colon \frac{R}{I\cap J}\to \frac{R}{I}\oplus\frac{R}{J} \qquad given\ by \qquad r+I\cap J\mapsto (r+I, r+J)$$

is a well-defined isomorphism.

Proof. Firstly suppose that $r_1+I\cap J=r_2\cap I+J$. Then $r_1-r_2=k$ for some integer k in both I and J. Hence $r_1+I=r_2+I$ and $r_1+J=r_2+J$. Hence ϕ is well-defined. Further

$$\begin{aligned}\phi\left((r_1+r_2)+I\cap J\right) &= ((r_1+r_2)+I,(r_1+r_2)+J)\\ &= (r_1+I,r_1+J)+(r_2+I,r_2+J)\\ &= \phi\left(r_1+I\cap J\right)+\phi\left(r_2+I\cap J\right).\end{aligned}$$

Similarly

$$\begin{aligned}\phi\left(r_1r_2+I\cap J\right) &= (r_1r_2+I,r_1r_2+J)\\ &= (r_1+I,r_1+J)\,(r_2+I,r_2+J)\\ &= \phi\left(r_1+I\cap J\right)\phi\left(r_2+I\cap J\right),\end{aligned}$$

and so ϕ is a ring homorphism. Finally we need to check that ϕ is 1-1 and onto. If

$$\phi\left(r_1+I\cap J\right)=\phi\left(r_2+I\cap J\right)$$

then $r_1 + I = r_2 + I$ and $r_1 + J = r_2 + J$. So $r_1 - r_2$ lies in both I and J and we have

$$r_1 + I \cap J = r_2 + I \cap J.$$

Hence ϕ is 1-1. Finally as I and J are coprime ideals then we know that there exist $i_0 \in I$ and $j_0 \in J$ such that $i_0 + j_0 = 1_R$. So for any $r_1, r_2 \in R$ we have

$$\begin{aligned}
\phi(r_1 j_0 + r_2 i_0 + I \cap J) &= (r_1 j_0 + r_2 i_0 + I, r_1 j_0 + r_2 i_0 + J) \\
&= (r_1 j_0 + I, r_2 i_0 + J) \\
&= (r_1(1 - i_0) + I, r_2(1 - j_0) + J) \\
&= (r_1 + I, r_2 + J)
\end{aligned}$$

and we see ϕ is also onto. [Note in the above how i_0 corresponds to $(0, 1)$ and j_0 corresponds to $(1, 0)$ under ϕ.] ■

Example 88 *Determine the inverse* $\mathbb{Z}_{11} \times \mathbb{Z}_{13} \to \mathbb{Z}_{143}$ *of the map*

$$r \bmod 143 \mapsto (r \bmod 11, r \bmod 13).$$

Solution. We see from the above proof that we first need to determine integers u and v such that $11u + 13v = 1$. By inspection we see that $u = 6$ and $v = -5$ work. Under the above map

$$11u = 66 \mapsto (0, 1), \qquad 13v = -65 \mapsto (1, 0)$$

so that

$$-65x + 66y \mapsto (x, y)$$

and the map's inverse is given by

$$(x \bmod 11, y \bmod 13) \mapsto 66y - 65x \bmod 143.$$

■

Example 89 *Note*

$$\frac{\mathbb{R}[x]}{\langle x^3 - x^2 + x - 1\rangle} = \frac{\mathbb{R}[x]}{\langle (x-1)(x^2+1)\rangle} = \frac{\mathbb{R}[x]}{\langle x - 1\rangle \cap \langle x^2 + 1\rangle} \cong \frac{\mathbb{R}[x]}{\langle x - 1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x^2 + 11\rangle} \equiv \mathbb{R} \oplus \mathbb{C}.$$

Example 90 *How many units are there in* $\mathbb{Z}_{120}$*?*

Solution. Using the Chinese remainder theorem note

$$\mathbb{Z}_{120} = \frac{\mathbb{Z}}{120\mathbb{Z}} \cong \frac{\mathbb{Z}}{8\mathbb{Z} \cap 15\mathbb{Z}} \cong \frac{\mathbb{Z}}{8\mathbb{Z}} \oplus \frac{\mathbb{Z}}{(3\mathbb{Z} \cap 5\mathbb{Z})} \equiv \mathbb{Z}_8 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_5.$$

There are 4 units in $\mathbb{Z}_8$, 2 units in $\mathbb{Z}_3$ and 4 units in $\mathbb{Z}_5$ and so there are $4 \times 2 \times 4 = 32$ units in $\mathbb{Z}_{120}$. In fact we can see that

$$\mathbb{Z}_{120}^* \cong \mathbb{Z}_8^* \times \mathbb{Z}_3^* \times \mathbb{Z}_5^* \cong C_2 \times C_2 \times C_2 \times C_4.$$

■

3. MORE ON IDEALS AND QUOTIENT RINGS. FIELD EXTENSIONS.

We now move from the isomorphism theorem to the more general connection between ideals and quotient rings.

Proposition 91 *Let R be a ring and I an ideal of R. Then there is a 1-1 correspondence between the ideals of R that contain I and the ideals of R/I.*

Proof. Let $\pi\colon R \to R/I$ denote the quotient map and let J be an ideal of R which contains I. Then I is an ideal of J as well as it is still closed under addition and multiplication by elements of J. Further $\bar{J} = \pi(J) = J/I$ is an ideal of R/I. Conversely say that $\bar{J}$ is an ideal of R/I. If we define

$$J = \pi^{-1}(\bar{J}) = \left\{ r \in R : \bar{r} \in \bar{J} \right\},$$

then this is an ideal of R and contains I as $\bar{0} \in \bar{J}$. The above maps $J \mapsto \pi(J)$ and $\bar{J} \mapsto \pi^{-1}(\bar{J})$ are inverse processes of one another and the result follows. ■

Example 92 *The ideals of $\mathbb{Z}_6 \cong \mathbb{Z}/6\mathbb{Z}$ are*

$$\{\bar{0}\}, \qquad \{\bar{0}, \bar{2}, \bar{4}\}, \qquad \{\bar{0}, \bar{3}\}, \qquad \mathbb{Z}_6.$$

These correspond to the ideals $6\mathbb{Z}, 2\mathbb{Z}, 3\mathbb{Z}, \mathbb{Z}$ of $\mathbb{Z}$ which contain $6\mathbb{Z}$.

Definition 93 *(a) A proper ideal I of a ring R is said to be **prime** if whenever $ab \in I$ then $a \in I$ or $b \in I$ (or both).*

*(b) A proper ideal I of a ring R is said to be **maximal** if the only ideal of R to strictly contain I is R itself. (Note therefore that R is not maximal.)*

Proposition 94 *Let I be a proper ideal of a CRI R. Then*

(a) I is prime if and only if R/I is an integral domain.

(b) I is maximal if and only if R/I is a field.

Solution. (a) Assume I is prime and say that $\bar{r}\bar{s} = \bar{0}$. Then $rs \in I$ and so $r \in I$ or $s \in \dot{I}$. But this is equivalent to $\bar{r} = \bar{0}$ or $\bar{s} = \bar{0}$.

Conversely, say that R/I is an integral domain. If $rs \in I$ then $\overline{rs} = \bar{r}\bar{s} = \bar{0}$ and, as R/I is an integral domain, then $\bar{r} = \bar{0}$ or $\bar{s} = \bar{0}$ or equivalentlt $r \in I$ or $s \in I$. Hence I is prime.

(b) Say that I is maximal. Say that $\bar{r} \neq \bar{0} \in R/I$. Then the ideal $J = \langle I, r \rangle$ strictly contains I and so by maximality $J = R$. Consequently we can write $1 = sr + i$ for some $s \in R$ and $i \in I$. We then have $\bar{s}\bar{r} = \bar{1}$. So $\bar{r}$ is a unit in R/I and R/I is a field.

Conversely say that R/I is a field. Then the only ideals are $\{\bar{0}\}$ and R/I. These correspond (under Proposition 91) to I and R being the only ideals of R containing I and hence I is maximal. ■

Corollary 95 *Maximal ideals are prime. (This is something that can also be proved directly in a relatively straightforward manner.)*

Example 96 *The prime ideals of $\mathbb{Z}$ are $p\mathbb{Z}$ where p is prime or $p = 0$.*
The maximal ideals of $\mathbb{Z}$ are $p\mathbb{Z}$ where p is prime.

Example 97 *(See Sheet 1, #7) The ideal $\langle x^2 - 1\rangle$ is not prime in $\mathbb{Q}[x]$ as $\mathbb{Q}[x]/\langle x^2 - 1\rangle \cong \mathbb{Q}^2$ is not an integral domain.*
The ideal $\langle x^2 - 2\rangle$ is maximal in $\mathbb{Q}[x]$ as $\mathbb{Q}[x]/\langle x^2 - 2\rangle \cong \mathbb{Q}[\sqrt{2}]$ is a field.

Example 98 *(See Sheet 1, #6) The ideal $\langle x^2+x+1\rangle$ is maximal in $\mathbb{Z}_2[x]$ as $\mathbb{Z}_2[x]/\langle x^2+x+1\rangle$ is a field.*

Proposition 99 *Let F be a field.*
(a) Every ideal of $F[x]$ is principal.
(b) An ideal $\langle p(x)\rangle$ is maximal if and only if $p(x)$ is irreducible (that is $f(x)$ cannot be written as a product of two non-constant polynomials).

Proof. (a) Let I be an ideal of $F[x]$. If $I = \{0\}$ then we have $I = \langle 0\rangle$. Otherwise there is a non-zero polynomial $f(x)$ in I of least degree. By the division algorithm, for any $g(x)$ in I we can write

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

As $r = g - qf \in I$ then we must have $r = 0$ and so $g(x) = q(x)f(x) \in \langle f(x)\rangle$. But $\langle f(x)\rangle \subseteq I$ by definition and hence $I = \langle f(x)\rangle$.

(b) Say that $p(x)$ is irreducible. Then for any $q(x) \notin \langle p(x)\rangle$, $p(x)$ and $q(x)$ are coprime and so by Bézout's lemma there exist $u(x), v(x) \in F[x]$ such that

$$u(x)p(x) + v(x)q(x) = 1.$$

But then $\langle p(x), q(x)\rangle = F[x]$ and so $\langle p(x)\rangle$ is maximal. Conversely say that $p(x)$ is not irreducible and $p(x) = a(x)b(x)$ where neither a nor b is a constant polynomial. Then $\langle p(x)\rangle$ is strictly contained in $\langle a(x)\rangle \neq F[x]$ and so $\langle p(x)\rangle$ is not maximal. ■

Proposition 100 *In an integral domain R in which every ideal is principal, then non-zero prime ideals are maximal.*

Proof. Let $I \lhd R$ with I prime. Then there exists $x \neq 0$ such that $I = \langle x\rangle$. For $r \notin I$ then $\langle x, r\rangle = \langle y\rangle$ for some y as every ideal is principal. This in particular means that $x = yu$. As $yu \in I$ and I is prime then either $y \in I$ or $u \in I$. Were it the case that $y \in I = \langle x\rangle$ then we'd have $y = xv$ for some v and so $x = xvu$; by the cancellation law u would be a unit and we'd have $I = \langle y\rangle = \langle x, r\rangle$ which is a contradiction as $\langle x, r\rangle$ strictly contains I. So $u \in I$ and y is a unit and hence $\langle x, r\rangle = R$. We have therefore shown that I is maximal. ■

Example 101 *As $x^3 + x + 1$ has no roots in $\mathbb{Z}_2$ then the polynomial is irreducible. (Any reduction of a cubic polynomial involves a linear factor.) So*

$$\mathbb{F}_8 = \frac{\mathbb{Z}_2[x]}{\langle x^3 + x + 1 \rangle}$$

is a field with 8 elements. Every coset has a representative $a + bx + cx^2$ where $a, b, c \in \mathbb{Z}_2$.

Similarly $x^2 + 1$ is irreducible over $\mathbb{Z}_3$ and

$$\mathbb{F}_9 = \frac{\mathbb{Z}_3[x]}{\langle x^2 + 1 \rangle}$$

is a field with 9 elements.

Note in the first case that we can identify $\mathbb{Z}_2$ with $\{0, 1\} \subseteq \mathbb{F}_8$, its prime subfield and similarly we can $\mathbb{Z}_3$ with $\{0, 1, 2\} \subseteq \mathbb{F}_9$.

The above is part of a more general approach to *adjoining a root.*

Proposition 102 *Let F be a field and $f(x)$ be an irreducible polynomial in $F[x]$. Then*

$$K = \frac{F[x]}{\langle f(x) \rangle}$$

contains an isomorphic copy of F and there is a root of $f(x)$ in K.

Proof. $\langle f(x) \rangle$ is maximal and so K is a field. K contains an isomorphic copy of F when we identify F with the cosets

$$a \leftrightarrow a + \langle f(x) \rangle = \bar{a}.$$

Consequently we can make sense of $p(k)$ where $k \in K$ and $p(x) \in F[x]$. In particular we have that

$$f(\bar{x}) = f(x + \langle f(x) \rangle) = f(x) + \langle f(x) \rangle = \langle f(x) \rangle = 0_K$$

and we see that f has a root $\bar{x}$ in K. ■

Example 103 *The polynomial $x^2 + x + 1$ is irreducible in $\mathbb{Z}_2[x]$. If we set*

$$\mathbb{F}_4 = \frac{\mathbb{Z}_2[y]}{\langle y^2 + y + 1 \rangle}$$

then we note that y is a root of $x^2 + x + 1$ in $\mathbb{F}_4$. We then note

$$x^2 + x + 1 = (x - y)(x + y + 1)$$

as expansion gives

$$x^2 + (y + 1 - y)x + (-y^2 - y) = x^2 + x + 1.$$

Example 104 *The polynomial $x^3 - 2$ is irreducible over $\mathbb{Q}$. If we set $\alpha = \sqrt[3]{2}$ then we see in*

$$K = \mathbb{Q}[\alpha] = \left\{q_1 + q_2\alpha + q_3\alpha^2 : q_i \in \mathbb{Q}\right\}$$

that in $K[x]$ we can factorize as

$$x^3 - 2 = x^3 - \alpha^3 = (x - \alpha)(x^2 + \alpha x + \alpha^2).$$

However we cannot factorize further over K – the other two roots are complex and K is a subfield of $\mathbb{R}$.

Definition 105 *Let F be a field. A **field extension** K of F is a field which contains F as a subfield. We denote this extension as $K\colon F$.*

*K can then be considered as a vector space over F. The **degree** of the field extension is denoted $|K\colon F|$ and equals $\dim_F K$.*

Example 106 *(a) $\mathbb{C}$ is a field extension of $\mathbb{R}$ of degree 2. A basis of $\mathbb{C}$ over $\mathbb{R}$ is $\{1, i\}$.*

(b) $\mathbb{R}$ is a field extension of $\mathbb{Q}$ of infinite degree. To see this note that any finite degree field extension of $\mathbb{Q}$ is isomorphic as a vector space to $\mathbb{Q}^n$ for some n and in particular is finite. As $\mathbb{R}$ is uncountable this cannot be the case.

(c) $\mathbb{Q}[\sqrt{2}]$ is a field extension of $\mathbb{Q}$ of degree 2. A basis of $\mathbb{Q}[\sqrt{2}]$ over $\mathbb{R}$ is $\{1, \sqrt{2}\}$.

(d) $\mathbb{Q}[\sqrt[3]{2}]$ is a field extension of $\mathbb{Q}$ of degree 3. A basis of $\mathbb{Q}[\sqrt[3]{2}]$ over $\mathbb{R}$ is $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$.

(e) $\mathbb{F}_8 = \mathbb{Z}_2[x]/\langle x^3 + x + 1\rangle$ is a field extenison over $\mathbb{Z}_2$ of degree 3 with a basis being $\{1, x, x^2\}$.

(f) $\mathbb{F}_9 = \mathbb{Z}_3[x]/\langle x^2 + 1\rangle$ is a field extenison over $\mathbb{Z}_3$ of degree 2 with a basis being $\{1, x\}$.

