Algebra I: Chapter 7
A Brief Introduction to Theory of Rings

7.1 Rings, Homomorphisms and Ideals.

A **ring** is a set R equipped with two operations $(+)$ and $(\cdot)$ having the following properties

(i) $(R,+)$ is an abelian group whose additive identity is the *zero element* $0 = 0_{_R}$ in the ring.

(ii) The multiplication operation $(\cdot)$ is associative: $x\cdot(y\cdot z) = (x\cdot y)\cdot z$

(iii) Distributive laws hold from both sides: $a\cdot(b+c) = a\cdot b + a\cdot c$ and $(a+b)\cdot c = a\cdot c + b\cdot c$

We do not assume multiplication is commutative, though it will be in the majority of our examples; if $ab = ba$ for all a, b then R is a **commutative ring**, as in Chapter 2. A ring has an **identity element** if there is some element $1 \in R$ such that

$$1\cdot a = a\cdot 1 = a \qquad \text{for all } a \in R$$

(a "two-sided" identity element); the identity is unique if it exists, and rings with identities are called **unital** rings. Two rings are **isomorphic**, indicated by writing $R \cong R'$, if there is a bijection $\phi : (R,+,\cdot) \to (R',\oplus,\odot)$ that intertwines the ring operations

$$\phi(a+b) = \phi(a)\oplus\phi(b) \qquad \text{and} \qquad \phi(a\cdot b) = \phi(a)\odot\phi(b)$$

It follows easily that if $R \cong R'$ and R has an identity so does R', with $1_{_{R'}} = \phi(1_{_R})$.

7.1.1 Definition. *In a commutative ring R with identity $1_{_R}$ the* **multiplicative units** *are the elements with multiplicative inverses, so that $x^{-1}x = 1_{_R}$. These elements form an abelian group $(\mathrm{U}_R,\cdot)$ that always includes $\pm 1_{_R}$. It is possible that these are the only units in R, as in $\mathbb{Z}_2$ and $\mathbb{Z}_4$. An element $x \in R$ is a* **prime** *if $x \neq 0$ and x cannot be factored $x = ab$ as the product of two* NON-UNITS *(a nontrivial factorization). The identity element $1_{_R}$ cannot be a prime, nor can any multiplicative unit because $a\cdot b = 1_{_R} \Rightarrow b = a^{-1}$ and a, b are both units.*

Two special types of rings will be of recurring interest. A commutative ring is:

- *An* **integral domain** *if it has an identity and no "zero divisors," which means that products of nonzero elements are always nonzero. Thus b must be zero if $ab = 0$ and $a \neq 0$.*
- *An integral domain R is a* **field** *if every nonzero element has a multiplicative inverse x^{-1} so that $x\cdot x^{-1} = 1_{_R}$.*

Division can be performed in any field if we let $a/b = a\cdot b^{-1}$ for $a, b \in R$ with $b \neq 0$. The ring of integers $(\mathbb{Z},+,\cdot)$ is a commutative ring with identity in which there is no division process, but in $\mathbb{Z}$ and other integral domains we may nevertheless perform "cancellation,"

(1) *If $a \neq 0$ and $ab = ac$ then $b = c$.*

in place of division. This works because

$$ab = ac \;\Leftrightarrow\; a\cdot(b-c) = 0 \;\Rightarrow\; b - c = 0 \;\Leftrightarrow\; b = c$$

if $a \neq 0$.

Examples of number fields include $\mathbb{C}, \mathbb{R}, \mathbb{Q}$ and the spaces $(\mathbb{Z}_p, +, \cdot)$ for prime $p > 1$. The rings $(\mathbb{Z}_n, +,)$ of (mod n) congruence classes have identities $1_{_R} = [1]$ but are not integral domains unless n is a prime. For instance $[2]\cdot[2] = [4] = [0]$ in $\mathbb{Z}_4$, and in general any non-prime $n > 1$ can be factored as $n = k\cdot\ell$ with $1 < k, \ell < n$; then $\mathbb{Z}_n$ has zero divisors because $[k], [\ell] \neq [0]$ but $[k]\cdot[\ell] = [n] = [0]$ in $\mathbb{Z}_n$.

More Examples.

- Zero Ring. The trivial ring, or "zero ring" is an amusing counterexample to many theorems unless it is explicitly excluded. It consists of a single element "0" such that

$$0 + 0 = 0 \qquad 0 \cdot 0 = 0$$

  This ring *has* an identity, namely $1_{_R} = 0_{_R}$ and is the only ring with identity such that $1_{_R} = 0_{_R}$.

- Another amusing example involves a nontrivial set R all of whose operations are trivial: $R = \{0, a\}$ with

$$0 + 0 = 0 \qquad a + a = 0 \qquad a\cdot a = 0 \qquad 0\cdot 0 = 0$$

  and $0 + a = a + 0 = a$ (as required by the commutative ring axioms, Chapter 2). This is commutative but has no identity element because $a\cdot x = 0 \neq a$ for all x. It is *not* isomorphic to the two-element ring $(\mathbb{Z}_2, +, \cdot)$.

- Boolean Algebra. Let S be a nonempty set and R = (all subsets $A \subseteq S$), equipped with the algebraic operations

$$\begin{aligned} A + B &= (A \sim B) \cup (B \sim A) \qquad (\textit{symmetric difference set}) \\ A\cdot B &= A \cap B \end{aligned}$$

  Then $(R, +, \cdot)$ is a commutative unital ring with identity $1_{_R} = S$, zero element $0_{_R} = \emptyset$, and $-A = S \sim A$. Note that every element is an *idempotent*, with $A^2 = A$. There are many zero divisors since $A\cdot B = 0_{_R} \Leftrightarrow A \cap B = \emptyset$.

- Matrix Rings. The set of all $n \times n$ matrices $\mathrm{M}(n, R)$ with entries in an integral domain R is a noncommutative ring with identity $I_{n\times n}$. There are many zero divisors, for instance

$$A^2 = 0 \text{ for } A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \left(\text{but } B^2 = B \neq 0 \text{ for } B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)$$

  Units in this ring, elements with two-sided multiplicative inverses, form a noncommutative group, the "*general linear group*"

$$\mathrm{GL} = \{A \in \mathrm{M}(n, R) : BA = AB = I \text{ for some } B\}$$

  under matrix multiply. If $R = \mathbb{F}$ (a field) the units in $\mathrm{M}(n, \mathbb{F})$ are the invertible matrices $\mathrm{GL}(n, \mathbb{F}) = \{A : \det(A) \neq 0\}$. The situation is more complicated if R is just an integral domain. For instance the units in $\mathrm{M}(n, \mathbb{Z})$ are the integral matrices with determinant ± 1; integral matrices with nonzero determinant have inverses in $\mathrm{M}(n, \mathbb{R})$, but by Cramer's Rule for computing matrix inverses the entries in A^{-1} won't be integers unless $\det = \pm 1$.