Proposition 107 *A finite field has order p^n for some prime p and a positive integer n.*

Proof. Let F be a finite field. As a field of characteristic 0 contains a prime subfield isomorphic to $\mathbb{Q}$ then F has characteristic p for some prime p and contains a copy of $\mathbb{Z}_p$ as its prime subfield. F is the a field extension of some degree n over its prime subfield and so F is isomorphic to $\mathbb{Z}_p^n$ *as a vector space.* Hence $|F| = \left|\mathbb{Z}_p^n\right| = p^n$. ■

Remark 108 ***(Off-syllabus)*** *There exists a field of order p^n for each prime p and positive integer n. One can show this by demonstrating that, for each n, there is an irreducible polynomial over $\mathbb{Z}_p$ of degree n. It can further be shown that, up to isomorphism, there is a unique field $\mathbb{F}_{p^n}$ of order p^n. Finite fields are commonly called **Galois fields** in honour of Évariste Galois who first proved these results. The multiplicative group of any finite field is cyclic though it can be computationally difficult to find a generator, and consequently finite fields can be used in this way in crytopgraphy.*

Example 109 *Give an example of a field $\mathbb{F}_{16}$ with 16 elements. How many generators does its multiplicative group have?*

Solution. We can produce such a field by finding an irreducible quartic over $\mathbb{Z}_2$. Consider x^4+x+1. Certainly neither 0 nor 1 are roots; it remains to show that x^4+x+1 cannot be written as a product of two irreducible quadratics. The only irreducible quadratic over $\mathbb{Z}_2$ is x^2+x+1 and we note that

$$(x^2+x+1)^2 = x^2+2x^3+3x^2+2x+1 = x^4+x^2+1,$$

which is not equal to our polynomial. Hence

$$\mathbb{F}_{16} = \frac{\mathbb{Z}_2[x]}{\langle x^4+x+1\rangle}$$

is a field of 16 elements. Given what we've been told above, $\mathbb{F}_{16}^* \cong \mathbb{Z}_{15}$ which has 8 generators, $1,2,4,7,8,11,13,14$. Note in $\mathbb{F}_{16}$ that $x^5 = x^2+x \neq 1$ and so x does not have order 3 or 5 and so must have order 15. So the 8 generators of $\mathbb{F}_{16}^*$ are

$$x, \quad x^2, \quad x^4, \quad x^7, \quad x^8, \quad x^{11}, \quad x^{13}, \quad x^{14},$$

whose cosets have "preferred" representatives

$$x, \quad x^2, \quad x+1, \quad x^3+x+1, \quad x^2+1, \quad x^3+x^2+x, \quad x^3+x^2+1, \quad x^3+1.$$

■

Definition 110 *Given a field extension $K\colon F$ and $\alpha \in K$, we say that α is **algebraic** over F if there exists a polynomial $f(x) \in F[x]$ such that $f(\alpha) = 0$.*

Proposition 111 *Let $K\colon F$ be a field extension and $\alpha \in K$ be algebraic over F.*

*(a) There exists a unique monic polynomial $m(x)$ in $F[x]$ of least degree such that $m(\alpha) = 0$. The polynomial $m(x)$ is known as the **minimal polynomial** of α.*

(b) For $f(x) \in F[x]$, we have $f(\alpha) = 0$ if and only if $m(x)$ divides $f(x)$. Further $m(x)$ is irreducible.

(c) $F[\alpha]$ is a subfield of K of degree equal to the degree of $m(x)$.

Proof. (a) As there exists a polynomial $f(x)$ such that $f(\alpha) = 0$ then there exists such a polynomial $m(x)$ of least degree. By dividing by its leading coefficient if necessary we can assume $m(x)$ to be monic. If $m_1(x)$ and $m_2(x)$ were monic polynomials of least degree such that $m_1(\alpha) = m_2(\alpha) = 0$ then $m_1 - m_2$ would be a polynomial of strictly lower degree with α as a root – a contradiction.

(b) By the division algorithm we can write $f(x) = q(x)m(x) + r(x)$ where $\deg r < \deg m$ or $r = 0$. We then have $0 = f(\alpha) = r(\alpha)$ and by the minimality of m's degree it must be that $r = 0$. Hence $f(x) = q(x)m(x)$ and we see $m(x)$ divides $f(x)$. The converse is obvious. Finally if $m(x)$ were reducible as $m(x) = g(x)h(x)$ then we'd have $g(\alpha)h(\alpha) = m(\alpha) = 0$. As K is a field, and so an integral domain, $g(\alpha) = 0$ or $h(\alpha) = 0$ and either would contradict the minimality of m.

(c) We already Certainly $F[\alpha]$ is a CRI, it only remains to show that non-zero elements are units. Say that $g(\alpha) \in F[\alpha]$ and $g(\alpha) \neq 0$. Then $m(x)$ does not divide $g(x)$ and as $m(x)$ is

irreducible then it follows that $g(x)$ and $m(x)$ are coprime. By Bézout's Lemma in $F[x]$ we know that there are polynomials $u(x)$ and $v(x)$ such that

$$u(x)g(x) + v(x)m(x) = 1.$$

But then $u(\alpha)g(\alpha) = 1$ and we see that $g(\alpha)$ is a unit. Finally, if $n = \deg m(x)$ and $f(\alpha) \in F[\alpha]$ we see from the division algorithm that $f(\alpha) = r(\alpha)$ for some $r(\alpha)$ in the span of $1, \alpha, \ldots, \alpha^{n-1}$. Further $1, \alpha, \ldots, \alpha^{n-1}$ are independent by the minimality of $m(x)$ and so a basis for $F[\alpha]$ over F. Hence $n = |F[\alpha] \colon F|$ as well. ■

Example 112 *Let $\mathbb{F}_{16}$ be as in Example 109. Find the minimal polynomials of (i) x, (ii) $x+1$, (iii) $x^2 + x$?*

Solution. (i) We know that x is a root of satisfies $y^4 + y + 1$ which is irreducible, so this must be its minimal polynomial.

(ii) We similarly then have that $x + 1$ satisfies

$$(y+1)^4 + (y+1) + 1 = y^4 + 4y^3 + 6y^2 + 4y + 1 + y + 1 + 1 = y^4 + y + 1$$

which again is irreducible and so $x + 1$ has this as its minimal polynomial.

(iii) Note with $\alpha = x^2 + x$ we have

$$\alpha^2 = x^4 + 2x^3 + x^2 = x + 1 + 0 + x^2 = \alpha + 1.$$

Hence the minimal polynomial of α is $y^2 + y + 1$. In particular this means that

$$\mathbb{Z}_2[\alpha] = \{0, 1, \alpha, \alpha + 1\} = \{0, 1, x^2 + x, x^2 + x + 1\}$$

is a subfield of $\mathbb{F}_{16}$. ■

Proposition 113 ***(Tower Law)*** *Let L, K, F be fields with $F \subseteq K \subseteq L$. Then L has finite degree over F if and only if $|L \colon K|$ and $|K \colon F|$ are finite. In this case*

$$|L \colon F| = |L \colon K| \, |K \colon F| \, .$$

Proof. Say that $|L \colon K| = m$ and $|K \colon F| = n$ are finite and $l_1, \ldots, l_m$ are a basis for L over K and $k_1, \ldots, k_n$ are a basis for K over F. Then we will show that

$$\{l_i k_j : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$$

is a basis for L over F. *Independence*: say that

$$0 = \sum_{i.j} f_{ij} l_i k_j = \sum_i \left(\sum_j f_{ij} k_j \right) l_i \qquad \text{for some } f_{ij} \in F.$$

As the l_i are independent over K then it follows that

$$\sum_j f_{ij} k_j = 0.$$

And as the k_j are independent over F then it follows that $f_{ij} = 0$ for all i, j. Consequently the $l_i k_j$ are linearly independent elements of L over F. *Spanning:* say that $l \in L$. Then there exists $\kappa_i \in K$ such that

$$l = \sum_i \kappa_i l_i.$$

Similarly there exist ϕ_{oj} such that

$$\kappa_i = \sum_j \phi_{ij} k_j$$

so that

$$l = \sum_i \sum_j \phi_{ij} k_j l_i$$

and we see that the $l_i k_j$ are spanning.

Conversely if $\dim_F L$ is finite then so is $\dim_F K$ as K is a subspace of L (over F) and $\dim_K L \leqslant \dim_F L$ as any set that spans L over F also spans L over K. ■

Example 114 *Find the degrees of the following extensions.*

$$\mathbb{Q}[\sqrt{2}, i] \colon \mathbb{Q} \qquad \textit{and} \qquad \mathbb{Q}[\sqrt{2}, \sqrt{3}].$$

Solution. The former is somewhat easier. The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$. It cannot be linear as $\sqrt{2}$ is irrational. So $\left|\mathbb{Q}[\sqrt{2}] \colon \mathbb{Q}\right| = 2$. Similarly the minimal polynomial of i over $\mathbb{Q}[\sqrt{2}]$ is $x^2 + 1$; it cannot be linear as $\mathbb{Q}[\sqrt{2}]$ is a real subfield and i is not real. So

$$\left|\mathbb{Q}[\sqrt{2}, i] \colon \mathbb{Q}\right| = \left|\mathbb{Q}[\sqrt{2}, i] \colon \mathbb{Q}[\sqrt{2}]\right| \left|\mathbb{Q}[\sqrt{2}] \colon \mathbb{Q}\right| = 2 \times 2 = 4.$$

Again $\left|\mathbb{Q}[\sqrt{2}] \colon \mathbb{Q}\right| = 2$. We need to take a little care to show that

$$\left|\mathbb{Q}[\sqrt{2}, \sqrt{3}] \colon \mathbb{Q}[\sqrt{2}]\right| = 2.$$

Certainly $x^2 - 3$ has $\sqrt{3}$ as a root, but is this polynomial irreducible over $\mathbb{Q}[\sqrt{2}]$? A general element of $\mathbb{Q}[\sqrt{2}]$ is $q_1 + q_2\sqrt{2}$ where $q_1, q_2 \in \mathbb{Q}$. If $x^2 - 3$ reduced in $\mathbb{Q}[\sqrt{2}]$ then we'd have

$$(q_1^2 - 2q_2^2) + 2q_1 q_2 \sqrt{2} = (q_1 + q_2\sqrt{2})^2 = 3$$

for some q_1, q_2. As 1 and $\sqrt{2}$ are independent over $\mathbb{Q}$ then we have $q_1 = 0$ or $q_2 = 0$ both of which lead to contradictions. Hence $x^2 - 3$ is the minimal polynomial of $\sqrt{3}$ over $\mathbb{Q}[\sqrt{2}]$ and we have

$$\left|\mathbb{Q}[\sqrt{2}, \sqrt{3}] \colon \mathbb{Q}\right| = \left|\mathbb{Q}[\sqrt{2}, \sqrt{3}] \colon \mathbb{Q}[\sqrt{2}]\right| \left|\mathbb{Q}[\sqrt{2}] \colon \mathbb{Q}\right| = 2 \times 2 = 4.$$

■

Example 115 *Show that* $\mathbb{F}_{16}$ *does not contain a subfield of order* 8.

Solution. If $\mathbb{F}_{16}$ has a subfield K of order 8 we would have $|K \colon \mathbb{Z}_2| = 3$ and

$$4 = |\mathbb{F}_{16} \colon \mathbb{Z}_2| = |\mathbb{F}_{16} \colon K| \, |K \colon \mathbb{Z}_2| = |\mathbb{F}_{16} \colon K| \times 3$$

which is a contradiction as 3 does not divide 4. ■

4. FACTORIZATION. EDS. PIDS. UFDS.

Throughout this section, rings will be assumed to be CRIs (commutative and having an identity) unless otherwise stated.

Definition 116 *Let R be a ring and let $a, b, c \in R$.*
*(i) If $a \neq 0$, we say that a **divides** b if there exists $r \in R$ such that $ra = b$. This is written*

$$a|b.$$

*Equivalently we say that a is a **factor** of b or that b is a **multiple** of a.*
*(ii) If $c \neq 0$, we say that c is a **common factor** of a and b if $c|a$ and $c|b$.*
*(iii) If c is a common factor of a and b, then we say that c is a **highest common factor** (or hcf) of a and b if whenever d is a common factor of a and b then $d|c$.*
*(iv) If $a, b \neq 0$ we say that c is a **common multiple** of a and b if $a|c$ and $b|c$.*
*(v) If c is a common multiple of a and b, then we say that c is a **least common multiple** (or lcm) of a and b if whenever d is a common multiple of a and b then $c|d$.*

Proposition 117 *In an integral domain, the hcf of two elements, if it exists, is unique up to multiplication by a unit. The corresponding result also holds for lcms.*

Proof. Let R be an integral domain and a, b be non-zero elements of R. Say that h_1 and h_2 are two hcfs of a and b. In particular h_1 and h_2 are common factors of a and b, so that $h_1|h_2$ as h_2 is an hcf and $h_2|h_1$ as h_1 is an hcf . So there exist $r_1, r_2 \in R$ such that

$$h_1 r_1 = h_2, \qquad h_2 r_2 = h_1.$$

Hence $h_1 (r_1 r_2 - 1_R) = 0$ and, as $h_1 \neq 0$, by the cancellation law $r_1 r_2 = 1_R$ and r_1 and r_2 are both units in R. ■

Example 118 *(i) Let $R = \mathbb{Z}$. Let $a = 105$ and $b = 441$. Then 21 and -21 are hcfs of a and b.*
(ii) Let $R = \mathbb{Q}[x]$. Let $a = x^2 - 3x + 4$ and $b = x^2 - 2x + 1$. Then the hcfs of a and b are of the form $c\,(x - 1)$ where c is a non-zero rational.
(iii) Let $R = \mathbb{Z}[x]$. Let $a = 2x - 2$ and $b = 4x^2 - 2x + 6$. Then the hcfs are 2 and -2.
(iv) Let $R = \mathbb{Z}[i] = \{x + yi : x, y \in \mathbb{Z}\}$. Let $a = 2$ and $b = -1 + 3i$. Then the hcfs are $1 + i$, $-1 - i$, $-1 + i$, $1 - i$.
(v) Let $R = \mathbb{Z}[\sqrt{-3}] = \left\{x + y\sqrt{-3} : x, y \in \mathbb{Z}\right\}$. Then $a = 4$ and $b = 2 + 2\sqrt{-3}$ have no highest common factor. To appreciate this:

$$\text{if} \quad 4 = (\alpha + \beta\sqrt{-3})(\gamma + \delta\sqrt{-3}) \quad \text{then} \quad 16 = (\alpha^2 + 3\beta^2)(\gamma^2 + 3\delta^2)$$

and, knowing the factorizations of 16 in $\mathbb{Z}$ we see that the factors of 4 are $\pm 1, \pm 1 \mp \sqrt{-3}, \pm 2, \pm 4$ and

$$\text{if} \quad 2 + 2\sqrt{-3} = (\alpha + \beta\sqrt{-3})(\gamma + \delta\sqrt{-3}) \quad \text{then} \quad 16 = (\alpha^2 + 3\beta^2)(\gamma^2 + 3\delta^2)$$

and so we see that the factors of $2+2\sqrt{-3}$ *are* $\pm 1, \pm 2, \pm 1 \mp \sqrt{-3}$. *So the common factors of* a *and* b *are*

$$\pm 1, \quad \pm 2, \quad \pm(1+\sqrt{-3}).$$

1 *cannot be a highest common factor as it divides the other two. However neither* 2 *nor* $1+\sqrt{-3}$ *is a highest common factor as neither divides the other.*

Definition 119 *In an integral domain, non-zero elements are said to be* ***coprime*** *if* 1_R *(or equivalently a unit) is an hcf of theirs.*

Proposition 120 *Let* R *be an integral domain and* $a, b \in R$. *If there exist* $u, v \in R$ *such that* $ua + vb = 1_R$ *then* a *and* b *are coprime.*

Proof. (a) Let c be a common factor of a and b. Then there exist $r, s \in R$ such that $a = cr$ and $b = cs$. Hence

$$1_R = ua + vb = (ur + vs)c$$

and we see that c is a unit. Finally, by definition, units are factors of all elements in a ring. ■

Definition 121 *Given a ring* R *we say that a non-zero, non-unit* $x \in R$ *is*

(a) a ***prime*** *element if whenever* $x|yz$ *then* $x|y$ *or* $x|z$.

(b) an ***irreducible*** *element if whenever* $x = yz$ *then* y *is a unit or* z *is a unit.*

Proposition 122 *(a) Given a non-zero, non-unit* x, *then the principal ideal* $\langle x \rangle$ *is prime if and only if* x *is a prime element.*

(b) In an integral domain, prime elements are irreducible.

(c) 2 *is irreducible but not prime in* $\mathbb{Z}[\sqrt{-3}]$.

Proof. (a) Say that x is prime and $ab \in \langle x \rangle$. Then $x|ab$ and as x is prime we have $x|a$ or $x|b$. That is $a \in \langle x \rangle$ or $b \in \langle x \rangle$. Conversely say that $\langle x \rangle$ is prime with x non-zero and a non-unit. If $x|ab$ then $ab \in \langle x \rangle$ and so $a \in \langle x \rangle$ or $b \in \langle x \rangle$ i.e. $x|a$ or $x|b$. As x is a non-unit then x is a prime element.

(b) Let R be an integral domain and x be prime. If $x = yz$ then $x|yz$ and so $x|y$ or $x|z$. If $x|y$ then $y = ux$ for some u and then $x = uxz$. By the cancellation rule $uz = 1$ and so z is a unit, showing that x is irreducible.

(c) Note that $2|(1+\sqrt{-3})(1-\sqrt{-3}) = 4$ but that 2 divides neither $1 \pm \sqrt{-3}$. So 2 is not prime. But if we could write

$$2 = (a + b\sqrt{-3})(c + d\sqrt{-3}).$$

Taking the modulus squared we see

$$4 = (a^2 + 3b^2)(c^2 + 3d^2).$$

If we have $a^2 + 3b^2 = 1$ then $a + b\sqrt{-3}$ is a unit. So a genuine reduction of 2 would mean

$$a^2 + 3b^2 = c^2 + 3d^2 = 2$$

and these equations have no solution. Hence 2 is irreducible. ■

4.1 Euclidean Domains (EDs)

Definition 123 *A **Euclidean domain** R is essentially an integral domain which admits the division algorithm. That is R is an integral domain together for which there exists a function, often called a **norm**, $d\colon R\backslash\{0\} \to \mathbb{N} = \{0, 1, 2, 3 \ldots\}$ such that*

(a) $d(a) \leqslant d(ab)$ for all $a, b \in R\backslash\{0\}$.

(b) given $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$ with $d(r) < d(b)$ or $r = 0$.

Example 124 *(a) The integers $\mathbb{Z}$ form an ED with $d(x) = |x|$ for $x \neq 0$.*

(b) Given a field F, the polynomial ring $F[x]$ forms an ED with $d(f) = \deg f$ for any $f \neq 0$.

Proposition 125 *The Gaussian integers $\mathbb{Z}[i]$ form an ED with*

$$d(a + bi) = |a + bi|^2 = a^2 + b^2 \qquad \text{where } a + bi \neq 0.$$

Proof. We immediately have for non-zero α, β in $\mathbb{Z}[i]$ that

$$d(\alpha\beta) = |\alpha\beta|^2 = |\alpha|^2 |\beta|^2 \geqslant |\alpha|^2 = d(\alpha).$$

Further we have that $\alpha/\beta \in \mathbb{C}$ and – noting the Gaussian integers form a grid of unit squares in $\mathbb{C}$ – there exists $q \in \mathbb{Z}[i]$ such that

$$\left|\frac{\alpha}{\beta} - q\right| \leqslant \frac{1}{\sqrt{2}}.$$

If we set $r = \alpha - q\beta \in \mathbb{Z}[i]$ then we have

$$d(r) = |\alpha - q\beta|^2 = \left|\frac{\alpha}{\beta} - q\right|^2 |\beta|^2 \leqslant \frac{1}{2} |\beta|^2 = \frac{1}{2} d(\beta) < d(\beta).$$

■

Example 126 *The ring $\mathbb{Z}[\sqrt{2}]$ is an ED with*

$$d(a + b\sqrt{2}) = \left|a^2 - 2b^2\right|.$$

Show that the equations $x^2 - 2y^2 = 1$ and $x^2 - 2y^2 = -1$ each have infinitely many integer solutions.

Solution. Firstly note that d is multiplication. If $\alpha = a_1 + a_2\sqrt{2}$ and $\beta = b_1 + b_2\sqrt{2}$ then

$$\begin{aligned}
d((a_1 + a_2\sqrt{2})(b_1 + b_2\sqrt{2})) &= d((a_1b_1 + 2a_2b_2) + (a_2b_1 + a_1b_2)\sqrt{2}) \\
&= \left|(a_1b_1 + 2a_2b_2)^2 - 2(a_2b_1 + a_1b_2)^2\right| \\
&= \left|a_1^2b_1^2 + 4a_1a_2b_1b_2 + 4a_2^2{b_2}^2 - 2a_2^2b_1^2 - 2a_1^2b_2^2 - 4a_1a_2b_1b_2\right| \\
&= \left|a_1^2 - 2a_2^2\right| \left|b_1^2 - 2b_2^2\right| \\
&= d(a_1 + a_2\sqrt{2}) d(b_1 + b_2\sqrt{2}).
\end{aligned}$$

Let $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]\backslash\{0\}$. So we have $d(\alpha) \leqslant d(\alpha)d(\beta) = d(\alpha\beta)$ as $d(\beta) \geqslant 1$. Also we have $\alpha/\beta = q_1 + q\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$ and so we can find $\gamma = c_1 + c_2\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$ such that $|q_i - c_1| \leqslant 1/2$ so that

$$\begin{aligned} d\left(\frac{\alpha}{\beta} - \gamma\right) &= \left|(q_1 - c_1)^2 - 2(q_2 - c_2)^2\right| \\ &\leqslant (q_1 - c_1)^2 + 2(q_2 - c_2)^2 \\ &\leqslant \frac{1}{4} + \frac{2}{4} = \frac{3}{4}. \end{aligned}$$

If we set $\delta = \alpha - \beta\gamma$ and so

$$d(\delta) = d\left(\alpha - \beta\gamma\right) = d\left(\frac{\alpha}{\beta} - \gamma\right) d(\beta) \leqslant \frac{3}{4} d(\beta) < d(\beta).$$

If instead we had defined $D(a + b\sqrt{2}) = a^2 - 2b^2$ then we see as above that D is multiplicative. As $D(1 + \sqrt{2}) = 1^2 - 2 \times 1^2 = -1$ it then follows that

$$D(x_n + \sqrt{2}y_n) = D((1 + \sqrt{2})^n) = (-1)^n$$

and so (x_{2n}, y_{2n}) is a solution of $x^2 - 2y^2 = 1$ and (x_{2n+1}, y_{2n+1}) is a solution of $x^2 - 2y^2 = -1$. ■

Proposition 127 *Let R be an ED and $I \lhd R$. Then I is principal.*

Proof. If $I = \{0_R\}$ then we are done as $I = \langle 0_R \rangle$. Otherwise there exists $x \in R$ with $x \neq 0$ and $d(x)$ minimal. Certainly $\langle x \rangle \subseteq I$. Conversely if $y \in I$ then there exist $q, r \in R$ such that $y = qx + r$ where $d(r) < d(x)$ or $r = 0$. As $r = y - qx \in I$ then $r = 0$ by the minimality of $d(x)$. So $y = qx$ and hence $I \subseteq \langle x \rangle$. Hence I is principal. ■

Corollary 128 *In an ED, irreducible elements are prime.*

Proof. Say that x is irreducible and $\langle x \rangle \subseteq I \lhd R$. There exists y such that $\langle y \rangle = I$ and so there exists z such that $x = yz$. By the irreducibility of x we either have y is a unit and $I = \langle y \rangle = R$ or z is a unit and we have $I = \langle y \rangle = \langle x \rangle$. Hence $\langle x \rangle$ is maximal, and so prime, and hence x is a prime element. ■

Example 129 *$\mathbb{Z}[x]$ and $\mathbb{Q}[x, y]$ are not Euclidean domains*

Solution. In the first case note that $\langle 2, x \rangle$ is not principal, and in the second case $\langle x, y \rangle$ is not principal. ■

4.2 Aside – the Euclidean Algorithm (Off-syllabus)

The Euclidean algorithm employs the division algorithm repeatedly to find the hcf of two integers a and b. The algorithm first appeared in Euclid's *Elements* Book VII in around 300 B.C.. First we describe an example in $\mathbb{Z}$ to see how the algorithm works.

Example 130 *Find the highest common factor of* 53714 *and* 30281.

Solution. Let $n_1 = 53714$ (the greater of the two numbers) and $n_2 = 30281$.
$n_1/n_2 = 1.77\ldots$ and so $q_1 = 1$ and we set $n_3 = r_1 = n_1 - 1 \times n_2 = 23433$;
$n_2/n_3 = 1.29\ldots$ and so $q_2 = 1$ and $n_4 = r_2 = n_2 - 1 \times n_3 = 6848$;
$n_3/n_4 = 3.42\ldots$ and so $q_3 = 3$ and $n_5 = r_3 = n_3 - 3 \times n_4 = 2889$;
$n_4/n_5 = 2.37\ldots$ and so $q_4 = 2$ and $n_6 = r_4 = n_4 - 2 \times n_5 = 1070$;
$n_5/n_6 = 2.7$ and so $q_5 = 2$ and $n_7 = r_5 = n_5 - 2 \times n_6 = 749$;
$n_6/n_7 = 1.42\ldots$ and so $q_6 = 1$ and $n_8 = r_6 = n_6 - 1 \times n_7 = 321$;
$n_7/n_8 = 2.33\ldots$ and so $q_7 = 2$ and $n_9 = r_7 = n_7 - 2 \times n_8 = 107$;
$n_8/n_9 = 3$ and so $q_8 = 3$ and $n_{10} = n_9 - 3 \times n_8 = 0$.
At this point the algorithm terminates and the output is the last positive number namely 107.
■

Algorithm 131 *(**Euclidean algorithm**) Let a_1 and a_2 be non-zero elements of an ED with $d(a_1) \geqslant d(a_2)$. Then the Euclidean algorithm uniquely defines two sequences q_i and a_i of integers by*

$$\begin{aligned} a_1 &= q_1 a_2 + a_3 \qquad \text{where } d(a_3) < d(a_2) \\ a_2 &= q_2 a_3 + a_4 \qquad \text{where } d(a_4) < d(a_3) \\ a_3 &= q_3 a_4 + a_5 \qquad \text{where } d(a_5) < d(a_4) \end{aligned}$$

and so on. The algorithm terminates if $a_k = 0$ for some k.

Theorem 132 *The Euclidean algorithm always terminates with $a_k = 0$ for some k and*

$$a_{k-1} = \text{hcf}\,(a_1, a_2)\,.$$

Proof. The sequence of positive integers $d(a_i)$ is strictly decreasing and bounded below, and so can only be finite in length. If $d(a_i) > 0$ then it possible to run (at least) one further application of the division algorithm and so the Euclidean algorithm terminates when $a_k = 0$ for some k.

Firstly, I claim a_{k-1} divides a_i for $1 \leqslant i \leqslant k$. The proof follows by reverse induction. Certainly $a_{k-1}|a_{k-1}$ and $a_{k-1}|a_k = 0$. Suppose (as an inductive hypothesis) that a_{k-1} divides a_r and a_{r+1}. Then

$$a_{r-1} = q_{r-1} a_r + a_{r+1}$$

is also divisible by a_{k-1}. Hence by induction a_{k-1} is a common factor of a_1 and a_2.

Secondly, we need to show that $a_{k-1} = \text{hcf}(a_1, a_2)$. Suppose that m is a common factor of a_1 and a_2. I claim that m is also a factor of a_{k-1} with the proof following by induction. Suppose (as an inductive hypothesis) that m divides a_r and a_{r+1}. Then m divides

$$a_{r+2} = a_r - q_r a_{r+1}$$

also. By induction m divides a_{k-1} also completing the proof. ■

Theorem 133 ***(Bézout's Lemma)*** *Let a, b be non-zero elements of an ED R with a highest common factor h. Then there exist u and v in R such that*

$$ua + vb = h.$$

Proof. Set $a_1 = a$ and $a_2 = b$. The proof uses *reverse induction* working backwards through the calculations performed in the Euclidean Algorithm. We will show that, for each $i = 1, 2, \ldots, k$, there exist u_i and v_i such that

$$u_i a_i + v_i a_{i+1} = h. \tag{4.1}$$

As $h = a_{k-1}$ then we can set $u_{k-2} = 0$ and $v_{k-2} = 1$ to see that (4.1) is true for $i = k-2$. Now suppose, as an inductive hypothesis, that (4.1) holds true for $i = I$, where $1 < I \leqslant k-2$, and we shall aim to show it's true for $i = I - 1$. We have

$$\begin{aligned} h &= u_I a_I + v_I a_{I+1} \\ &= u_I a_I + v_I (a_{I-1} - q_{I-1} a_I) \\ &= v_I a_{I-1} + (u_I - q_{I-1} v_I) a_I \end{aligned}$$

thus proving that (4.1) holds true for $i = I - 1$ with $u_{I-1} = v_I$ and $v_{I-1} = u_I - q_{I-1} v_I$. By induction (4.1) holds true when $i = 1$ which is the required result for the case of a and b. ■

Example 134 *Find integers u and v such that* $53714u + 30281v = 107$.

Solution. Recall from our earlier calculations that

$$\begin{aligned} n_3 &= n_1 - n_2; \qquad n_4 = n_2 - n_3; \qquad n_5 = n_3 - 3n_4; \\ n_6 &= n_4 - 2n_5; \qquad n_7 = n_5 - 2n_6; \qquad n_8 = n_6 - n_7; \end{aligned}$$

with $n_9 = n_7 - 2n_8 = 107$ being the highest common factor. We will "reverse" the above equations to write n_9 in terms of n_7 and n_8, then in terms of n_6 and n_7, and so on repeatedly until we have n_8 in terms of n_1 and n_2. We see

$$\begin{aligned} 107 &= n_7 - 2n_8 \\ &= n_7 - 2(n_6 - n_7) = -2n_6 + 3n_7 \\ &= -2n_6 + 3(n_5 - 2n_6) = 3n_5 - 8n_6 \\ &= 3n_5 - 8(n_4 - 2n_5) = -8n_4 + 19n_5 \\ &= -8n_4 + 19(n_3 - 3n_4) = 19n_3 - 65n_4 \\ &= 19n_3 - 65(n_2 - n_3) = -65n_2 + 84n_3 \\ &= -65n_2 + 84(n_1 - n_2) = 84n_1 - 149n_2 \end{aligned}$$

■

Example 135 *Find the multiplicative inverse of* 2167 *in* mod 65537 *arithmetic.*

Solution. We first find integers u and v such that $2167u + 65537v = 1$. Applying first the Euclidean algorithm we set:

Let $n_1 = 65537$ and $n_2 = 2167$.
As $n_1/n_2 = 30.24\ldots$ and so $q_1 = 30$ and we set $n_3 = r_1 = n_1 - 30 \times n_2 = 527$.
As $n_2/n_3 = 4.11\ldots$ then $q_2 = 4$ and $n_4 = r_2 = n_2 - 4 \times n_3 = 59$.
As $n_3/n_4 = 8.93\ldots$ then $q_3 = 8$ and $n_5 = r_3 = n_3 - 8 \times n_4 = 55$.
As $n_4/n_5 = 1.07\ldots$ then $q_4 = 1$ and $n_6 = r_4 = n_4 - n_5 = 4$.
As $n_5/n_6 = 13.75$ then $q_5 = 13$ and $n_7 = r_5 = n_5 - 13 \times n_6 = 3$.
As $n_6/n_7 = 1.25$ then $q_6 = 1$ and $n_8 = r_6 = n_6 - n_7 = 1$.

Hence $\text{hcf}\,(65537, 2167) = 1$ as required and we can work backwards to find

$$\begin{aligned}
1 &= n_6 - n_7 \\
&= n_6 - (n_5 - 13n_6) = -n_5 + 14n_6 \\
&= -n_5 + 14\,(n_4 - n_5) = 14n_4 - 15n_5 \\
&= 14n_4 - 15\,(n_3 - 8n_4) = -15n_3 + 134n_4 \\
&= -15n_3 + 134\,(n_2 - 4n_3) = 134n_2 - 551n_3 \\
&= 134n_2 - 551\,(n_1 - 30n_2) = -551n_1 + 16664n_2.
\end{aligned}$$

Hence we have that

$$(-551) \times (65537) + (16664) \times (2167) = 1.$$

So in mod 65537 arithmetic we have

$$16664 \times 2167 = 1 \bmod 65537 \quad \text{or equally} \quad 2167^{-1} = 16664 \bmod 65537.$$

■

4.3 Unique Factorization in PIDs.

Definition 136 *An integral domain R is said to be a* ***principal ideal domain*** *(or PID) if every ideal is principal.*

Example 137 *We know that every Euclidean domain is a PID. These include (with F a field)*

$$\mathbb{Z}, \qquad \mathbb{Z}[i], \qquad F[x], \qquad \mathbb{Z}[\sqrt{2}]$$

but not $\mathbb{Z}[x]$ nor $F[x, y]$.