- POLYNOMIAL RINGS $\mathbb{F}[x]$. If $\mathbb{F}$ is a field the space $\mathbb{F}[x]$ of polynomials in one indeterminate consists of *finite* sums

$$f(x) = \sum_{k\geq 0} c_k x^k \qquad (c_k \in \mathbb{F}),$$

This is a commutative and unital ring under the usual $(+)$ and $(\cdot)$ operations on polynomials. The *zero element* is the polynomial with all $c_k = 0$ and the multiplicative identity $\mathbb{1}$ has $c_0 = 1, c_k = 0$ for $k > 0$. Every nonzero polynomial has a *degree*

$$\deg\, f = m \text{ if } x^m \text{ is the highest power with nonzero coefficient}$$

Constant polynomials $c\mathbb{1}$ have degree zero, except for the zero polynomial $(c = 0)$ whose degree canot be defined. The following important property

$$\deg(f\cdot h) = \deg(f) + \deg(h) \qquad \text{for } f, h \neq 0 \text{ in } \mathbb{F}[x] \tag{2}$$

follows because the leading nonzero term of $f\cdot h$ is $a_m b_n x^{m+n}$ if $f = \sum_{k=0}^{m} a_k x^k$ and $h = \sum_{k=0}^{n} b_k x^k$ with $a_m, b_n \neq 0$. Obviously $\mathbb{F}[x]$ has no zero divisors and is an integral domain. Its group of units is the set of nonzero constant polynomials $\mathrm{U}_{\mathbb{F}[x]} = \{c\mathbb{1} : c \neq 0 \text{ in } \mathbb{F}\}$. The primes in this ring are the **irreducible polynomials**, those that cannot be written as a product $f = h_1 \cdot h_2$ whose factors have $\deg(h_i) < \deg(f)$. Nonzero constant polynomials are units and cannot be prime.

7.1.2 Exercise. If $\mathbb{F}$ is a field and $f \in \mathbb{F}[x]$, explain why $f(x)$ cannot have roots in $\mathbb{F}$ (values where $f(\lambda) = 0$) if f is a prime in $\mathbb{F}[x]$. Give an example in $\mathbb{R}[x]$ showing that the converse is false: produce an $f \in \mathbb{R}[x]$ that has no real roots but is not prime.
Hint: If $f(\lambda) = 0$ you can long divide by $(x - \lambda)$ without remainder. $\square$

- POLYNOMIAL RINGS $R[x]$. These properties remain true if we allow polynomials to have coefficients chosen from an arbitrary integral domain R, not necessarily a field, defining $f + h$ in the obvious way and taking

$$\Big(\sum_{k=0}^{m} a_k x^k\Big)\cdot\Big(\sum_{k=0}^{n} b_k x^k\Big) \;=\; \sum_{k=0}^{m+n}\Big(\sum_{i+j=k} a_i b_j\Big)x^k$$

for the product. The identity element in $R[x]$ is the constant polynomial $\mathbb{1}$ whose only nonzero coefficient is $a_0 = 1_R$ (identity element in the coefficient ring R). The degree formula (2) remains valid because $a_m, b_n \neq 0 \Rightarrow a_m b_n \neq 0$ in the leading term of $f\cdot h$, hence $R[x]$ is also an integral domain.

One example in which R is not a field is the ring $\mathbb{Z}[x]$ of polynomials with integer coefficients. If f, h have leading nonzero terms $a_m x^m$ and $b_n x^n$ the leading term in their product is $a_m b_n \cdot x^{m+n}$, with $a_m b_n \neq 0$, so $fh \neq 0$. Thus $\mathbb{Z}[x]$ is also an integral domain.

7.1.3 Exercise. If R is an integral domain verify that the group of units in $R[x]$ is the set of constant polynomials $\mathrm{U}_{R[x]} = \{c\cdot\mathbb{1} : c \text{ a unit in } R\} = \mathrm{U}_R\cdot\mathbb{1}$. $\square$

- POLYNOMIAL RINGS $R[\mathbf{x}] = R[x_1, \ldots, x_n]$. If R is a field or integral domain the ring $R[\mathbf{x}]$ of polynomials in several indeterminates $x_1, \ldots, x_n$ consists of finite sums with coefficients in R

$$f(\mathbf{x}) = \sum_{k_1\geq 0}\cdots\sum_{k_n\geq 0} c_{(k_1,\ldots,k_n)}\, x_1^{k_1}\ldots x_n^{k_n}$$

Efficient discussion of these rings requires use of "**multi-index notation**." A multi-index is an ordered n-tuple of nonnegative integers $\alpha = (\alpha_1, \ldots, \alpha_n)$ in $\mathbb{Z}_+^n$ for which we define a *degree* $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and *sums*

$$\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$$

Each multi-index corresponds to a "monomial" in $R[x_1, \ldots, x_n]$

$$x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n} \qquad (\text{with } x_k^{\alpha_k} = 1 \text{ if } \alpha_k = 0)$$

We assign a degree $\deg(x^\alpha) = |\alpha|$ to each monomial. By convention $x^{(0,\ldots,0)} = 1\!\!1$, the constant polynomial whose only nonzero coefficient is $c_{(0,\ldots,0)} = 1_R$. Every polynomial in $R[\mathbf{x}]$ is a finite linear combination $f = \sum_{\alpha \in \mathbb{Z}_+^n} c_\alpha x^\alpha$ with coefficients $c_\alpha \in R$.

Sums $f + h$ are defined by adding coefficients of like monomials; products are formed by multiplying monomials appearing in f and h, assuming that the indeterminates commute $x_i x_j = x_j x_i$ so that

$$c_\alpha x^\alpha \cdot d_\beta x^\beta = c_\alpha d_\beta \cdot x^{\alpha+\beta} \qquad (\text{which has degree } |\alpha + \beta| = |\alpha| + |\beta|),$$

and then adding all terms involving the same monomial. In multi-index notation the resulting product formula is

$$\Big(\sum_{|\alpha| \geq 0} c_\alpha x^\alpha \Big) \cdot \Big(\sum_{|\beta| \geq 0} d_\alpha x^\beta \Big) = \sum_{|\gamma| \geq 0} \Big(\sum_{\alpha+\beta=\gamma} c_\alpha \cdot d_\beta \Big) \cdot x^\gamma \tag{3}$$

This will look reasonably familiar if your write it out when there are just $n = 2$ variables x, y.

Every nonzero polynomial in $\mathbb{F}[x_1, \ldots, x_n]$ has a well-defined **degree**

$$\deg(f) = \max\{|\alpha| : c_\alpha \neq 0\}$$

with $\deg(f) = 0 \Leftrightarrow f$ is a nonzero constant polynomial; no meaningful degree can be assigned to the zero polynomial $0 \cdot 1\!\!1$. Because a nonzero f might have several terms of highest degree $c_\alpha x^\alpha$ with $|\alpha| = m$, the proof of the degree formula for polynomials of several variables is much more challenging than that for polynomials in one variable. The desired result is

(4) THEOREM (DEGREE FORMULA). *If R is an integral domain and $f, h \in R[\mathbf{x}]$ are nonzero then $f \cdot h$ is nonzero and* $\deg(f \cdot h) = \deg(f) + \deg(h)$.