Definition 138 *An integral domain R is said to be a* ***unique factorization domain*** *(or UFD) if every non-zero, non-unit element x can be written*

$$x = p_1 p_2 \cdots p_r$$

where $p_1, \ldots, p_r$ are irreducible elements and further this factorization is unique in the sense that if

$$x = p_1 p_2 \cdots p_r = q_1 q_2 \ldots q_s$$

are two factorizations into irreducible elements then $r = s$ and (with a possible reordering of the factors) we have

$$p_i = u_i q_i$$

for each i where each u_i is a unit.

Remark 139 *Though we shall not prove it explicitly, the above definition is equivalent to requiring that (i) non-zero elements can be written as the product of irreducible elements and (ii) irreducible elements are prime.*

Example 140 *(i)* $4 = (1+\sqrt{-3})(1-\sqrt{-3}) = 2 \times 2$ *are two essentially different factorizations in* $\mathbb{Z}[\sqrt{-3}]$.

(ii) $5 = (1+2i)(1-2i) = (2-i)(2+i)$ *are essentially the same factorizations in* $\mathbb{Z}[i]$.

(iii) $2x^2 - 6x + 4 = (2x-2)(x-2) = (x-1)(2x-4)$ *are essentially the same factorization in* $\mathbb{Q}[x]$.

(iv) $x^2 - 1 = (x-1)(x+1) = (x-3)(x-5)$ *are two essentially different factorizations in* $\mathbb{Z}_8[x]$.

Example 141 *Factorize* $17 - i$ *into irreducibles in* $\mathbb{Z}[i]$.

Solution. Note that $d(17-i) = 17^2 + 1^2 = 290 = 2 \times 5 \times 29$ so that any factorization into irreducibles can involve at most three factors. Up to multiplication by units, the only elements with a norm of 2 are $1 \pm i$. Now note that

$$\frac{17-i}{1+i} = \frac{(17-i)(1-i)}{2} = \frac{16-18i}{2} = 8 - 9i.$$

Likewise the only Gaussian integers with a norm of 5 are $1 \pm 2i$ up to units. Note that

$$\frac{8-9i}{1+2i} = \frac{(8-9i)(1-2i)}{5} = \frac{-10-25i}{5} = -2 - 5i.$$

Hence

$$17 - i = (1+i)(1+2i)(-2-5i)$$

is a factorization into irreducibles. ■

Proposition 142 *In a UFD two non-zero elements have a unique hcf.*

Proof. Let x, y be non-zero elements of a UFD R. By considering the irreducible elements which divide a or b we may write

$$x = u p_1^{\alpha_1} \cdots p_r^{\alpha_r} \qquad \text{and} \qquad y = v p_1^{\beta_1} \cdots p_r^{\beta_r}$$

where u and v are units and $\alpha_i, \beta_i \geqslant 0$. Then it can be shown that

$$\text{hcf}(x, y) = p_1^{\gamma_1} \cdots p_r^{\gamma_r} \qquad \text{where } \gamma_i = \min\{\alpha_i, \beta_i\}$$

and that this hcf is unique up to multiplication by a unit. Details are omitted here. ■

We know that in an ID a prime element is irreducible, but that the converse need not hold (e.g. 2 in $\mathbb{Z}[\sqrt{-3}]$). We see now that in UFDs and PIDs that irreducibles are prime. As we shall see, PIDs are UFDs and so the latter should not be surprising. But we shall need the equivalence of primes and irreducibles in PIDs to actually show that PIDs are UFDs.

Proposition 143 *(a) In a UFD an irreducible element is prime.*
(b) In a PID an irreducible element is prime.

Proof. (a) Say that x is irreducible and that $x|yz$. We then have that $xv = yz$ for some v. The elements v, y, z can each be factorized into irreducibles, and by the uniqueness of the factorizations of $xv = yz$ it must be that x (up to a unit) is present in the factorizations of y or z (or both). So $x|y$ or $x|z$ and we see that x is prime.

(b) Say that x is irreducible and $\langle x\rangle \subseteq I \lhd R$. There exists y such that $\langle y\rangle = I$ and so there exists z such that $x = yz$. By the irreducibility of x we either have y is a unit and $I = \langle y\rangle = R$ or z is a unit and we have $I = \langle y\rangle = \langle x\rangle$. Hence $\langle x\rangle$ is maximal, and so prime, and hence x is a prime element. ■

Proposition 144 *PIDs are Noetherian. That is given an increasing sequence of ideals*

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

in a PID R, there exists N such that $I_n = I_N$ for $n \geqslant N$.

Proof. This is left to Sheet 2, Exercise 5(i). ■

Theorem 145 *A PID is a UFD.*

Proof. *Existence*: Let R be a PID and x be non-zero. If x is irreducible then we are done; otherwise we may write $x = yz$ where y and z are not units. If y and z are irreducible then we are done, and otherwise we may continue factorizing the composite elements that arise and, say, write next $y = ab$. If the process terminates (through lack of any genuinely composite remaining factors) then we have written x as a product of irreducibles and we are done. However if the process does not terminate then we could produce an infinite strictly increasing sequence of ideals

$$\langle x\rangle \subseteq \langle y\rangle \subseteq \langle a\rangle \subseteq \cdots .$$

However as PIDs are Noetherian then this cannot occur and the above factorization will terminate.

Uniqueness: Say that we have

$$x = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

are two factorizations of x into irreducible elements. As p_1 is irreducible, and so prime, then $p_1|q_1 q_2 \cdots q_s$ implies that $p_1|q_i$ for some i. As q_i is irreducible and p_1 is not a unit we then have $q_i = up_1$ for some unit u. By renumbering $q_1, \ldots, q_s$ and incorporating the unit u into one of the other factors we may assume that $p_1 = q_1$ and then by the cancellation law we have that

$$p_2 \cdots p_r = q_2 \cdots q_s$$

and may proceed along similar lines again, ultimately showing that $r = s$ and that the two factorizations are essentially the same. ■

4.4 Factorization in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$

Factorization in $\mathbb{C}[x]$ is, in principle at least, very straightforward. The fundamental theorem of algebra tells us that any complex polynomial can be uniquely written as

$$qc(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)$$

for some $\alpha_1,\ldots,\alpha_n,c\in\mathbb{C}$. These linear factors are irreducible and so no further factorization is possible.

The above also helps us appreciate how factorization works in $\mathbb{R}[x]$ when one recalls that a real polynomial's non-real roots will arise as conjugate pairs. Hence we can factorize any real polynomial can be unqiuely factorized as

$$c(x-a_1)(x-a_2)\cdots(x-a_r)q_1(x)q_2(x)\cdots q_s(x)$$

where $a_1,\ldots,a_s,c$ are real and $q_1,\cdots,q_s$ are irreducible monic quadratics.

We know from our earlier discussion that $\mathbb{Q}[x]$ is a UFD but we are yet to prove this for $\mathbb{Z}[x]$ (though this is indeed the case). But we have no general theorem like that above to help us determine these rings' irreducible elements. We begin with a naive treatment of the following example.

Example 146 *Show that the cubic x^3-2x+3 is irreducible in $\mathbb{Z}[x]$.*

Solution. If the cubic did factorize then its factors would include a linear factor $ax+b$. Note that a would have to divide 1 and that b would have to divide 3. So up to units, the only linear factors could be $x-1,x+1,x-3,x+3$. However as none of $1,-1,3,-3$ is a root of the cubic then we can see that the cubic is irreducible over $\mathbb{Z}$. ■

How might we have approached this problem in $\mathbb{Q}[x]$? Certainly not so naively treating a finite number of possibilities. Reassuringly Gauss showed these problems to be largely equivalent. We first note the following useful and quite general approach involving the polynomial rings $\mathbb{Z}_p[x]$.

Proposition 147 *Let $f(x)$ be an integer polynomial whose leading coefficient is not divisible by the prime p. If $\overline{f(x)}$ is irreducible in $\mathbb{Z}_p[x]$ then $f(x)$ is irreducible in $\mathbb{Z}[x]$.*

Proof. Suppose that $f=gh$ be a proper factorization in $\mathbb{Z}[x]$. Then g and h are both of positive degree and we also have $\deg f=\deg f+\deg g$. In $\mathbb{Z}_p[x]$ we have $\overline{f}=\overline{g}\overline{h}$ and

$$\deg f=\deg\overline{f}=\deg\overline{g}+\deg\overline{h}\leqslant\deg g+\deg h=\deg f$$

as the leading coefficient of f is not divisible by p, as $\mathbb{Z}_p$ is a field and as $\deg\overline{p}\leqslant\deg p$ in general. Hence it must be the case that $\deg\overline{g}$ and $\deg\overline{h}$ are both positive and $\overline{f}$ is reducible in $\mathbb{Z}_p[x]$. ■

Example 148 *Show that the following polynomials are irreducible over $\mathbb{Z}[x]$.*

$$f_1(x)=17x^3+7x+3,\qquad f_2(x)=2x^3+3x^2+x-2,\qquad f_3(x)=7x^4+5x-3.$$

Solution. $f_1(x) \bmod 2$ equals $x^3 + x + 1$ which is irreducible over $\mathbb{Z}_2$ as it has no roots in $\mathbb{Z}_2$. Hence $f_1(x)$ is irreducible over $\mathbb{Z}$.

We cannot use $\mathbb{Z}_2$ for $f_2(x)$ as 2 divides the leading coefficient. However $\bmod 3$ we obtain the polynomial

$$2x^3 + x + 1$$

which does not have a root in $\mathbb{Z}_3$ and so $f_2(x)$ is irreducible over $\mathbb{Z}$.

$f_3(x) \bmod 2$ equals $x^4 + x + 1$ which is irreducible over $\mathbb{Z}_2$ as $\overline{f_3(x)}$ has no roots in $\mathbb{Z}_2$ and as $x^2 + x + 1$, the only irreducible quadratic $\bmod 2$ is not a factor. So $f_3(x)$ is irreducible over $\mathbb{Z}$. ■

Whilst thinking along these lines we will also introduce the following criterion for irreducibility. The criterion may seem somewhat contrived but can in fact be very useful particularly for cyclotomic polynomials.

Proposition 149 ***(Eisenstein's Criterion)*** *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be an integer polynomial and p be a prime such that*

$$p \text{ does not divide } a_n; \qquad p \text{ divides } a_0, a_1, \ldots, a_{n-1}; \qquad p^2 \text{ does not divide } a_0.$$

Then $f(x)$ is irreducible in $\mathbb{Z}[x]$.

Proof. Note that $\bmod p$ we have

$$\overline{f(x)} = \overline{a_n} x^n$$

where $\overline{a_n} \neq \bar{0}$. If $f(x) = g(x)h(x)$ is a genuine reduction into integer polynomials, then, as $\mathbb{Z}_p[x]$ is a UFD, we have

$$\overline{g(x)} = \bar{b}x^r \qquad\qquad \overline{h(x)} = \bar{c}x^s$$

where $r + s = n$ and $r, s \geqslant 1$. So p divides the constant coefficients of $g(x)$ and $h(x)$ and hence p^2 divides the constant constant of $f(x)$, but this is the required contradiction. ■

Example 150 *Show that $x^4 + x^3 + x^2 + x + 1$ is irreducible over $\mathbb{Z}$.*

Solution. A standard trick for such "cyclotomic" polynomials (irreducible factors of $x^n - 1$) is to set $x = u + 1$. The new polynomial in u will be irreducible if and only if the original one in x is. Now

$$(u+1)^4 + (u+1)^3 + (u+1)^2 + (u+1) + 1 = u^4 + 5u^3 + 10u^2 + 10u + 5.$$

We see that the new polynomial is irreducible – by Eisenstein with $p = 5$ – and so the original polynomial also is. ■

We return now to discussion relating factorization over $\mathbb{Q}$ to factorization over $\mathbb{Z}$.

Definition 151 *Let $f(x)$ be a polynomial over the integers. The* ***content*** *$c(f)$ is the hcf of the coefficients of f. A polynomial is said to be* ***primitive*** *is it has content equal to* 1.

Note that any integer polynomial can be written uniquely as $f(x) = cp(x)$ where $c = c(f)$ and $p(x)$ is a primitive polynomial.

Lemma 152 *(**Gauss Lemma**) The product of two primitive polynomials is primitive. Consequently*

$$c(fg) = c(f)c(g)$$

for any two integer polynomials f and g.

Proof. Say that f and g are primitive polynomials and suppose for a contradiction that fg is not primitive. Let p be a prime factor of $c(fg)$. We then have that

$$\overline{f(x)g(x)} = 0 \bmod p.$$

As $\mathbb{Z}_p[x]$ is an ID then either $\overline{f(x)} = 0$ or $\overline{g(x)} = 0$ in $\mathbb{Z}_p[x]$. But then p divides the coefficients of f or divides the coefficients of g. This would contradict f and g being primitive.

More generally for integer polynomials f and g we have

$$f = c(f)p_f, \qquad g = c(g)p_g, \qquad fg = c(fg)p_{fg}.$$

Then

$$c(fg)p_{fg} = c(f)c(g)p_f p_g.$$

As $p_f p_g$ is primitive, then by the uniqueness of the above expressions we have $c(fg) = c(f)c(g)$ and $p_{fg} = p_f p_g$. ■

Gauss' lemma plays a key role in relating factorization in $\mathbb{Z}[x]$ with factorization in $\mathbb{Q}[x]$. Note

- $2x - 6$ is irreducible in $\mathbb{Q}[x]$ but not in $\mathbb{Z}[x]$ as $2x - 6 = 2(x - 3)$ and 2 is not a unit in $\mathbb{Z}[x]$.
- But if $f(x)$ in $\mathbb{Z}[x]$ reduces into positive degree polynomials in $\mathbb{Z}[x]$ then it is reducible in $\mathbb{Q}[x]$ as $\mathbb{Z}[x] \subseteq \mathbb{Q}[x]$.
- In general polynomials in $\mathbb{Q}[x]$ are not in $\mathbb{Z}[x]$ but can be multiplied by a non-zero integer to form a polynomial in $\mathbb{Z}[x]$.

Given the example that started this section we would much rather work with irreducibility in $\mathbb{Z}[x]$ rather than in $\mathbb{Q}[x]$ if at all possible. The following result addresses precisely this point.

Theorem 153 *Let f be a primitive non-constant polynomial in $\mathbb{Z}[x]$. Then $f(x)$ is irreducible in $\mathbb{Z}[x]$ if and only if $f(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. If f is irreducible over $\mathbb{Q}$ then it is irreducible over $\mathbb{Z}$, given the comments above. Say now that $f(x)$ is genuinely reducible in $\mathbb{Q}[x]$, say $f(x) = q_1(x)q_2(x)$ for polynomials in $\mathbb{Q}[x]$. We then have $g_1 = d_1 q_1$ and $g_2 = d_2 q_2$ are integer polynomials where d_i is the lcm of the denominators of the coefficients in q_i. Finally we have $g_i = c(g_i)p_i$ where p_i are primitive integer polynomials. So

$$d_1 d_2 f = g_1 g_2 = c(g_1)c(g_2)p_1 p_2.$$

As f and $p_1 p_2$ are both primitive then we have, by uniqueness, that $f = p_1 p_2$ and so f is reducible in $\mathbb{Z}[x]$. ■

Proposition 154 *$\mathbb{Z}[x]$ is a UFD.*

Proof. Any $f(x)$ in $\mathbb{Z}[x]$ can be written uniquely as $f(x) = c(f)p(x)$ where p is primitive. As $\mathbb{Z}$ is a UFD then $c(f)$ can be uniquely written as a product of prime (integers). As $\mathbb{Q}[x]$ is a UFD then $p(x)$ can be written as a product of irreducible rational polynomials. As seen in the proof of Theorem 153 any reduction of $p(x)$ into rational polynomials can be replaced with a reduction of $p(x)$ into integer polynomials, again in a unique fashion (up to the units of $\mathbb{Z}$). ■

In fact the above theorems generalize quite naturally to demonstrate the following result. As UFDs have well defined hcfs then we can introduce the notion of content to polynomials in $R[x]$. Replacing $\mathbb{Z}$ with R and $\mathbb{Q}$ with the field of fractions of R then we can rewrite the above arguments to show:

Proposition 155 *If R is a UFD then so is $R[x]$.*

5. MODULES

Modules are essentially the equivalent of vector spaces when the scalars come from a ring rather than a field. Throughout we will consider only modules over Euclidean domains. (The results that follow can all be proved over PIDs, though some of the proofs can be somewhat more laborious and in any case the important examples we will meet are all over EDs.)

Definition 156 *Let R be an ED. An R-**module** M is an abelian group together with scalar multiplication $R \times M \to M$, denoted as $(r, m) \mapsto rm$, such that*

- $r(m_1 + m_2) = rm_1 + rm_2$ *for all* $r \in R$ *and* $m_1, m_2 \in M$.
- $(r_1 + r_2)m = r_1m + r_2m$ *for all* $r_1, r_2 \in R$ *and* $m \in M$.
- $(r_1r_2)m = r_1(r_2m)$ *for all* $r_1, r_2 \in R$ *and* $m \in M$.
- $1_Rm = m$ *for all* $m \in M$.

Example 157 ***Modules over*** $\mathbb{Z}$. *A $\mathbb{Z}$-module is an abelian group and vice versa an abelian group is a $\mathbb{Z}$-module. Scalar multiplication by integers is entirely determined by the abelian group structure as for any positive integer n we have*

$$n.m = (1 + \cdots + 1).m = (1.m) + \cdots + (1.m) = m + \cdots + m$$

and $(-n).m = (-1).(n.m) = -(n.m)$.

Example 158 ***Modules over fields***. *A module over a field is a vector space.*

Example 159 ***Modules over polynomial rings***. *Say that M is a module over $F[x]$ where F is a field. Then M is a vector space over F when considered as the constant polynomials in $F[x]$.*

Multiplication by x, that is $T(m) = xm$ has the effect of an F-linear map $T\colon M \to M$.

Note that the entire $F[x]$-module structure on V is entirely determined by this map T as by the module axioms we have

$$p(x).m = p(T)m$$

for any polynomial $p(x)$ in $F[x]$.

Conversely given a vector space M over F and a linear map $T\colon M \to M$ we can define M as a $F[x]$-module by defining scalar multiplication as

$$x.m = T(m).$$

Example 160 *Given a field F and a square matrix A over F then $F[A]$ is an $F[x]$-module with*

$$x.p(A) = (xp)(A).$$

N.B. *In general though $F[A]$ is a different module to the $F[x]$-module defined by A.*

For example, consider the real matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

The $\mathbb{R}[x]$-module defined by A is a 3-dimensional real vector space. By comparison

$$\mathbb{R}[A] \cong \frac{\mathbb{R}[x]}{\langle (x-1)^2 \rangle},$$

because $m_A(x) = (x-1)^2$, and is a two-dimensional real vector space. $\mathbb{R}[A]$ is spanned by I, A.

Example 161 *Given two R-modules M and N then we can form the **direct sum** $M \oplus N$ as an R-module in the natural way by component-wise addition and scalar multiplication.*

*Generally for a positive integer n we can define the R-module $R^n = R \oplus \cdots \oplus R$ (n times). These are the **free modules** over R.*

Example 162 *If R is an ED and I is an ideal of R then R/I is naturally an R-module. For example with $R = \mathbb{Q}[x]$ and $I = \langle x^2 + 1 \rangle$ then*

$$\frac{\mathbb{Q}[x]}{\langle x^2 + 1 \rangle}$$

also has the structure of a two-dimensional vector space over $\mathbb{Q}$ with basis $1, x$. Scalar multiplication is defined by

$$x.1 = x, \qquad x.x = x^2 = -1$$

so that multiplication by x is represented by the matrix

$$B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

with respect to the above basis. Hopefully unsurprisingly

$$\frac{\mathbb{Q}[x]}{\langle x^2 + 1 \rangle} \qquad \text{and} \qquad \mathbb{Q}[B]$$

are isomorphic as $\mathbb{Q}[x]$-modules, with such an isomorphism being $a + bx \mapsto a + bB$.

Here we make rigorous the idea of being isomorphic as R-modules and also introduce the idea of a module homomorphism.

Definition 163 *Let M and N be R-modules. A map $\phi\colon M \to N$ is a* ***module homomorphism*** *if*

$$\phi(m_1 + m_2) = \phi(m_1) + \phi(m_2), \qquad \phi(rm_1) = r\phi(m_1)$$

where $r \in R, m_1, m_2 \in M$. We say that ϕ is a ***module isomorphism*** *if ϕ is a bijection.*

Example 164 *A module homomorphism between $\mathbb{Z}$-modules is a group homomorphism.*

Example 165 *If R is a field then the module homomorphisms are precisely the linear maps.*

Example 166 *Let*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Find all module homomorphisms from $\mathbb{R}[C]$ to $\mathbb{R}[D]$.

Solution. Every element of $\mathbb{R}[C]$ can be written as $\alpha I + \beta C$ and every of $\mathbb{R}[D]$ can be written $\alpha I + \beta D$. So if we take initial basis $\{I, C\}$ and final basis $\{I, D\}$ then a module homomorphism ϕ can must be represented by a 2×2 real matrix. Say

$$\phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

so that $\phi(I) = aI + cD$ and $\phi(C) = bI + dD$. Our further requirements on ϕ are that

$$\begin{aligned} bI + dD &= \phi(C) = \phi(x.I) = x.\phi(I) = D(aI + cD) = -3cI + (a + 4c)D; \\ (-2a + 3b)I + (-2c + 3d)D &= \phi(3C - 2I) = \phi(C^2) \\ &= \phi(x.C) = x.\phi(C) = D(bI + dD) = -3dI + (b + 4d)D, \end{aligned}$$

noting that $C^2 = 3C - 2I$ and $D^2 = 4D - 3I$. Comparing coefficients we have that

$$b = -3c, \qquad d = a + 4c, \qquad -2a + 3b = -3d, \qquad -2c + 3d = b + 4d.$$

Solving these we see there is a one-parameter family of solutions with

$$a = b = -3c \qquad \text{and} \qquad d = c.$$

So

$$\phi = c \begin{pmatrix} -3 & -3 \\ 1 & 1 \end{pmatrix}.$$

We might have done the above calculation in a slightly more systematic way. Under the identification $(\alpha, \beta)^T \leftrightarrow \alpha I + \beta C$ of $\mathbb{R}^2$ with $\mathbb{R}[C]$ we note

$$x. \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \leftrightarrow x.(\alpha I + \beta C) = \alpha C + \beta C^2 = -2\beta I + (\alpha + 3\beta)C \leftrightarrow \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

so that multiplication by x in $\mathbb{R}[C]$ is given by the matrix

$$\begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$$

with respect to our basis, and similarly multiplication by x in $\mathbb{R}[D]$ is given by the matrix

$$\begin{pmatrix} 0 & -3 \\ 1 & 4 \end{pmatrix}$$

with respect to $\{I, D\}$. So the matrices for ϕ that we found are precisely those that satisfy

$$\begin{pmatrix} 0 & -3 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$$

as a consequence of requiring $x.\phi(v) = \phi(x.v)$ for all v in $\mathbb{R}[C]$. (We will see in due course that these two matrices are the companion matrices for $m_C(x)$ and $m_D(x)$.)

Why are there limited module homomorphisms here? And how might we better understand them? The second question will have a clearer answer once we better understand the structure of modules. However, to answer the first question, note that scalar multiplication by x^2-3x+2 is the same as multiplication by zero in $\mathbb{R}[C]$ as $C^2-3C+2I = 0$. However this is not the case in $\mathbb{R}[D]$ as $D^2 - 3D + 2I \neq 0$. ■

Definition 167 *A non-empty subset N of a module M is a **submodule** if N is closed under addition and scalar multiplication.*

Example 168 *When we consider R as an R-module the submodules of R are the ideals.*

Example 169 *The submodules of a $\mathbb{Z}$-module are the subgroups.*

Example 170 *Let V be a vector space over a field F and $T\colon V \to V$ be a linear map defining V as an $F[x]$-module. Then the submodules of V as the T-invariant subspaces – i.e. those subspaces U of V such that $T(U) \subseteq U$.*

Example 171 *Given a module homomorphism $\phi\colon M \to N$ then*

$$\begin{aligned} \ker\phi &= \{m \in M : \phi(m) = 0\} \qquad \text{is a submodule of } M; \\ \operatorname{Im}\phi &= \{\phi(m) : m \in M\} \qquad \text{is a submodule of } N. \end{aligned}$$

Example 172 *The rational matrix*

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

defines a $\mathbb{Q}[x]$ module structure on $\mathbb{Q}^2$. The submodules of $\mathbb{Q}^2$ are the F-invariant subspaces and so include the eigenspaces

$$E_2 = \langle (1,1)^T \rangle, \qquad E_0 = \langle (1,-1)^T \rangle$$

and we see that, as a direct sum of $\mathbb{Q}[x]$-modules,

$$\mathbb{Q}^2 = E_2 \oplus E_0, \qquad \begin{pmatrix} a \\ b \end{pmatrix} = \left(\frac{a+b}{2}\right)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \left(\frac{a-b}{2}\right)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Definition 173 *(a) We say that elements $x_1, x_2, \ldots, x_n$ of an R-module M are **linearly independent** if the only solution of the equation*

$$r_1x_1 + r_2x_2 + \cdots + r_nx_n = 0, \qquad r_i \in R$$

is $r_1 = r_2 = \cdots = r_n = 0$.

*(b) We say that elements $x_1, x_2, \ldots, x_n$ of an R-module M **generate** or **span** M if every element x of M can be written*

$$x = r_1x_1 + r_2x_2 + \cdots + r_nx_n$$

for some $r_1, r_2, \ldots, r_n$ in R.

(c) We say that elements $x_1, x_2, \ldots, x_n$ of an R-module M form a basis for M if the elements are linearly independent and span M.

*(d) We say that a module M is **finitely generated** if there is a finite subset of M that generates M.*

*(e) A module with a basis is called a **free module**. **N.B. Most modules don't have bases.***

Example 174 *The elements $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ are a basis for R^n.*

Example 175 *The set $\{2\}$ in $\mathbb{Z}$ is linearly independent but not a basis.*

Example 176 *$\mathbb{Z}_2$ is not free as a $\mathbb{Z}$-module as $2.\bar{1} = \bar{0}$ but $2 \neq 0$. But $\mathbb{Z}_2$ is free as a $\mathbb{Z}_2$-module – of course it remains the case that $\bar{2}.\bar{1} = \bar{0}$ but now the scalar $\bar{2}$ is the zero scalar.*

Proposition 177 *Any basis of R^n contains n elements. n is known as the **rank** of R^n. Note an R-module with a basis containing n elements is isomorphic as an R-module to R^n.*

Proof. Say that $b_1, \ldots, b_m$ is a basis for R^n. Then we may write each b_i as

$$b_i = b_{1i}e_1 + b_{2i}e_2 + \cdots + b_{ni}e_n$$

so that the $n \times m$ matrix $B = (b_{ij})$ is the change of basis matrix from the e_i to the b_i. However as the b_i are also a basis, we can write the e_i in terms of the b_i and hence there is an $m \times n$ change of basis matrix A. We then have that

$$BA = I_n \qquad \text{and} \qquad AB = I_m.$$

If $m \neq n$ then we may assume without loss of generality that $m > n$. We can introduce the $m \times m$ matrices

$$A' = (A \,|\, 0_{m(m-n)}), \qquad B' = \begin{pmatrix} B \\ 0_{(m-n)m} \end{pmatrix}$$

which still satisfy $A'B' = I_m$. However we would then have

$$1 = \det I_m = \det A' \det B' = 0 \times 0 = 0.$$

A contradiction. (Note that the veracity of the determinant product rule relies only on the commutativity of the matrices' entries.)

If a module M has a basis $b_1, \ldots, b_n$ then every element x of M can be uniquely written as

$$x = r_1 b_1 + r_2 b_2 + \cdots + r_n b_n$$

and the identification $x \leftrightarrow (r_1, r_2, \ldots, r_n)$ is a module isomorphism with R^n. ■

Proposition 178 *A submodule M of the free module R^n is finitely generated.*

Proof. We shall prove this by induction on n. The submodule of R are the ideals which are principal, and so generated by a single element, proving the case $n = 1$. Say now that the result holds for R^{n-1} and let W be a submodule of R^n. Then

$$W_0 = \{\mathbf{w} = (w_1, \ldots, w_n) \in W : w_n = 0\}$$

is isomorphic to a submodule of R^{n-1} and so is finitely generated. If $W = W_0$ then we are done. Otherwise there is $\mathbf{x} \in W$ with $x_n \neq 0$ and with norm $d(x_n)$ being minimal. Say that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ generate W_0 By minimality of $d(x_n)$ for any $\mathbf{w} \in W$ we have $w_n = r x_n$ for some r and hence $\mathbf{w} - r\mathbf{x} \in W_0$. It follows that $\mathbf{y}_1, \ldots, \mathbf{y}_k, \mathbf{x}$ generate W and so W is finitely generated and the result follows by induction. ■

Corollary 179 *A submodule M of R^n is free and* $\operatorname{rank}(M) \leqslant n$.

Proof. From the proof of the proposition, we can also see how one might produce a basis for W with n or fewer elements. We can create a nested sequence of submodules

$$W_0^{(1)} = W_0, \qquad W_0^{(k)} = (W_0^{(k-1)})_0, \qquad \text{and} \qquad \{\mathbf{0}\} = W_0^{(n)} \subseteq W_0^{(n-1)} \subseteq \cdots \subseteq W_0^{(1)} \subseteq W,$$

so that $W_0^{(k)}$ consists of those elements of W whose last k entries are 0. From the previous proof we see that at each stage we need add at most one generator to the generators of $W_0^{(k)}$ to produce a set of generators for $W_0^{(k-1)}$ and won't need to add any generator if $W_0^{(k)} = W_0^{(k-1)}$. Further by the nature of how these generators are constructed they are linearly independent. Hence we can produce a basis for W containing at most n elements. ■

Remark 180 *Do make note of the important differences between the theory of vector spaces and the more general theory of modules.*

- *Most modules don't have bases – only the free modules.*
- *A linearly independent set cannot always be extended to a basis (even in a free module).*
- *A spanning set need not contain a basis (even in a free module)..*
- *A proper submodule of a free module can have the same rank.*

Definition 181 *A module M is said to be **cyclic** if it is generated by a single element. For an $F[x]$-module such a generator is called a **cyclic vector**.*

Example 182 *The cyclic $\mathbb{Z}$-modules are the cyclic groups.*

Example 183 *Given a square matrix A over a field F then $F[A]$ is a cyclic module as it is generated by I.*

Example 184 *The $\mathbb{R}[x]$-module defined on $\mathbb{R}^3$ by*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

is not cyclic. Note that $m_A(x) = (x-1)^2$ and so for any $\mathbf{v} \in \mathbb{R}^3$ we see $\mathbf{v}$ generates $\langle \mathbf{v}, A\mathbf{v} \rangle \neq \mathbb{R}^3$ as $A^2\mathbf{v} = 2A\mathbf{v} - \mathbf{v}$ and so there is no cyclic vector.

The x-axis M_1 and yz-plane M_2 however are submodules (they are A-invariant subspaces) and are cyclic as they are respectively generated by $\mathbf{i}$ and $\mathbf{k}$. So we can decompose the module

$$\mathbb{R}^3 = M_1 \oplus M_2$$

as the direct sum of cyclic submodules.

Proposition 185 *Let V be an n-dimensional vector space over F with an $F[x]$-module structure defined by $T\colon V \to V$. Say that V is cyclic as an $F[x]$-module, and that $v \in V$ is a cyclic vector. Then the vectors*

$$v, Tv, T^2v, \ldots, T^{n-1}v$$

form a basis for V as a vector space. With respect to this basis T has matrix

$$C(f) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \ddots & \vdots & -a_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix}$$

where the minimal and characteristic polynomials of T both equal

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0.$$

The matrix $C(f)$ is called the ***companion matrix*** *of $f(x)$.*

Proof. As v generates V as an $F[x]$-module then the set

$$v, Tv, T^2v, \ldots$$

spans V as a vector space. If k is the first occasion that

$$v, Tv, T^2v, \ldots, T^kv$$

are linearly dependent as vectors, then we have that $v, Tv, T^2v, \ldots, T^{k-1}v$ are independent and spanning and so $k = \dim V = n$.