For $n \geq 2$ this is a tricky result involving "lexicographic ordering" of the monomials x^α; we won't go into the proof here. Once established it implies that $R[\mathbf{x}]$ is an integral domain whose units are the particular constant polynomials

$$\mathrm{U}_{R[x_1,\ldots,x_n]} = \mathrm{U}_R \cdot 1\!\!1$$

When $n \geq 2$ the set of primes in $R[x_1, \ldots, x_n]$ is not so easy to identify even if the coefficient ring is a field.

7.1.4 Exercise. Which of the following polynomials (if any)

(a) $x^2 - y^2$ (c) $x^2 + xy + y^2$

(b) $x^2 + y^2$ (d) $x^2 - y^2 + 1$

are primes in the polynomial ring $\mathbb{R}[x,y]$? In $\mathbb{C}[x,y]$? $\square$

Here is an instructive example of a polynomial ring with exotic coefficients.

7.1.5 Example. If $\mathbb{F}$ is a field and we take $R = \mathbb{F}[x]$ as the coefficients in the polynomial ring $R[y]$, every element in $\mathbb{F}[x,y]$ can be uniquely be rewritten as an element of $R[y]$:

$$f(x,y) = \sum_{i,j\geq 0} c_{ij}x^i y^j = \sum_{j\geq 0} \Big(\sum_{i\geq 0} c_{ij}x^i \Big) \cdot y^j = \sum_{j\geq 0} p_j(x) y^j$$

with $p_j \in R[x]$. Letting $\Phi(f) \in R[y]$ be the right hand sum, it is easy to verify that $\Phi : \mathbb{F}[x,y] \to R[y]$ is a bijection and an isomorphism of rings, so that

$$\mathbb{F}[x,y] \cong \big(\mathbb{F}[x]\big)[y] \quad \square$$

7.1.6 Example (Adjoining Roots to $\mathbb{Q}$). The set of $\mathbb{Q}$-linear combinations

$$\mathbb{E} = \mathbb{Q} + \sqrt{2}\cdot\mathbb{Q} = \{a + \sqrt{2}b \in \mathbb{R} : a,b \in \mathbb{Q}\} = \mathbb{Q}\text{-span}\{1, \sqrt{2}\}$$

is a commutative unital ring if we define

$$\begin{aligned} (a+\sqrt{2}b) + (a'+\sqrt{2}b') &= (a+a') + \sqrt{2}(b+b') \\ (a+\sqrt{2}b)\cdot(a'+\sqrt{2}b') &= (aa'+2bb') + \sqrt{2}(ab'+a'b) \end{aligned}$$

with $1_{_R} = 1+\sqrt{2}\cdot 0$ and $0_{_R} = 0+\sqrt{2}\cdot 0$. It contains a copy of the rationals $\mathbb{Q} \cong \mathbb{Q}+\sqrt{2}\cdot 0$ and hence may be regarded as an "extension" of $\mathbb{Q}$ to a larger system $\mathbb{E}$. This extension is actually a field, for if $z = a+\sqrt{2}b \neq 0+\sqrt{2}\cdot 0$, then

$$\frac{1}{z} = \frac{1}{a+\sqrt{2}b}\cdot\frac{a-\sqrt{2}b}{a-\sqrt{2}b} = \Big(\frac{a}{a^2-2b^2}\Big) + \sqrt{2}\Big(\frac{-b}{a^2-2b^2}\Big)$$

is its multiplicative inverse within $\mathbb{E}$. (The denominator a^2-2b^2 is nonzero because there is no $\sqrt{2}$ in $\mathbb{Q}$.)

Since $\mathbb{Q} \subseteq \mathbb{E}$ we may regard $\mathbb{Q}[x]$ as the subset of polynomials in $\mathbb{E}[x]$ that happen to have coefficients in $\mathbb{Q}$. The polynomial $f = x^2-2 \in \mathbb{Q}[x]$ has no roots in $\mathbb{Q}$ and is irreducible, so it is a prime in the ring $\mathbb{Q}[x]$. But it does have roots $\pm\sqrt{2}$ in the larger field $\mathbb{E}$, and splits into linear factors $x^2-2 = (x-\sqrt{2})\cdot(x+\sqrt{2})$ in $\mathbb{E}[x]$. Essentially, $\mathbb{E}$ was obtained by "adjoining a root of x^2-2" to the original field $\mathbb{Q}$. In a natural sense $\mathbb{E}$ is the smallest field containing $\sqrt{2}$ and $\mathbb{Q}$. $\square$

7.1.7 Example (The Gaussian Integers $\mathbb{Z}[\mathbf{i}]$). The "integral points" in the system of complex numbers

$$\mathbb{Z}[i] = \{m+in \in \mathbb{C} : m,n \in \mathbb{Z}\} \tag{5}$$

form a commutative unital ring with $1_{_R} = 1+i0$ because this system is closed under the usual operations $(+)$ and $(\cdot)$ in $\mathbb{C}$. It is an integral domain but not a field. In fact there are just four units in this system,

$$\mathrm{U}_R = \{1, -1, i, -i\}$$

because if $z\cdot w = 1$ in $\mathbb{C}$ and z,w are both integral, the identity $1 = |zw| = |z|\cdot|w|$ can only hold when both z and w have absolute value 1, which means $z = \pm 1$ or $\pm i$, and likewise for w.

We will soon see that the map $d : \mathbb{Z}[i] \to \mathbb{Z}_+$

$$d(m+in) = m^2+n^2 = |m+in|^2$$

shares many properties with the degree map on polynomial rings $\mathbb{F}[x]$, except that

$$d(zw) = d(z)\cdot d(w) \qquad \text{instead of} \qquad d(f\cdot h) = d(f) + d(h)$$

The ring $\mathbb{Z}[i]$ and its degree map are important in number theory and in later discussion of the "prime factorization problem" for integral domains. $\square$

Subrings. Consider a ring R that is not necessarily commutative. A **subring** is a subset R' such that

$$0_{_R} \in R' \qquad R' + R' \subseteq R' \qquad R' \cdot R' \subseteq R' \qquad -R' = R'$$

Then $(R', +, \cdot)$ is a ring in its own right with $0_{_{R'}} = 0_{_R}$, in which the additive inverse $-a$ of an element in R' agrees with the additive inverse in R. As examples we have (i) $R' = 2\mathbb{Z}$ or $R' = n\mathbb{Z}$ in $R = \mathbb{Z}$; (ii) the set $R' = x\cdot R[x]$ of polynomials in $R[x]$ without constant term; (iii) the set R' of strictly upper triangular matrices in $\mathrm{M}(n, \mathbb{F})$, those with zeros on and below the diagonal; (iv) $D =$ all diagonal matrices in $\mathrm{M}(n, \mathbb{F})$; and (v) $D_0 =$ the diagonal matrices with lower right entry $a_{nn} = 0$.