Hence $v, Tv, T^2v, \ldots, T^{n-1}v$ is a basis for V as a vector space. That $T(T^kv) = T^{k+1}v$ for $0 \leqslant k \leqslant n-2$ accounts for the first $n-1$ columns of $C(f)$. Then for some $a_0, a_1, \ldots, a_{n-1} \in F$ we have

$$T^nv = -a_0v - a_1Tv - \cdots - a_{n-1}T^{n-1}v.$$

If we define $f(x)$ as above then we have $f(T)v = 0$. But as polynomials in T commute we also have $f(T)T^kv = 0$ for all k and hence $f(T) = 0$. It must be the case that $\det m_T \geqslant n$ as $v, Tv, T^2v, \ldots, T^{n-1}v$ are independent and also the case that $m_T | f$. As f is monic then $m_T = f$. As c_T has degree n and $m_T | c_T$ then we also have $m_T = c_T$. ■

Definition 186 *Given an R-module M and submodule N then we can form the* ***quotient R-module*** *M/N by*

$$M/N = \{m + N : m \in M\}$$

with

$$(m_1 + N) + (m_2 + N) \colon = (m_1 + m_2) + N, \qquad r.(m + N) \colon = r.m + N.$$

Example 187 *Whilst somewhat cumbersome in notation expressing a module as a quotient module can make transparent the module structure. For example*

$$\frac{\mathbb{R}[x]}{\langle x + 1 \rangle}$$

is isomorphic to the $\mathbb{R}[x]$-module defined on $\mathbb{R}$ by $-I_1 = (-1)$. All this information is captured in writing the module as above – namely that scalar multiplication by x acts in the same way as multiplication by -1.

Example 188 *Note that for any polynomial $p(x)$ in $F[x]$ that*

$$\frac{F[x]}{\langle p(x) \rangle}$$

is a cyclic $F[x]$-module as it is generated by 1.

Example 189 *The Chinese Remainder Theorem still holds for modules and so we have for example that*

$$\frac{\mathbb{R}[x]}{\langle x^2 - 1 \rangle} \cong \frac{\mathbb{R}[x]}{\langle x - 1 \rangle} \oplus \frac{\mathbb{R}[x]}{\langle x + 1 \rangle}$$

with a module isomorphism being given by

$$a + bx + \langle x^2 - 1 \rangle \mapsto (a + bx + \langle x - 1 \rangle, a + bx + \langle x + 1 \rangle) = (a + b + \langle x - 1 \rangle, a - b + \langle x + 1 \rangle).$$

Theorem 190 ***(First Isomorphism Theorem)*** *Let $\phi \colon M \to N$ be a module homomorphism. Then the map*

$$\tilde{\phi} \colon \frac{M}{\ker \phi} \to \operatorname{Im} \phi \qquad \textit{where} \quad \tilde{\phi}(m + \ker \phi) = \phi(m)$$

is a module isomorphism.

Proof. This proof is almost identical to the first isomorphism theorem for rings, with an extra line to check the R-module structures align. ■

Example 191 *We return here to the modules* $\mathbb{R}[C]$ *and* $\mathbb{R}[D]$ *introduced earlier where*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

We can decompose these modules in such a way that should make clearer the nature of the homomorphisms we found between them. We shall show that

$$\mathbb{R}[C] \cong \frac{\mathbb{R}[x]}{\langle x^2 - 3x + 2\rangle} \cong \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x-2\rangle}.$$

These isomorphisms are given by

$$p(C) \mapsto p(x) + \langle x^2 - 3x + 2\rangle \qquad \text{and by} \qquad p(x) + \langle x^2 - 3x + 2\rangle \mapsto (p(x) + \langle x-1\rangle, p(x) + \langle x-2\rangle)\,.$$

In a similar fashion

$$\mathbb{R}[D] \cong \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x-3\rangle}.$$

Whilst these may look somewhat cumbersome decompositions they contain explicitly and transparently the $\mathbb{R}[x]$*-module structure. Thus finding the module homomorphisms between these two modules is a lot more straightforward and the answer a lot clearer. Note for example that*

$$x.(\alpha, \beta) = (\alpha, 2\beta) \qquad \text{in} \quad \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x-2\rangle}.$$

A module homomorphism

$$\phi\colon \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x-2\rangle} \to \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle x-3\rangle}$$

is, in particular, an $\mathbb{R}$*-linear map and so is determined by* $\phi(1,0)$ *and* $\phi(0,1)$*. Say that* $\phi(1,0) = (a,b)$ *and* $\phi(0,1) = (c,d)$*; then, as* ϕ *is a module homomorphism, we have*

$$\begin{aligned} (a, 2b) &= x.(a,b) = x.\phi(1,0) = \phi(x.(1,0)) = \phi(1,0) = (a,b); \\ (c, 3d) &= x.(c,d) = x.\phi(0,1) = \phi(x.(0,1)) = \phi(0,2) = (2c, 2d). \end{aligned}$$

So we see that $b = c = d = 0$ *and that the only homomorphisms are of the form*

$$(\alpha, \beta) \mapsto (a\alpha, 0).$$

These module homomorphisms correspond to a scaling in the $\mathbb{R}[x]/\langle x-1\rangle$ *factor and a collapsing of the* $\mathbb{R}[x]/\langle x-2\rangle$*, hopefully not surprisingly as scalar multiplication by* x *in* $\mathbb{R}[D]$ *never corresponds to multiplication by* 2 *as it does in the second summand of* $\mathbb{R}[C]$*.*

Example 192 *We return to the $\mathbb{R}[x]$-module $\mathbb{R}^3$ defined by the matrix*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have already seen that this ideal is not cyclic but can be decomposed as the direct sum of the x-axis M_1 and the yz-plane M_2 which are cyclic submodumles. Somewhat more transparently we can represent the module structure by writing

$$M_1 \cong \frac{\mathbb{R}[x]}{\langle x-1 \rangle} \cong \mathbb{R}[I_1] \qquad \textit{and} \qquad M_2 \cong \frac{\mathbb{R}[x]}{\langle (x-1)^2 \rangle} \cong \mathbb{R}\left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right].$$

6. SMITH NORMAL FORM. PRESENTATIONS.

Throughout let R denote a Euclidean domain.

Definition 193 *(a) A square matrix P with entries in R is said to be **invertible** if there is a matrix Q with entries in R such that $PQ = I = QR$.*

*(b) Two $m \times n$ matrices A and B are said to be **equivalent** if there exist an invertible $m \times m$ matrix P and an invertible $n \times n$ matrix Q such that $PAQ = B$. A simple check shows that equivalence is indeed an equivalence relation.*

Example 194 *The matrix* $\mathrm{diag}(1, 2)$ *is not invertible over $\mathbb{Z}$ but would be, say, over $\mathbb{Q}[x]$.*

Theorem 195 ***(Smith Normal Form)*** *Let A be an $m \times n$ matrix with entries in R. Then there exist elements $d_1, d_2, \ldots, d_r$, known as **invariant factors**, and unique up to multiplication by units, such that A is equivalent to*

$$\begin{pmatrix} \mathrm{diag}(d_1, d_2, \ldots, d_r) & 0_{r(n-r)} \\ 0_{(m-r)r} & 0_{(m-r)(n-r)} \end{pmatrix}$$

where

$$d_1 | d_2 | d_3 | \cdots | d_r.$$

Proof. *Existence*: Our first aim is to employ EROs and ECOs to show that A is equivalent to

$$\begin{pmatrix} d_1 & 0_{1(n-1)} \\ 0_{(m-1)1} & M \end{pmatrix}$$

where d_1 divides every entry of M.

Step One: By permuting rows and columns we can move an entry d of least norm to the first-row-first-column entry. We now aim to clear out the first row using ECOs: note every entry of the first row can be written in the form $a_{1j} = q_j d + r_j$ and if a remainder r_j of smaller norm is produced we permute it to the top left entry and begin again. This process must terminate as the norm we are continually decreasing takes a positive integral value.

Step Two: We then proceed similarly seeking to clear out the first column using EROs. Should we produce an entry of smaller norm then we move that row to the first row and return to Step One.

Step Three: Eventually the first row and first column have been cleared and it follows that the matrix A can be put in the above form but it may not be the case that d divides every entry of M. If all of the entries of M are divisible by d then we are done. If not then by EROs and ECOs we can again produce an element of smaller norm than d and we return to Step One.

The process must ultimately terminate as at each state we are producing entries of strictly smaller norm. Thus we have demonstrated that A is equivalent to a matrix of the above form and by repeating this process on M and so on we have shown the existence of the Smith normal form.

Uniqueness: To prove uniqueness, we introduce the notion of *determinantal divisors*. The ith determinantal divisor $D_i(A)$ is the highest common factor of the determinants of all $i \times i$ submatrices of A. Note that $D_i(A)$ is only defined up to multiplication by units and is invariant under EROs and ECOs. When the matrix is in Smith normal form, as above, then we see that

$$D_i(A) = d_1 d_2 \cdots d_i \qquad \text{for } i \leqslant r \qquad \text{and} \qquad D_i(A) = 0 \qquad \text{for } i > r.$$

Hence $d_k = D_k(A)/D_{k-1}(A)$ is invariant under EROs and ECOs and so the invariant factors are unique. ■

Corollary 196 *(**Submodules of Free Modules**) Let M be a submodule of R^n. Then there exist elements $d_1, d_2, \ldots, d_r$, where r is the rank of M, and a basis $f_1, \ldots, f_n$ for R^n such that $d_1 f_1, d_2 f_2, \ldots, d_r f_r$ is a basis for M and $d_1 | d_2 | d_3 | \cdots | d_r$.*

To see why this corollary follows we will apply the method to the following matrix.

Example 197 *Put the following matrix in Smith normal form*

$$\begin{pmatrix} 12 & 6 & 4 & 8 \\ 3 & 9 & 6 & 12 \\ 2 & 16 & 14 & 28 \\ 20 & 10 & 10 & 20 \end{pmatrix}.$$

Solution. We will do more than simply put the matrix into Smith normal form and keep track in particular of the ECOs that we are using. We find

$$\begin{pmatrix} e_1 & e_2 & e_3 & e_4 \\ 12 & 6 & 4 & 8 \\ 3 & 9 & 6 & 12 \\ 2 & 16 & 14 & 28 \\ 20 & 10 & 10 & 20 \end{pmatrix} \sim \begin{pmatrix} e_1 & e_2 & e_3 + 2e_4 & e_4 \\ 12 & 6 & 4 & 0 \\ 3 & 9 & 6 & 0 \\ 2 & 16 & 14 & 0 \\ 20 & 10 & 10 & 0 \end{pmatrix} \sim \begin{pmatrix} e_1 & e_2 & e_3 + 2e_4 & e_4 \\ 12 & 6 & 4 & 0 \\ 1 & -7 & -8 & 0 \\ 2 & 16 & 14 & 0 \\ 20 & 10 & 10 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} e_1 & e_2 & e_3 + 2e_4 & e_4 \\ 0 & 90 & 100 & 0 \\ 1 & -7 & -8 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 150 & 170 & 0 \end{pmatrix} \sim \begin{pmatrix} e_1 & e_2 & e_3 + 2e_4 & e_4 \\ 0 & 0 & 10 & 0 \\ 1 & -7 & -8 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 0 & 20 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} e_1 & e_2 & e_3 + 2e_4 & e_4 \\ 0 & 0 & 10 & 0 \\ 1 & -7 & -8 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} e_1 - 7e_2 - 8e_3 - 16e_4 & e_3 + 2e_4 & e_2 & e_4 \\ 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence the Smith normal form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

■

But we are also able to answer the following.

Example 198 *Find a basis f_1, f_2, f_3, f_4 for $\mathbb{Z}^4$ such that $d_1 f_1, \ldots, d_r f_r$ is a basis for*

$$M = \langle (12,6,4,8),\ (3,9,6,12),\ (2,16,14,28), (20,10,10,20) \rangle$$

where $r = \operatorname{rank}(M)$ and $d_1 | d_2 | \cdots | d_r$.

Solution. In our previous calculation we saw that

$$\begin{pmatrix} e_1 & e_2 & e_3 & e_4 \\ 12 & 6 & 4 & 8 \\ 3 & 9 & 6 & 12 \\ 2 & 16 & 14 & 28 \\ 20 & 10 & 10 & 20 \end{pmatrix} \sim \begin{pmatrix} e_1 - 7e_2 - 8e_3 - 16e_4 & e_3 + 2e_4 & e_2 & e_4 \\ 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The column headings are still a basis for $\mathbb{Z}^4$ as we began with a basis e_1, e_2, e_3, e_4 and each ECO is invertible. The rows (with co-ordinates understood wrt this new basis) still span M and are now clearly independent. So we have shown that

$$f_1 = e_1 - 7e_2 - 8e_3 - 16e_4, \qquad f_2 = e_3 + 2e_4, \qquad f_3 = e_2, \qquad f_4 = e_4$$

is a basis for $\mathbb{Z}^4$ and $f_1, 10f_2, 30f_3$ is a basis for M. ■

Before we move on to the Structure Theorem we will need to introduce the ideas of generators, relations and presentations. We begin with some motivational examples.

Example 199 *Consider the following modules.*

$$M_1 = \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_5, \qquad M_2 = \mathbb{R}[x] \oplus \frac{\mathbb{R}[x]}{\langle x-1 \rangle} \oplus \frac{\mathbb{R}[x]}{\langle x^2+1 \rangle}.$$

The $\mathbb{Z}$-module M_1 is generated by

$$a = (1,0,0), \qquad b = (0,1,0), \qquad c = (0,0,1).$$

The module is not free and we see for example the "relations" $2a = 4b = 5c = 0$. We could have chosen other generators for M_1 such as a and $b+c$ which satisfy the relations $2a = 20(b+c) = 0$. These facts are represented in the isomorphisms

$$M_1 \cong \frac{\mathbb{Z}^3}{\langle (2,0,0), (0,4,0), (0,0,5) \rangle} \cong \frac{\mathbb{Z}^2}{\langle (2,0), (0,20) \rangle}.$$

Note that every relation involving a and $b+c$ can be deduced from $2a = 20(b+c) = 0$ with these relations generating all relations.

For M_2 we note again that a, b, c are generators with $(x-1)b = (x^2+1)c = 0$ or again might have been generated by a and $b+c$. Again we have

$$M_2 \cong \frac{\mathbb{R}[x]^3}{\langle (0, x-1, 0), (0, 0, x^2+1) \rangle} \cong \frac{\mathbb{R}[x]^2}{\langle (0,0), (0, (x-1)(x^2+1)) \rangle}.$$

For each module, and for each different set of generators, the relations these generators satisfy can be captured in a "presentation" matrix. In the case of M_1 these would be the matrices

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix} \qquad \textit{and} \qquad \begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix}$$

where the columns in the first matrix relate to a, b, c and in the second to $a, b+c$. For M_2 the two presentation matrices would be

$$\begin{pmatrix} 0 & x-1 & 0 \\ 0 & 0 & x^2+1 \end{pmatrix} \qquad \textit{and} \qquad \begin{pmatrix} 0 & 0 \\ 0 & (x-1)(x^2+1) \end{pmatrix}.$$

Definition 200 *(a) As we have defined already the elements $x_1, \ldots, x_n$ in the R-module M are **generators** if every element x of M can be written in the form*

$$x = r_1x_1 + r_2x_2 + \cdots + r_nx_n.$$

*(b) A **relation** in the generators $x_1, \ldots, x_n$ is any (trivial or non-trivial) combination that adds to 0, that is*

$$r_1x_1 + r_2x_2 + \cdots + r_nx_n = 0.$$

Proposition 201 *Given a finitely generated R-module M with generators $x_1, \ldots, x_n$ there is an onto module homomorphism $\phi\colon R^n \to M$ and the relations in $x_1, \ldots, x_n$ correspond to $\ker\phi$.*

Proof. There is a module homomorphism

$$\phi\colon R^n \to M \qquad \text{given by} \quad e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \mapsto x_i.$$

As the x_i generate M then ϕ is onto and we have, by the first isomorphism theorem, that

$$\frac{R^n}{\ker\phi} \cong \operatorname{Im}\phi = M.$$

Further

$$r_1x_1 + r_2x_2 + \cdots + r_nx_n = 0$$

is a relation if and only if $(r_1, r_2, \ldots, r_n) \in \ker\phi$. ■

Corollary 202 *A finitely generated module is isomorphic to a quotient of a free module.*

Proof. This is immediate from the above by the First Isomorphism Theorem as $M \cong R^n/\ker\phi$. ■

Corollary 203 *A submodule of a finitely generated module is finitely generated.*

Proof. If M is finitely generated then $M \cong R^n/K$ for some submodule K of R^n. There is a correspondence between the submodules of R^n/K and the submodules of R^n which contain K (as with rings). So a submodule N of M is of the form P/K where P is a submodule of R^n. As submodules of free modules are free, and so finitely generated, then $N = P/K$ is finitely generated. ■

Definition 204 *Let M be an R-module generated by $x_1, \ldots, x_n$ and $\phi\colon R^n \to M$ be as above. Then the* ***relation module*** $\ker\phi$ *is finitely generated, say by the relations*

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \qquad \cdots \qquad a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0.$$

Then the ***presentation matrix*** *for M with respect to these generators and relations is $A = (a_{ij})$. As $\ker\phi$ is free the relations can be chosen to be a basis (but do not need to be).*

Example 205 *Say that A is the abelian group generated by a, b, c subject to the relations*

$$6a + 18b + 12c = 0, \qquad 12a - 9b + 15c = 0, \qquad 9a - 12b - 24c = 0.$$

Show that A is isomorphic to $\mathbb{Z}_3 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{414}$ and find, in terms of a, b, c all elements of order 3.

Solution. This module is

$$A = \frac{\mathbb{Z}^3}{\langle (6,18,12), (12,-9,15), (9,-12,-24)\rangle}$$

though this is not a particularly informative or transparent presentation. Rather we put the presentation matrix into Smith normal form as follows:

$$\begin{pmatrix} a & b & c \\ 6 & 18 & 12 \\ 12 & -9 & 15 \\ 9 & -12 & 24 \end{pmatrix} \sim \begin{pmatrix} a & b & c \\ 6 & 18 & 12 \\ 0 & -45 & -9 \\ 3 & -30 & 12 \end{pmatrix} \sim \begin{pmatrix} a & b & c \\ 3 & -30 & 12 \\ 0 & 45 & 9 \\ 0 & 78 & -12 \end{pmatrix} \sim$$

$$\begin{pmatrix} a-10b+4c & b & c \\ 3 & 0 & 0 \\ 0 & 45 & 9 \\ 0 & 78 & -12 \end{pmatrix} \sim \begin{pmatrix} a-10b+4c & b & 5b+c \\ 3 & 0 & 0 \\ 0 & 0 & 9 \\ 0 & 138 & -12 \end{pmatrix} \sim \begin{pmatrix} a-10b+4c & 5b+c & b \\ 3 & 0 & 0 \\ 0 & -3 & 138 \\ 0 & 9 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} a-10b+4c & 5b+c & b \\ 3 & 0 & 0 \\ 0 & 3 & -138 \\ 0 & 0 & 414 \end{pmatrix} \sim \begin{pmatrix} a-10b+4c & c-41b & b \\ 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 414 \end{pmatrix}$$

Hence if we set $\alpha = a - 10b + 4c, \beta = c - 41b, \gamma = b$ then we see

$$A = \langle \alpha, \beta, \gamma : 3\alpha = 3\beta = 414\gamma = 0\rangle \cong \mathbb{Z}_3 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_{414}.$$

We can see then that there are 26 elements of order three and that these are of the form

$$\varepsilon_1\alpha + \varepsilon_2\beta + 138\varepsilon_3\gamma \qquad \text{where } \varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{0,1,2\} \text{ without all being zero.}$$

■

Example 206 *Consider the $\mathbb{R}[x]$-module M defined on $\mathbb{R}^3$ by*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have already noted that this module is not cyclic; but we can see that e_1, e_3 are generators as $x.e_3 = Ae_3 = e_2 + e_3$.

Now note that $x.e_1 = e_1$ so that $(x-1).e_1 = 0$ and we also have

$$x^2.e_3 = x.(e_2 + e_3) = 2e_2 + e_3 = 2x.e_3 - e_3$$

so that $(x-1)^2.e_3 = 0$. So

$$(x-1).e_1 = 0, \qquad (x-1)^2.e_3 = 0$$

are relations. In fact these relations are sufficient to generate all relations. To appreciate this note that $\dim_{\mathbb{R}} M = 3$ and that the map

$$\phi\colon M \to \frac{\mathbb{R}[x]^2}{\langle((x-1),0),(0,(x-1)^2)\rangle} \cong \frac{\mathbb{R}[x]}{\langle x-1\rangle} \oplus \frac{\mathbb{R}[x]}{\langle (x-1)^2\rangle}$$

given by $\phi(e_1) = (1,0)$, $\phi(e_2) = (0, x-1)$, $\phi(e_3) = (0,1)$ is an $\mathbb{R}[x]$-homomorphism and bijective as the LHS and RHS are 3-dimensional real vector spaces.

Given a vector space V over a field F with basis $e_1, \ldots, e_n$ and defined as a $F[x]$-module via a matrix A, then V is generated by $e_1, e_2, \ldots, e_n$ and we have the relations $xe_i - Ae_i = 0$ for each i. In fact these relations generate the relation module as the entire module structure is a consequence of linearity and inductive use of these relations. Thus the presentation matrix for V with respect to generators $e_1, e_2, \ldots, e_n$ and the relations $xe_i - Ae_i = 0$ is the matrix

$$xI_n - A.$$

Example 207 *Let $V = \mathbb{C}^3$ and*

$$T = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Put the presentation matrix $xI - T$ into Smith normal form and find the invariant factors. Find generators for V. What is the minimal polynomial of T?

Solution. We now put the generators on the left like so

$$\begin{pmatrix} e_1 & x-1 & 1 & -1 \\ e_2 & 0 & x & -1 \\ e_3 & 0 & -1 & x \end{pmatrix}$$

so that the first column represents $(x-1).e_1 = 0$ and the second that $x.e_2 + e_1 - e_3 = 0$ etc. So now EROs will change the generators and ECOs have no effect. Then, placing this in Smith Normal form we find

$$\begin{pmatrix} e_1 & x-1 & 1 & -1 \\ e_2 & 0 & x & -1 \\ e_3 & 0 & -1 & x \end{pmatrix} \sim \begin{pmatrix} e_1 & x-1 & 0 & x-1 \\ e_2 & 0 & 0 & x^2-1 \\ e_3 - e_1 - xe_2 & 0 & -1 & x \end{pmatrix}$$

$$\sim \begin{pmatrix} e_1 & x-1 & 0 & 0 \\ e_2 & 0 & 0 & x^2-1 \\ e_1 + xe_2 - e_3 & 0 & 1 & 0 \end{pmatrix} \sim\sim \begin{pmatrix} e_1 + xe_2 - e_3 & 1 & 0 & 0 \\ e_1 & 0 & x-1 & 0 \\ e_2 & 0 & 0 & x^2-1 \end{pmatrix}$$

Hence the invariant factors are $x-1$ and x^2-1. This also means that the module defined by T is isomorphic to

$$V = \frac{\mathbb{C}[x]}{\langle x-1 \rangle} \oplus \frac{\mathbb{C}[x]}{\langle x^2-1 \rangle}$$

with generators being e_1 and e_2 and relations being

$$(x-1).e_1 = 0, \qquad (x^2-1).e_2 = 0.$$

It also follows that the minimal polynomial of T is x^2-1. ■

7. STRUCTURE THEOREM. APPLICATIONS.

Definition 208 *Let M be a finitely-generated R-module. An element $m \in M$ is said to be a* ***torsion element****, if $r.m = 0$ for some $r \neq 0$. M is said to have* ***torsion*** *if it has any non-zero torsion elements. M is said to be torsion-free if 0 is the only torsion element.*

Proposition 209 *(a) Let M be a finitely-generated R-module. The torsion elements form a submodule T of M.*

(b) A finitely generated torsion-free R-module is free.

(c) The module $F = M/T$ is a free R-module and $M \cong F \oplus T$.

Proof. (a) Note that $0 \in T$. If $m_1, m_2 \in T$ then there exist non-zero r_1, r_2 such that $r_1.m_1 = r_2.m_2 = 0$. Then $r_1 r_2 \neq 0$ and we have

$$r_1 r_2.(m_1 + m_2) = 0.$$

Further for $r \in R$ we have that $rm_1 \in T$ as

$$r_1.(rm) = r_1 r.m = r r_1.m = r.r_1 m = 0.$$

Hence T is a submodule.

(b) Let N be a finitely generated torsion-free R-module. Say that $x_1, \ldots, x_n$ generate N and that (by reordering if necessary) $x_1, \ldots, x_m$ is a maximal linearly independent subset of the x_i. We set F to be the span of $x_1, \ldots, x_m$, noting that F is free. Note that there exist $r_i, r_{ij} \in R$ not all zero and such that

$$r_i x_i + \sum_{j=1}^{m} r_{ij} x_j = 0.$$

This is clear for $i \leqslant m$ and follows by the maximality of m for $i > m$. Further as $x_1, \ldots, x_m$ are independent then $r_i \neq 0$ for all i. We set

$$r = r_1 r_2 \cdots r_n \neq 0.$$

Note that $r_i x_i$ is in F for all i and hence $r x_i \in F$ for all i. But then $rN \subseteq F$. Note that the map $n \mapsto r.n$ is an injective module homomomorphism as N is torsion-free. So N is isomorphic to its image which is a submodule of the free module F and so free itself.

(c) As M is finitely generated then $F = M/T$ is finitely generated and is torsion free, and so free, by construction. Let $x_1, \ldots, x_n$ be a basis for F. Then any element of M/T can be uniquely written as

$$r_1 x_1 + \cdots + r_n x_n + T.$$

Then the map ϕ: $M/T \oplus T \to M$ given by

$$\phi(r_1 x_1 + \cdots + r_n x_n + T, t) = r_1 x_1 + \cdots + r_n x_n + t$$

is a module isomorphism. ■

Example 210 *A linear map $T\colon V \to V$ on a finite dimensional vector space V over F defines an $F[x]$-module. In this module every element is a torsion element as $m_T(x).v = 0$ for all v yet $m_T(x) \neq 0 \in F[x]$.*

Theorem 211 ***(Structure Theorem for Finitely Generated Modules)*** *Let M be a finitely-generated R-module. Then there exists a non-negative integer r, called the (torsion-free)* ***rank*** *of M and non-zero, non-unit elements $d_i \in R$, known as the* ***invariant factors*** *such that*

$$d_1|d_2|d_3|\cdots|d_k$$

and such that

$$M \cong R^r \oplus \frac{R}{\langle d_1\rangle} \oplus \frac{R}{\langle d_2\rangle} \oplus \cdots \oplus \frac{R}{\langle d_k\rangle}.$$

The rank r is unique and $d_1, \ldots, d_k$ unique up to multiplication by units.

Proof. Say that M, which is finitely generated, is generated by $x_1, \ldots, x_n$. There is then a module homomorphism

$$\phi\colon R^n \to M \qquad \text{given by} \quad e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \mapsto x_i.$$

As the x_i generate M then ϕ is onto and we have, by the first isomorphism theorem, that

$$\frac{R^n}{\ker \phi} \cong \operatorname{Im} \phi = M.$$

Now by the corollary to the Smith normal form, we know that there is a basis $f_1, \ldots, f_n$ for R^n and d_i as above with

$$\ker \phi = 0 \oplus \cdots \oplus 0 \oplus \langle d_1\rangle \oplus \cdots \oplus \langle d_k\rangle$$

and hence we have

$$M \cong \frac{R^n}{0 \oplus \cdots \oplus 0 \oplus \langle d_1\rangle \oplus \cdots \oplus \langle d_k\rangle} \cong R^{n-k} \oplus \frac{R}{\langle d_1\rangle} \oplus \frac{R}{\langle d_2\rangle} \oplus \cdots \oplus \frac{R}{\langle d_k\rangle}.$$

Had any of the d_i been units then we would have $R/\langle d_i\rangle = R/R \cong 0$ and we can just omit such factors.

Now note that $d_k M = d_k R^r$ is a free module and so the rank of M is uniquely determined. However we will postpone for now proving the uniqueness of invariant factors until we have introduced the notion of elementary divisors. ■

Example 212 *Present each of the following modules as described in the structure theorem.*

(a) $\mathbb{Z}_6 \oplus \mathbb{Z}_{12} \oplus \mathbb{Z}_{16} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_{16} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_{12} \oplus \mathbb{Z}_{48}$.

(b)

$$\frac{\mathbb{Q}[x]}{\langle (x^2-4)(x^3-8)\rangle}.$$

This is already cyclic (it is generated by 1) and so is in the required form.

(c)

$$\frac{\mathbb{Z}[i]}{\langle 2\rangle} \oplus \frac{\mathbb{Z}[i]}{\langle 4\rangle} \oplus \frac{\mathbb{Z}[i]}{\langle 5\rangle} \cong \frac{\mathbb{Z}[i]}{\langle 2\rangle} \oplus \frac{\mathbb{Z}[i]}{\langle 20\rangle}.$$

Note that all of the above three modules all have zero rank.

There is an alternative form of the structure theorem. The above decomposition might be viewed as a minimal decomposition into cyclic submodules. Alternatively we can decompose some of the summands yet further by using the Chinese remainder theorem applied to the coprime factors of the invariant factors. Let $p_1, \ldots, p_n$ be the prime factors of at least one of the d_i. Then we may write

$$d_i = p_1^{\alpha_{1i}} p_2^{\alpha_{2i}} \cdots p_n^{\alpha_{ni}}$$

where $0 \leqslant \alpha_{ki} \leqslant \alpha_{(k+1)i}$. Then we have

$$\frac{R}{\langle d_i \rangle} \cong \frac{R}{\langle p_1^{\alpha_{1i}} \rangle} \oplus \frac{R}{\langle p_2^{\alpha_{2i}} \rangle} v \cdots \oplus \frac{R}{\langle p_n^{\alpha_{ni}} \rangle}$$

so that

$$M \cong \left(\bigoplus_{i=1}^{k} \frac{R}{\langle p_1^{\alpha_{1i}} \rangle} \right) \oplus \left(\bigoplus_{i=1}^{k} \frac{R}{\langle p_2^{\alpha_{2i}} \rangle} \right) \oplus \cdots \oplus \left(\bigoplus_{i=1}^{k} \frac{R}{\langle p_n^{\alpha_{ni}} \rangle} \right).$$

The elements $p_j^{\alpha_{ji}}$, where $\alpha_{ji} > 0$, are known as the **elementary divisors**.

Example 213 *Applying this alternative decomposition to the previous three examples we would write*

$$\begin{array}{rcl} \mathbb{Z}_6 \oplus \mathbb{Z}_{12} \oplus \mathbb{Z}_{16} & \cong & \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_{16} \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_3. \\ \dfrac{\mathbb{Q}[x]}{\langle (x^2-4)(x^3-8) \rangle} & \cong & \dfrac{\mathbb{Q}[x]}{\langle (x-2)^2 \rangle} \oplus \dfrac{\mathbb{Q}[x]}{\langle x+2 \rangle} \oplus \dfrac{\mathbb{Q}[x]}{\langle x^2+2x+4 \rangle}. \\ \dfrac{\mathbb{Z}[i]}{\langle 2 \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle 4 \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle 5 \rangle} & \cong & \dfrac{\mathbb{Z}[i]}{\langle 1+i \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle (1+i)^2 \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle 1-i \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle (1-i)^2 \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle 1+2i \rangle} \oplus \dfrac{\mathbb{Z}[i]}{\langle 1-2i \rangle}. \end{array}$$

Proposition 214 *The elementary divisors and invariant factors are unique (up to multiplication by units).*

Proof. We can separately consider the different irreducibles p for which there is some non-zero element x of M satisfying $p^n x = 0$ for a power of p, because different (non-associate) irreducibles will be coprime. So say that M is a module with $p^n M = 0$ for some positive integer n, and choose n to be the least such n. Then we have that $p^{n-1}M$ is a non-zero vector space over the field $R/\langle p \rangle$ for if $\overline{r_1} = \overline{r_2}$ in $R/\langle p \rangle$ then $r_1 - r_2 = rp$ for some r and we have

$$r_1.p^{n-1}m = (r_2 + rp).p^{n-1}m = r_2.p^{n-1}m.$$

The dimension of $p^{n-1}M$ as an $R/\langle p \rangle$-vector space is the number of copies of $R/\langle p^{n-1} \rangle$ in the decomposition and so recoverable from M. In a similar fashion $p^{n-2}M/p^{n-1}M$ is an $R/\langle p \rangle$-vector space and its dimension is the total number of copies of $R/\langle p^{n-2} \rangle$ and $R/\langle p^{n-1} \rangle$ in the decomposition, and subtracting our previously found dimension we now know the number of summands of $R/\langle p^{n-2} \rangle$ in the decomposition. Continuing in this fashion we are able to determine the number of each different summand. The invariant factors are then recoverable from the elementary divisors in a straightforward, but notationally painful manner. Begin by putting the highest power of each irreducible among the elementary divisors into d_k and keep repeating this process to produce all the invariant factors. ■

Example 215 *To help understand the previous proposition, here is the argument made for a specific $\mathbb{Z}$-module*

$$M = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_8 \oplus \mathbb{Z}_8.$$

Note that $8M = 0$ and that

$$4M = 0 \oplus 0 \oplus 0 \oplus \langle 4 \rangle \oplus \langle 4 \rangle \cong \mathbb{Z}_2^2.$$

This dimension, 2 is the number of $\mathbb{Z}_8$ summands. Now

$$\frac{2M}{4M} \cong \frac{0 \oplus 0 \oplus \langle 2 \rangle \oplus \langle 2 \rangle \oplus \langle 2 \rangle}{0 \oplus 0 \oplus 0 \oplus \langle 4 \rangle \oplus \langle 4 \rangle} \cong \mathbb{Z}_2^3.$$

This dimension, 3, is the number of $\mathbb{Z}_4$ and $\mathbb{Z}_8$ summands in total. Finally

$$\frac{M}{2M} = \frac{\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \langle 1 \rangle \oplus \langle 1 \rangle \oplus \langle 1 \rangle}{0 \oplus 0 \oplus \langle 2 \rangle \oplus \langle 2 \rangle \oplus \langle 2 \rangle} \cong \mathbb{Z}_2^5$$

and this dimension 5 is the total number of $\mathbb{Z}_2, \mathbb{Z}_4, \mathbb{Z}_8$ summands. So we see that the number of each elementary divisor is recoverable from these dimensions.