Identity elements in subrings require careful handling. R' need not have an identity even if R does [as in (ii), (iii), (v)], and R' can have its own identity even if R does not. Moreover, even if there are identities $1_{_R} \in R$ and $1_{_{R'}} \in R'$, these might not agree. [The identity in D is the identity matrix $I = I_{n\times n}$ while the identity in the subring D_0 is $E = \mathrm{diag}(1, \ldots, 1, 0)$, with $E^2 = E \neq I$ in the larger ring D.]

Now shift attention to commutative rings to keep things simple. A nonempty subset of a commutative ring R generates a subring $R' = \langle S\rangle$ just as a subset of a group generates a subgroup $H = \langle S\rangle$

(6) GENERATED SUBRING: *The subring $\langle S\rangle$ generated by a nonempty subset of a ring R is the smallest subring $R' \subseteq R$ that contains S.*

This "top down" definition makes sense because the intersection of an arbitrary family of subrings is again a subring, so $\langle S\rangle$ is the intersection of all subrings containing S (there is at least one, namely R itself). But, as with groups, there is a more informative "bottom up" construction as sums of "words" in the generators:

$$\langle S\rangle = \left\{ \sum_{i=1}^{r} s_1 \cdot \ldots \cdot s_r : r < \infty, s_i \in S \cup (-S) \right\} \tag{7}$$

7.1.8 Exercise. Verify that the set of elements specified in (7) actually is the minimal subring containing the generators S. Why must we allow $s_i \in -S$ in forming the "words" whose finite sums make up $\langle S\rangle$? Construct a simple example illustrating what goes wrong if you only allow $s_i \in S$. $\square$

7.1.9 Exercise. If $R = \mathbb{Z}$ and $S = \{15, 18\}$ what is $\langle S\rangle$?

(a) If $R = \mathbb{F}[x]$ what is the subring generated by $\{x^2, x^3\}$? By $\{x^4, x^6\}$?

(b) Show that the subring generated by $f = 1 + x^2$ consists of the polynomials in $(1 + x^2)\cdot \mathbb{F}[x^2]$, where $\mathbb{F}[x^2]$ is the subring of *even* polynomials – those in which only even powers x^{2k} apppear.

Note: In (b) you should verify that $(1 + x^2)\mathbb{F}[x^2]$ actually is a subring in $\mathbb{F}[x]$.

Now assume there are only finitely many generators, so $S = \{s_1, \ldots, s_n\}$. If R is assumed commutative every word $s_1 \ldots s_r$ in (7) may be rewritten as a "reduced word" $\pm s_1^{k_1} \ldots s_n^{k_n}$ (with $k_i \geq 0$ and $k_1 + \ldots + k_n = r$) by gathering together all occurrences of

the same generator s_i. The sum in (7) can be further simplified by bringing together all reduced words that differ only by a $\pm$ sign; using multi-index notation we get

$$\langle S\rangle = \sum_{\alpha\in\mathbb{Z}_+^n} m_\alpha \cdot s_1^{\alpha_1}\dots s_n^{\alpha_n} = \sum_{\alpha\in\mathbb{Z}_+^n} m_\alpha s^\alpha \qquad (m_\alpha\in\mathbb{Z}) \tag{8}$$

Thus $\langle S\rangle$ consists of finite sums of monomials s^α in the (commuting) generators $s_1,\dots,s_n$ with coefficients $m_\alpha\in\mathbb{Z}$. Note the resemblance between (8) and elements of the polynomial ring $\mathbb{Z}[x_1,\dots,x_n]$, which are described in multi-index notation as the finite sums

$$f = \sum_{\alpha\in\mathbb{Z}_+^n} m_\alpha x^\alpha = \sum_{\alpha\in\mathbb{Z}_+^n} m_\alpha \cdot x_1^{\alpha_1}\cdot\ldots\cdot x_n^{\alpha_n} \qquad (m_\alpha\in\mathbb{Z})$$

This similarity is no accident. The generated ring $\langle S\rangle$ is obtained by substituting $x_1 = s_1,\dots,x_n = s_n$ in arbitrary polynomials $f\in\mathbb{Z}[x_1,\dots,x_n]$. In this description there is no need to worry about the role of $-S$ because the integer coefficients m_α in (8) can be positive or negative.

If R is a commutative ring with identity and S a nonempty subset, the subring it generates $R' = \langle S\rangle$ might not contain the identity element $1_{_R}$. The subring $R'' = \langle 1_{_R}, S\rangle$ contains the identity and is just large enough to pick up $1_{_R}$ and all the elements in S.

7.1.10 Exercise. If R is a commutative ring with identity $1_{_R}$ show that

(a) The subring $\langle 1_{_R}\rangle$ generated by $1_{_R}$ is the set of "integer multiples" $\mathbb{Z}\cdot 1_{_R}$.

(b) If S is a nonempty subset of R show that

$$\langle S, 1_{_R}\rangle = \mathbb{Z}\cdot 1_{_R} + \langle S\rangle = \{a+b : a\in\mathbb{Z}\cdot 1_{_R}, b\in\langle S\rangle\} \quad \square$$

Here the action $\mathbb{Z}\times R\to R$ of the natural integers on elements of a ring must be defined with some care. For $x\in R$ and $n\in\mathbb{Z}$ we define

$$n\cdot x = \begin{cases} x+\ldots+x & (n \text{ times}) \quad \text{if } n>0 \text{ in } \mathbb{Z} \\ (-x)+\ldots+(-x) & (|n| \text{ times}) \quad \text{if } n<0 \text{ in } \mathbb{Z} \end{cases} \tag{9}$$

taking $0\cdot x = 0_{_R}$ when $n=0$.

BEWARE: The outcome is not always what you might expect. For instance if $R=\mathbb{Z}_n$ we get $n\cdot 1_{_R} = [1]+\ldots+[1]$ (n terms) $=[0]$ in $\mathbb{Z}_n$, so $\mathbb{Z}\cdot 1_{_R} = \mathbb{Z}\cdot[1] = \mathbb{Z}_n$ is finite! The subring $\mathbb{Z}\cdot 1_{_R}$ can be finite even if the over-ring R is infinite, as in the next example.