Corollary 216 ***(Classification Theorem for Finitely Generated Abelian Groups)*** *Let A be a finitely generated abelian group. Then there exist unique non-negative r and integers $d_i \geqslant 2$ with $d_1 | d_2 | \cdots | d_k$ such that*

$$A \cong \mathbb{Z}^r \oplus \mathbb{Z}_{d_1} \oplus \mathbb{Z}_{d_2} \oplus \cdots \oplus \mathbb{Z}_{d_k}.$$

Proof. This is simply a statement of the Structure Theorem for $\mathbb{Z}$-modules. ■

Example 217 *Find, up to isomorphism, all abelian groups of order 360. What are their elementary divisors?*

If not explicitly on your list, explain which of your groups is isomorphic to $\mathbb{Z}_4 \oplus \mathbb{Z}_{90}$?

Solution. Note that $360 = 2^3 3^2 5$. Hence $k \leqslant 3$ and we must have $2 \times 3 \times 5 = 30 | d_k$. We see that the only abelian groups, up to isomorphism, are

$$\mathbb{Z}_{360},, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_{180}, \qquad \mathbb{Z}_3 \oplus \mathbb{Z}_{120}, \qquad \mathbb{Z}_6 \oplus \mathbb{Z}_{60}, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_{90}, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_6 \oplus \mathbb{Z}_{30}.$$

The elementary divisors of these groups are respectively

$$8, 9, 5, \qquad 2, 4, 9, 5, \qquad 8, 3, 3, 5, \qquad 2, 4, 3, 3, 5, \qquad 2, 2, 2, 9, 5, \qquad 2, 2, 2, 3, 3, 5.$$

Now note that

$$\mathbb{Z}_4 \oplus \mathbb{Z}_{90} \cong \mathbb{Z}_4 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_{45} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_{180}$$

or we might have noted its elementary divisors to be $4, 2, 5, 9$ and so it is the second listed group. ■

Example 218 *Identify the abelian group generated by four elements a, b, c, d subject to the relations*

$$12a{+}6b{+}4c{+}8d = 0, \quad 3a{+}9b{+}6c{+}12d = 0, \quad 2a{+}16b{+}15c{+}28d = 0, \quad 20a{+}10b{+}10c{+}20d = 0.$$

Solution. We know that we can use EROs and ECOs to put the matrix into Smith normal form, and from our previous calculation we have

$$\begin{pmatrix} a & b & c & d \\ 12 & 6 & 4 & 8 \\ 3 & 9 & 6 & 12 \\ 2 & 16 & 14 & 28 \\ 20 & 10 & 10 & 20 \end{pmatrix} \sim \begin{pmatrix} a-7b-8c-16d & c+2d & b & d \\ 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Hence if we set

$$\alpha = a-7b-8c-16d, \qquad \beta = c+2d, \qquad \gamma = b, \qquad \delta = d$$

then the described abelian group is

$$\langle \alpha, \beta, \gamma, \delta : \alpha = 10\beta = 30\gamma = 0 \rangle \cong \mathbb{Z}_{10} \oplus \mathbb{Z}_{30} \oplus \mathbb{Z}.$$

There are two elements of order 3 for example, namely $(0, \pm 10, 0)$ or in terms of the generators these are

$$\pm 10\beta = \pm(10c + 20d).$$

■

Corollary 219 *(**Rational Canonical Form**) Let A be an $n \times n$ matrix over a field F. Then there A is similar to a matrix in the form*

$$\mathrm{diag}(C(d_1), C(d_2), \ldots C(d_k))$$

where $d_i \in F[x]$ are monic polynomials, $C(d_i)$ denotes the companion matrix of d_i and

$$d_1 | d_2 | \cdots | d_k.$$

The d_i are unique up to multiplication by units. Note further that

$$m_A(x) = d_k(x) \qquad \textit{and that} \qquad c_A(x) = d_1(x)d_2(x)\cdots d_k(x).$$

*The above matrix representative of A is known as its **rational canonical form** or **Frobenius normal form**.*

Remark 220 *Equivalently, in terms of the $F[x]$-module structure defined on F^n by A, the above says that*

$$F^n \cong F[C(d_1)] \oplus F[C(d_2)] \oplus \cdots \oplus F[C(d_k)],$$

hence decomposing the $F[x]$-module defined by A, which in general will not be cyclic, into cyclic $F[x]$-modules defined by the above companion matrices.

Proof. Consider the $F[x]$-module structure defined on F^n by the linear map $T(\mathbf{v}) = A\mathbf{v}$. This module has rank zero and every element is a torsion element. By the Structure Theorem we know that there exist invariant factors $d_i \in F[x]$ such that

$$d_1|d_2|\cdots|d_k$$

with

$$F^n \cong \frac{F[x]}{\langle d_1(x)\rangle} \oplus \frac{F[x]}{\langle d_2(x)\rangle} \oplus \cdots \oplus \frac{F[x]}{\langle d_k(x)\rangle}.$$

If $\deg d_1 = n_1$ then $1, x, \ldots, x^{n_1-1}$ is a basis for $F[x]/\langle d_1(x)\rangle$ as a vector space and with respect to this basis multiplication by x, or equivalently by A, on $F[x]/\langle d_1(x)\rangle$ is given by the companion matrix $C(d_1)$. Taking such a basis for each summand, the union of these bases is a basis for F^n and with respect to this basis T has matrix representative

$$\mathrm{diag}(C(d_1), C(d_2), \ldots C(d_k))$$

to which A is similar. (In choosing the above basis we have found a change of basis matrix P such that $P^{-1}AP$ equals the above matrix representative.) Conversely any such matrix representation leads to a decomposition of the $F[x]$-module as above and we know that the invariant factors are unique up to multiplication by units.

For elements in the final summand $F[x]/\langle d_k(x)\rangle$ we know that the minimal polynomial is $d_k(x)$. As $d_i|d_k$ for all i then $d_k(A)$ also annihilates the other summands and we see that $m_A(x) = d_k(x)$. We also have that the characteristic polynomial of $C(d_i)$ is $d_i(x)$; as the characteristic polynomial of the above matrix representative equals the product of the characteristic polynomials of the blocks we have

$$c_A(x) = d_1(x)d_2(x)\cdots d_k(x).$$

■

Corollary 221 ***(Jordan Normal Form)*** *Let A be a complex $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then A is similar to a matrix of the form*

$$\mathrm{diag}(J(\lambda_1, r_{11}), J(\lambda_1, r_{12}), \ldots J(\lambda_1, r_{1n_1}), \ldots J(\lambda_k, r_{k1}), J(\lambda_1, r_{k2}), \ldots J(\lambda_k, r_{kn_k}))$$

where $J(\lambda, r)$ denotes the $r \times r$ Jordan block matrix

$$J(\lambda, r) = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 1 & \lambda & 0 & \ddots & \vdots \\ 0 & 1 & \lambda & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \lambda \end{pmatrix}$$

and

$$r_{i1} \leqslant r_{i2} \leqslant \cdots \leqslant r_{in_i} \qquad \text{for each } i.$$

Note that

$$\mathrm{nullity}(A - \lambda_i I)^j - \mathrm{nullity}(A - \lambda_i I)^{j-1} = \text{number of } J(\lambda_i, r) \text{ blocks with } r \geqslant j.$$

Remark 222 *Equivalently, in terms of the $\mathbb{C}[x]$-module structure defined on $\mathbb{C}^n$ by A, the above says that*

$$\mathbb{C}^n \cong \mathbb{C}[J(\lambda_1, r_{11})] \oplus \mathbb{C}[J(\lambda_1, r_{12})] \oplus \cdots \oplus \mathbb{C}[J(\lambda_k, r_{kn_k})],$$

hence decomposing the $\mathbb{C}[x]$-module defined by A, which in general will not be cyclic, into cyclic $\mathbb{C}[x]$-modules defined by the above Jordan block matrices.

Proof. Consider the $\mathbb{C}[x]$-module structure defined on $\mathbb{C}^n$ by A. We have that

$$c_A(x) = d_1(x)d_2(x)\cdots d_k(x)$$

and by the Fundamental Theorem of Algebra the elementary divisors are all of the form $(x-\lambda)^r$ for some r and λ an eigenvalue. So the alternative statement of the Structure Theorem gives us

$$\mathbb{C}^n \cong \frac{\mathbb{C}[x]}{\langle (x-\lambda_1)^{r_{11}} \rangle} \oplus \frac{\mathbb{C}[x]}{\langle (x-\lambda_1)^{r_{12}} \rangle} \oplus \cdots \oplus \frac{\mathbb{C}[x]}{\langle (x-\lambda_1)^{r_{1n_1}} \rangle} \oplus \cdots \oplus \frac{\mathbb{C}[x]}{\langle (x-\lambda_1)^{r_{k1}} \rangle} \oplus \cdots \oplus \frac{\mathbb{C}[x]}{\langle (x-\lambda_1)^{r_{kn_k}} \rangle}.$$

Note that is a basis for $\mathbb{C}[x]/\langle (x-\lambda)^r \rangle$ as a vector space, and with respect to this basis we see that multiplication by x is represented by the matrix $J(\lambda, r)$. To see this note that

$$x(x-\lambda)^s = \lambda(x-\lambda)^s + (x-\lambda)^{s+1} \qquad \text{for } 0 \leqslant s < r-1$$

and that $x(x-\lambda)^{r-1} = \lambda(x-\lambda)^{r-1}$ as $(x-\lambda)^r = 0$. Hence with respect to the union of these bases multiplication by A is represented by the above matrix of Jordan blocks.

We know that the elementary divisors are unique (up to units) and so the Jordan normal form is unique. However we can further appreciate that the number of $J(\lambda, r)$ blocks equals $\text{nullity}(A - \lambda I)$, a basis for $\ker(A - \lambda I)$ consisting of the last basis vector associated with each such block. A basis for $\ker(A-\lambda I)^2$ consists of those vectors just described and the penultimate basis vectors of those $J(\lambda, r)$ blocks for which $r \geqslant 2$. etc. ■

Remark 223 *If we instead worked with the (ordered) basis $(x-\lambda)^{r-1}, \ldots, (x-\lambda)^2, (x-\lambda), 1$, then the Jordan blocks would have the form*

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & 0 & \lambda & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 0 & \lambda \end{pmatrix}$$

which is a form just as commonly used.

Example 224 *Find the RCFs and JNFs of the following complex matrices.*

$$X = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Solution. Note that $m_X(x) = c_X(x) = x^3$. Hence the rational canonical form is $C(x^3)$. The geometric multiplicity of 0 is 1 and so we have just one Jordan block $J(0,3)$. So the RCF and JNF of X are both

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note that $m_Y(x) = c_Y(x) = x^3 - 1$. This has three distinct complex roots 1, ω and ω^2 where $\omega = \text{cis}(2\pi/3)$. So the RCF equals $C(x^3-1)$ and the JNF equals $\text{diag}(J(1,1), J(1,\omega), J(1,\omega^2))$. Explicitly these are

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}.$$

Finally note that C is already in JNF equalling $\text{diag}(J(2,1), J(2,2))$. We have $m_Z(x) = (x-2)^2$ and $c_Z(x) = (x-2)^3$ so that the invariant factors are $(x-2), (x-2)^2$ and the RCF equals $\text{diag}(C(x-2), C((x-2)^2))$. Explicitly the RCF and JNF are

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -4 \\ 0 & 1 & 4 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

■

Example 225 *(a) Find the RCF and JNF of*

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

by finding the minimal and characteristic polynomials.

(b) What are the decompositions of the module defined by T that are associated with the RCF and JNF?

(c) Rederive the invariant factors by finding the Smith normal form of $xI - T$

Solution. (a) We have $c_T(x) = x^3(x-4)$ and $m_T(x) = x(x-4)$. So the invariant factors are

$$x, \qquad x, \qquad x(x-4).$$

Hence the RCF and JNF are respectively

$$\text{diag}(C(x), C(x), C(x(x-4)) \qquad \text{and} \qquad \text{diag}(J(0,1), J(0,1), J(0,1), J(4,1)).$$

Explicitly these are

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

(b) Two bases that the above matrix representatives for T are with respect to are

$$\{(1,-1,0,0)^T,(0,1,-1,0)^T,(0,0,1,0)^T,(1,1,1,1)^T\};$$
$$\{(1,-1,0,0)^T,(0,1,-1,0)^T,(0,0,1,-1)^T,(1,1,1,1)^T\}.$$

These correspond respectively to the following decompositions of $\mathbb{C}^4$ into submodules as

$$\begin{aligned}\mathbb{C}^4 &= \langle(1,-1,0,0)^T\rangle \oplus \langle(0,1,-1,0)^T\rangle \oplus \langle(0,0,1,0)^T,(1,1,1,1)^T\rangle \\ &= \langle(1,-1,0,0)^T\rangle \oplus \langle(0,1,-1,0)^T\rangle \oplus \langle(0,0,1,0)^T\rangle \oplus \langle(1,1,1,1)^T\rangle.\end{aligned}$$

or we might write these as

$$\mathbb{C}^4 \cong \frac{\mathbb{C}[x]}{\langle x\rangle} \oplus \frac{\mathbb{C}[x]}{\langle x\rangle} \oplus \frac{\mathbb{C}[x]}{\langle x(x-4)\rangle} \cong \frac{\mathbb{C}[x]}{\langle x\rangle} \oplus \frac{\mathbb{C}[x]}{\langle x\rangle} \oplus \frac{\mathbb{C}[x]}{\langle x\rangle} \oplus \frac{\mathbb{C}[x]}{\langle x-4\rangle}$$

or as further alternatives as

$$\mathbb{C}^4 \cong \mathbb{C}\left[0_1\right] \oplus \mathbb{C}\left[0_1\right] \oplus \mathbb{C}\left[\begin{pmatrix} 0 & 0 \\ 1 & 4 \end{pmatrix}\right] \cong \mathbb{C}\left[0_1\right] \oplus \mathbb{C}\left[0_1\right] \oplus \mathbb{C}\left[0_1\right] \oplus \mathbb{C}\left[4I_1\right].$$

(c) If we put $xI - T$ into Smith normal form then we find

$$\begin{pmatrix} x-1 & -1 & -1 & -1 \\ -1 & x-1 & -1 & -1 \\ -1 & -1 & x-1 & -1 \\ -1 & -1 & -1 & x-1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1-x \\ -1 & x-1 & -1 & -1 \\ -1 & -1 & x-1 & -1 \\ x-1 & -1 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1-x \\ 0 & x & 0 & -x \\ 0 & 0 & x & -x \\ 0 & -x & -x & x^2-2x \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x & 0 & -x \\ 0 & 0 & x & -x \\ 0 & -x & -x & x^2-2x \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x & 0 & -x \\ 0 & 0 & x & -x \\ 0 & 0 & 0 & x^2-4x \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x^2-4x \end{pmatrix}.$$

■

Example 226 *Let $V = \mathbb{C}^3$ and*

$$T = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Find the RCF and JNF of T.

Solution. In Example 207 we put $xI - T$ into Smith normal form and found the invariant factors to be $x-1$ and x^2-1. So the RCF and JNF are therefore respectively

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

■