7.1.11 Example. If p is a prime the polynomial ring $\mathbb{Z}_p[x]$ is an infinite dimensional vector space with basis vectors $\mathfrak{1}, x, x^2, \dots$ and coefficients in the *finite* field $\mathbb{Z}_p$. The set of constant polynomials $c\cdot\mathfrak{1}$ ($c\in\mathbb{F}=\mathbb{Z}_p$) is finite and equal to

$$\mathbb{Z}_p\cdot\mathfrak{1} = \{[k]\cdot\mathfrak{1} : [k]\in\mathbb{Z}_p\}.$$

By the distributive laws this is the same as the set $\mathbb{Z}\cdot\mathfrak{1}$ of finite sums

$$\mathfrak{1}, \quad \mathfrak{1}+\mathfrak{1} = [2]\cdot\mathfrak{1}, \quad \mathfrak{1}+\mathfrak{1}+\mathfrak{1} = [3]\cdot\mathfrak{1}, \quad \ldots$$

If $R=\mathbb{Z}_p[x]$ the subring $\langle\mathfrak{1}\rangle = \mathbb{Z}\cdot\mathfrak{1}$ generated by the identity element reduces to $\mathbb{Z}_p\cdot\mathfrak{1}$. $\square$

Quotient Rings and Ideals. Again we allow noncommutative rings for a while, before focusing on the commutative rings that are our main interest. A **homomorphism** $\phi : R\to R'$ between rings is a map such that

$$\phi(a+b) = \phi(a)\oplus\phi(b) \qquad \text{and} \qquad \phi(a\cdot b) = \phi(a)\odot\phi(b) \qquad \text{for all } a,b\in R$$

Here the operations in R' are being used on the right. An **isomorphism** is a bijective homomorphism between rings; its inverse $\phi^{-1} : R' \to R$ is automatically a homomorphism too. We write $R \cong R'$ if the rings are isomorphic, and this RST equivalence relation partitions the family of all rings into disjoint "isomorphism classes."

Things go differently for rings than they do for groups. Even if R, R' have identities there is no guarantee that $\phi(1_{_R}) = 1_{_{R'}}$ unless this property is specified separately. Homomorphisms between unital rings are called **unital homomorphisms** if $\phi(1_{_R}) = 1_{_{R'}}$.

7.1.12 Exercise. If $\phi : R \to R'$ is a homomorphism of rings prove that

$$\text{(a)}\ \ \phi(0_{_R}) = 0_{_{R'}} \qquad\qquad \text{(b)}\ \ \phi(-a) = -\phi(a) \text{ in } R'$$

7.1.13 Exercise. If $\phi : R \to R'$ is a *bijective* homomorphism between rings, prove that the inverse map $\phi^{-1} : R' \to R$ is automatically a homomorphism of rings. □

7.1.14 Exercise. Invent a homomorphism from $\mathrm{M}(n, \mathbb{F}) \to \mathrm{M}(n+1, \mathbb{F})$ such that $\phi(I_n) \neq I_{n+1}$. □

7.1.15 Exercise. If $\phi : R \to R'$ is a *surjective* homomorphism and R has a multiplicative identy $1_{_R}$, explain why R' also has an identity, and ϕ is a *unital* homomorphism. □

Note that a homomorphism $\phi : R \to R'$ is one-to-one if and only if its $\ker(\phi)$ is trivial because

$$\phi(a) = \phi(b) \ \Leftrightarrow\ \phi(b-a) = 0_{_{R'}} \ \Leftrightarrow\ b - a \in K(\phi)\ \ ,$$

so $b = a$ if $\ker(\phi) = \{0_{_R}\}$.

The **zero homomorphism** $\phi : R \to R'$ kills everything: $\phi(a) = 0_{_{R'}}$ for all a. This example also shows that $\mathrm{range}(\phi)$ can be the trivial ring, and that $\phi(1_{_R})$ need not be equal to $1_{_{R'}}$. The **kernel** $K(\phi) = \{x \in R : \phi(x) = 0_{_{R'}}\}$ of a homomorphism $\phi : R \to R'$ is a subring of R with the special property

$$a \cdot K(\phi) \subseteq K(\phi) \quad \text{ and } \quad K(\phi) \cdot a \subseteq K(\phi) \quad \text{ for all } a \in R,$$

which follows because $\phi(x) = 0 \Rightarrow \phi(ax) = \phi(a)\phi(x) = 0_{_{R'}}$, and similarly if we multiply by a on the right. The **range** of ϕ, the image of R in R', is a subring of R' without any special properties.

As with groups, it is useful to abstract the properties of kernels of homomorphisms that distinguish them from mere subrings. But in rings there are complications since rings have two operations $(+)$ and $(\cdot)$, and R need not be commutative.

7.1.16 Definition. *Let R be a nontrivial ring. A (left/right/two-sided)* **ideal** *is a subset $I \subseteq R$ such that*

(i) *I is a subring (possibly trivial, or all of R).*

(ii) *For all $a \in R$ we have, respectively, $aI \subseteq I$; or $Ia \subseteq I$; or $aI \subseteq I$ and $Ia \subseteq I$.*

An ideal is **proper** *if $I \neq (0)$ and $I \neq R$. Subsets of the form $a + I = \{a + x : x \in I\}$ are the* **additive cosets** *of the ideal. The* **quotient space** *R/I is the set of* ADDITIVE *cosets $x + I$, $x \in R$.*

Of course if R is commutative there is no distinction between left-, right-, or two-sided ideals, and in that case we simply speak of "ideals."

7.1.17 Exercise. If I is a two-sided ideal in a ring R and $a \in R$, show that

(a) $a + I = I \Leftrightarrow a \in I$

(b) $I + I = \{a + b : a, b \in I\}$ is equal to I.

(c) $a + I = b + I \Leftrightarrow b - a \in I$.

(d) If $a, b \in R$ either $a + I = b + I$ or these two cosets are disjoint. □

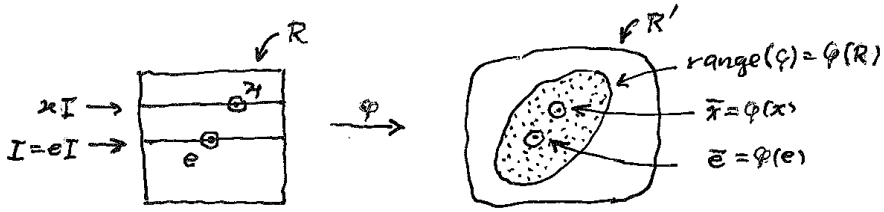

Figure 7.1. A homomorphism $\phi : R \to R'$ sends each coset $x + I$ ($I = \ker\phi$), to a single point in the target space R', and distinct cosets have distinct images. The range of ϕ need not be all of R', as shown here.

Thus the additive cosets of a two-sided ideal partition R into disjoint subsets. The **quotient space** R/I consists of the distinct additive cosets of I.

The preceding remarks show that kernels $K(\phi)$ of homomorphisms $\phi : R \to R'$ between rings are two-sided ideals. We will soon see that all two sided ideals arise this way. The action of a typical homomorphism $\phi : R \to R'$ has the behavior shown in Figure 7.1.

7.1.18 Lemma. *If $\phi : R \to R'$ is a homomorphism between rings and $I = \ker\phi$, then*

(i) *ϕ is constant on cosets: each coset $a + I$ "collapses" to a single point in R'.*

(ii) *ϕ maps distinct cosets $a + I \neq b + I$ to different points in R'.*

(iii) *The map ϕ is one-to-one $\Leftrightarrow$ the kernel $I = \ker\phi = \{0_{_R}\}$.*

PROOF: In (i) we have $\phi(a + I) = \phi(a) + \phi(I) = \phi(a)$; furthermore

$$\begin{aligned} \phi(a) &= \phi(b) \Leftrightarrow \phi(a-b) = 0_{_R} \Leftrightarrow a - b \in I \\ &\Leftrightarrow a \in b + I \Leftrightarrow a + I = b + I \quad \text{(by 7.1.17(b))} \end{aligned}$$

proving (ii). Obviously (iii) follows from (i) + (ii). Remember: $a + I \neq b + I \Leftrightarrow$ they are disjoint. □

The Quotient Ring R/I. If N is a normal subgroup in a group G the space G/N of left cosets xN inherits a natural group structure, with $(xN) \cdot (yN) = xyN$. For a two-sided ideal I in a ring R the operations $(+)$ and $(\cdot)$ in R pass down to operations that make the quotient space R/I a ring in its own right.

7.1.19 Definition. *Let R be a nontrivial ring, I a two-sided ideal, and let $\pi : R \to R/I$ be the* **quotient map** *sending $a \in R$ to $\pi(a) = a + I \in R/I$. Then the operations*

$$(a + I) \oplus (b + I) = (a + b) + I \qquad (a + I) \odot (b + I) = ab + I \tag{10}$$

are well-defined and make R/I into a ring. The quotient map becomes a surjective ring homomorphism; if R has an identity element then $\phi(1_{_R}) = 1_{_R} + I$ is an identity element for R/I.

PROOF: All choices of coset representatives a, b yield the same outcome in (10). In fact $a' + I = a + I \Rightarrow a' = a + k$ for some $k \in I$ and similarly $b' = b + \ell$, with $\ell \in I$, so

$$\begin{aligned} (a' + b') + I &= (a + b) + \big((k + \ell) + I\big) = (a + b) + I \\ (a'b') + I &= ab + \big((a\ell + kb + k\ell) + I\big) = ab + I \end{aligned}$$

By definition π is surjective and the ring axioms are easily verified via calculations involving representatives, once the operations are known to make sense. If R is unital $\overline{1} = \pi(1_R) = 1_R + I$ is obviously a two-sided identity for the quotient ring and $\pi : R \to R/I$ is a unital homomorphism. $\square$

This also shows that every two-sided ideal I in a ring R is the kernel of some ring homomorphism, namely the quotient map $\pi : R \to R/I$.

7.1.20 Example. (Ideals in $\mathbb{Z}$). If $m > 1$ in $\mathbb{Z}$ then $I = (m) = m\mathbb{Z} = \{mk : k \in \mathbb{Z}\}$ is a (two-sided) ideal in $\mathbb{Z}$. The quotient ring $R/I = \mathbb{Z}/(m)$ is the familiar space of (mod m) congruence classes in $\mathbb{Z}$, and the operations in R/I coincide with the operations defined earlier on congruence classes in $\mathbb{Z}_m$. Thus R/I and $\mathbb{Z}_m$ are *isomorphic* rings under the bijection $f(k + (m)) = [k]_m = k + m\mathbb{Z}$

Excluding the ideals $I = (0)$ and $I = \mathbb{Z}$, all *proper* ideals have the form $I = m\mathbb{Z}$ for some $m > 1$. In fact, $I \cap \mathbb{N}$ must contain a smallest element m and if $m = 1$ the ideal coincides with $\mathbb{Z}$. When $m > 1$ the ideal contains $m\mathbb{Z}$, but if I contained any other element $n_0 \notin m\mathbb{Z}$ we could adjust this by adding a multiple of m to get an element in I such that $0 < n_0 < m$. That would contradict minimality of $m = \min(I \cap \mathbb{N})$. $\square$

7.1.21 Example (A Substitution Principle). Let $R \neq (0)$ be a *commutative* unital ring without zero divsiors (an integral domain) and consider a polynomial $f(x) \in R[x]$ with coefficients in R. If we fix some $a \in R$ we can substitute $x = a$ in $f(x)$ to get an **evaluation map** $\epsilon_a : R[x] \to R$

$$\epsilon_a\Big(\sum_{i=0} c_i x^i\Big) = \Big(\sum_{i=0} c_i a^i\Big) \in R$$

It is trivial to verify that ϵ_a is a homomorphism, surjective since $\epsilon_a(c\mathbb{1}) = c$ for $c \in R$; it is unital because $\epsilon_a(\mathbb{1}) = \mathbb{1}_R$. Computing the kernel $\ker(\epsilon_a) = \{f \in R[x] : f(a) = 0_R\}$ explicitly can be a challenge if R is a peculiar ring, but is easy when $R = \mathbb{F}$ is a field. $\square$

7.1.22 Lemma. If $\mathbb{F}$ is a field and $f \in \mathbb{F}[x]$ a nonzero polynomial such that $f(a) = 0$ for some $a \in \mathbb{F}$, then $(x - a)$ divides f without remainder:

$$\text{There is some } h \in \mathbb{F}[x] \text{ such that } f = (x - a)\cdot h(x).$$

Thus $\ker(\epsilon_a) = (x - a)\mathbb{F}[x]$, the ideal in $\mathbb{F}[x]$ generated by $(x - a)$.

PROOF: Since f is nonzero in $\mathbb{F}[x]$ but $f(a) = 0$, f cannot be a constant polynomial, so $\deg(f) \geq 1$ and

$$f = c_m x^m + \ldots + c_1 x + c_0 \qquad \text{with } c_m \neq 0, m \geq 1$$

If $m = 1$ then $f = c_1 x + c_0$ and since $f(a) = 0$ we have $c_0 = -c_1 a$, so $f = c_1(x - a)$ and $(x-a)$ divides $f(x)$. Arguing inductively, assume $(x-a)$ divides $h(x)$ for all $h(x)$ of degree $\leq m$ such that $h(a) = 0$. If $\deg(f) = m+1$ and $f(a) = 0$ with $f(x) = c_{m+1}x^{m+1} + \ldots + c_0$, let $g(x) = c_{m+1}x^m(x - a) = c_{m+1}x^{m+1} - ac_{m+1}x^m$, chosen to have the same leading term as $f(x)$. Subtract to get $r(x) = f(x) - g(x)$, which has lower degree. Then

$$f(x) = g(x) + r(x) = h(x)(x - a) + r(x) \qquad \text{where } h(x) = c_{m+1}x^m$$

Here the remainder is either zero (and $(x-a)$ divides $f(x)$), or is nonzero with degree $\leq m$. In this case $r(a) = f(a) - h(a)(a - a) = 0$, and by the induction hypothesis there is an $h'(x) \in \mathbb{F}[x]$ such that $r(x) = h'(x)(x - a)$. Then

$$f(x) = \big[h(x) - h'(x)\big](x - a)$$

is also divisible by $(x - a)$ as required. $\square$

7.1.23 Corollary. *If $\mathbb{F}$ is a field any nonconstant polynomial $f \in \mathbb{F}[x]$ has at most d distinct roots, where $d = \deg(f)$.*

For every root a we can split off a factor $(x-a)$, lowering deg f by one. Repeating this process we can write $f(x) = (x-a)^m h(x)$ where a is not a root of $h(x)$; the exponent m is the **algebraic multiplicity** of a. Extending 7.1.23, if each root in $\mathbb{F}$ of $f(x)$ is counted according to its multiplicity $m_i = m(a_i)$ we see that

$$f(x) = h(x) \cdot \prod_{i=1}^{r} (x-a_i)^{m_i} \qquad \text{where } h(x) \text{ has no roots in } \mathbb{F} \tag{11}$$

Even if each root is counted according to its multiplicity the number of roots cannot exceed deg f,

$$\deg f \geq m(a_1) + \ldots + m(a_r) \qquad a_1, \ldots a_r \text{ the distinct roots of } f$$

Of course it is possible that $f(x)$ has *no* roots in $\mathbb{F}$, as when $\mathbb{F} = \mathbb{R}$ and $f = x^2 + 1$ or $\mathbb{F} = \mathbb{Q}$ and $f = x^2 - 2$ (since there is no $\sqrt{2}$ in the rationals).

From here on we will focus on *commutative* rings R, and some of the special ideals they contain.

7.1.24 Definition. (Principal Ideals). *Let R be a nontrivial commutative ring and $a \in R$. This element determines two ideals in R, the ideal $I = \langle\langle a \rangle\rangle$* **generated by** *$a$, and the* **principal ideal** *$I = (a)$ determined by a.*

PRINCIPAL IDEAL: *The subset* $(a) = aR = \{ar : r \in R\}$

GENERATED IDEAL: *The smallest ideal I containing a,* (12)

$$\langle\langle a \rangle\rangle = \mathbb{Z} \cdot a + aR = \{na + ar : n \in \mathbb{Z}, r \in R\}$$

We do not assume that there is an identity in R. If not, the principal ideal (a) is contained in, but could differ from, the ideal $I = \langle\langle a \rangle\rangle$ generated by a, and the principal ideal (a) might not contain the element by which it is determined! If R has an identity 1_R then $a = a \cdot 1_R \in aR = (a)$ and in this case $(a) = \langle\langle a \rangle\rangle$. Below we will show that all ideals in the polynomial ring $\mathbb{F}[x]$ are principal ideals when $\mathbb{F}$ is a field, a fact that is no longer true for polynomials $\mathbb{F}[x_1, \ldots, x_n]$ in more than one unknown.

7.1.25 Exercise. In the ring $\mathbb{F}[x, y]$ prove that the ideal $I = \{f(x, y) : f(0, 0) = 0\}$ is *not* a principal ideal – i.e. $I \neq (h) = h(x, y) \cdot \mathbb{F}[x, y]$ for any polynomial $h(x, y)$.
Hint: The polynomials $p(x, y) = x$ and $q(x, y) = y$ are in I, and the degree formula (4)

$$\deg(f \cdot g) = \deg(f) + \deg(g)$$

holds for any $f, g \neq 0$ in $\mathbb{F}[x, y]$. What does this tell you about a possible generator $h(x, y)$? □

7.1.26 Exercise. If a commutative ring $R \neq (0)$ does not have an identity show that

(a) The set $(a) = aR$ is an ideal in R.

(b) The set $\langle\langle a \rangle\rangle = \mathbb{Z}a + aR = \{na + ar : n \in \mathbb{Z}, r \in R\}$ is an ideal.

(c) Explain why $\langle\langle a \rangle\rangle$ is the *smallest* ideal in R containing a. □

7.1.27 Example. In the unital ring $\mathbb{F}[x]$ the sets $x \cdot \mathbb{F}[x]$ = (polynomials without constant term) and $(1 + x^2) \cdot \mathbb{F}[x]$ are examples of principal ideals, generated by the elements $g = x$ and $h = 1 + x^2$. Principal ideals in $\mathbb{F}[x]$ can arise in other ways, for example if $a \in \mathbb{F}$

the kernel of the evaluation map $\epsilon_a : f \to f(a) \in \mathbb{F}$ is a principal ideal – and in fact by 7.1.22 we have

$$\ker(\epsilon_a) = \{f : f(a) = 0\} = (x-a)\cdot\mathbb{F}[x] = \langle\langle x-a \rangle\rangle$$

so $\ker(\epsilon_a)$ is precisely the principal ideal $I = (x-a)$ consisting of polynomials divisible by $(x-a)$. $\square$

Polynomials $f = \sum_{k\geq 0} c_k x^k$ (finite formal sums of powers x^k) are easily confused with the scalar valued functions $\psi_f : \mathbb{F} \to \mathbb{F}$ they determine, namely

$$\psi_f(a) = \sum_{k=0} c_k a^k \qquad \text{(for all } a \in \mathbb{F})$$

When $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, $\mathbb{F}[x]$ and $\mathcal{P}_\mathbb{F}$ are essentially the same thing, but as the next example shows they are very different animals when $\mathbb{F}$ is a finite field such as $\mathbb{Z}_p$ ($p > 1$ prime).

7.1.28 Example (Formal Sums vs. Polynomial Functions). Let $\mathbb{F}$ be a field and $\mathcal{P}_\mathbb{F}$ the set of "polynomial maps" $\psi : \mathbb{F} \to \mathbb{F}$, so

$$\psi \in \mathcal{P}_\mathbb{F} \;\Leftrightarrow\; \exists\, f \in \mathbb{F}[x] \text{ such that } \psi(t) = \psi_f(t) = \sum_{k=0} c_k t^k \text{ for all } t \in \mathbb{F}$$

$\mathcal{P}_\mathbb{F}$ is a commutative unital ring of functions on $\mathbb{F}$ under the usual pointwise operations $(+)$ and $(\cdot)$; the identity element in $\mathcal{P}_\mathbb{F}$ is the constant function everywhere equal to $1_\mathbb{F}$. Every polynomial $f \in \mathbb{F}[x]$ yields a polynomial map $\psi_f : \mathbb{F} \to \mathbb{F}$ under the natural correspondence

$$\Psi : \mathbb{F}[x] \to \mathcal{P}_\mathbb{F} \qquad \text{such that} \qquad \Psi(f) = \psi_f(t) \text{ for all } t \in \mathbb{F}$$

The map Ψ is surjective by definition of $\mathcal{P}_\mathbb{F}$, and is easily seen to be a unital homorphism between rings.

What is often not recognized is that Ψ might not be a ring isomorphism because it *need not be one-to-one*. Different polynomial expressions $f = \sum_{i=0} c_i x^i$ may collapse to the same scalar-valued function $\psi_f : \mathbb{F} \to \mathbb{F}$, or equivalently a nonzero polynomial $f(x)$ might yield the zero function $\psi_f(t) \equiv 0_\mathbb{F}$ on $\mathbb{F}$.

First note that $\ker(\Psi) = \{f \in \mathbb{F}[x] : f(t) \equiv 0 \text{ on } \mathbb{F}\}$, so Ψ *is* bijective if the field has infinitely many elements. [Then $f \in \ker(\Psi) \Rightarrow f(t)$ would be zero for infinitely many $t \in \mathbb{F}$, which by 7.1.23 is impossible unless f is the zero polynomial.] Thus $\ker(\Psi)$ is trivial and $\mathbb{F}[x] \cong \mathcal{P}_\mathbb{F}$ if $|\mathbb{F}| = \infty$.

Now consider the polynomial $x^p - x$ in the ring $\mathbb{Z}_p[x]$ of polynomials with coefficients in the finite field $\mathbb{Z}_p$ (p a prime). We invoke the following well-known result from Number Theory.

7.1.29 Theorem (Fermat's Little Theorem). *If $p > 1$ is a prime then $t^p = t$ for all $t \in \mathbb{Z}_p$, so the polynomial $f = x^p - x \in \mathbb{Z}_p[x]$ is in the kernel of the homomorphism $\Psi : \mathbb{Z}_p[x] \to \mathcal{P}_\mathbb{F}$ when $\mathbb{F} = \mathbb{Z}_p$.*

PROOF: The equation $x^p - x = 0$ is satisfied if we set $x = 0$. Since p is a prime $\mathbb{Z}_p$ is a field and every nonzero element is a multiplicative unit. These units form a multiplicative group with $p-1$ elements. By the Lagrange Theorem we must have $u^{p-1} = 1$ for every unit, so every $u \neq 0$ in $\mathbb{Z}_p$ is a root of the polynomial $x^{p-1} - 1$. Thus $f(x) = x^p - x$ is zero for every element in $\mathbb{Z}_p$, making $\psi_f \equiv 0$. $\square$

When $\mathbb{F} = \mathbb{Z}_p$ the kernel

$$\ker(\Psi) = \{f \in \mathbb{Z}_p[x] : f(t) \equiv 0 \text{ on } \mathbb{Z}_p\}$$

obviously contains the principal ideal $I = (x^p - x) \cdot \mathbb{Z}_p[x]$. Later on with additional tools it will be easy to show that $\ker(\Psi)$ is actually equal to $(x^p - x) \cdot \mathbb{Z}_p[x]$. Can you devise a proof now, using brute force?

7.2. General Properties of Quotient Ring Construction.

We begin with further comments valid for arbitrary, not necessarily commutative rings.

7.2.1 Theorem (First Isomorphism Theorem). *Let R be a nonzero ring and ϕ a homomorphism from R to R' with $I = \ker(\phi)$ and let $A = \phi(R)$ be the image of R in R'. If $\pi : R \to R/I$ is the quotient homomorphism there is an induced isomorphism of rings $\tilde{\phi} : R/I \to A$ given by*

$$\tilde{\phi}(x + I) = \phi(x) \text{ for all } x \in R \tag{13}$$

The map (13) *is well-defined, the diagram in Figure 7.2 commutes (with $\phi = \tilde{\phi} \circ \pi$), and $\phi : R/I \to A$ is a ring isomorphism. If ϕ is surjective then $R' \cong R/I$.*

PROOF: The definition $\tilde{\phi}(x + I) = \phi(x)$ makes sense: if $I = \ker(\phi)$ then ϕ is constant

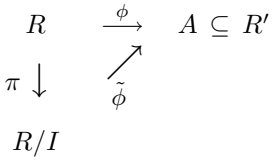

Figure 7.2. The situation in proving the First Isomorphism Theorem. Here $\phi : R \to R'$ is a homomorphism and $A = \phi(R)$ its range, a subring in R'; π is the quotient map. Note that (i) ϕ is constant on cosets $x + I$ in R, and (ii) $\ker(\phi) = \ker(\pi)$.

on each coset $x + I$. It is a ring homomorphism because

$$\begin{aligned}
\phi\big[(x+I) \oplus (y+I)\big] &= \phi\big((x+y)+I\big) = \phi(x+y) \\
&= \phi(x) \oplus \phi(y) = \phi(x+I) \oplus \phi(y+I) \\
\phi\big[(x+I) \odot (y+I)\big] &= \phi\big((xy)+I\big) = \phi(xy) \\
&= \phi(x) \odot \phi(y) = \phi(x+I) \odot \phi(y+I)
\end{aligned}$$

The diagram commutes (with $\phi = \tilde{\phi} \circ \pi$) by definition of $\tilde{\phi}$.

Map $\tilde{\phi}$ is surjective because $a \in A \Leftrightarrow a = \phi(x)$ for some $x \in R$ and then $\tilde{\phi}(x + I) = \phi(x) = a$. The map is one-to-one $\Leftrightarrow \ker(\tilde{\phi})$ is trivial, but $\tilde{\phi}(x + I) = \phi(x) = 0_{R'} \Rightarrow x + I = 0_R + I = I$ (the zero element in R/I), so $\tilde{\phi}$ is one-to-one. □

Sometimes we want to show that two rings resulting from different quotient constructions are actually isomorphic. That can be hard, but if A, B come from the same "mother ring" R by surjective homomorphisms ϕ, ψ as shown below then $A \cong B$ if the maps have the same kernel in R.

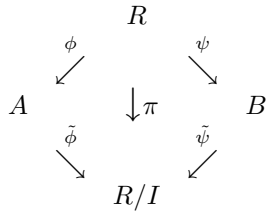

Figure 7.3. The situation in Corollary 7.2.2. Here $\ker(\phi) = I = \ker(\psi)$ for *surjective* homomorphisms $\phi : R \to A$ and $\psi : R \to B$.

7.2.2 Corollary. *If ϕ, ψ are surjective homomorphisms from a ring R to other rings A, B and if their kernels agree $\ker(\phi) = \ker(\psi)$, then $A \cong B$ as rings.*

PROOF: Let $I \subseteq R$ be their common kernel and $\pi : R \to R/I$ the quotient homomorphism. Applying the First Isomorphism Theorem twice we get $A \cong R/I \cong B$. $\square$

Ideals in R vs. Ideals in R/I. If I is a two-sided ideal in an arbitrary ring R there is a natural bijective correspondence

$$\left(\begin{array}{c}\text{two-sided ideals } \overline{J} \\ \text{in quotient ring } R/I\end{array}\right) \longleftrightarrow \left(\begin{array}{c}\text{two-sided ideals } J \subseteq R \\ \text{that contain } I\end{array}\right)$$

The following facts are easily verified.

7.2.3 Exercise. If R is a nonzero ring, I a two-sided ideal, and $\pi : R \to R/I$ is the quotient homorphism, prove that

(a) If J is a two-sided ideal in R such that $J \supseteq I$ then

$$\overline{J} = \pi(J) = J/I\,\big(\{u + I; u \in J\}\big)$$

is a two-sided ideal in R/I.

(b) The correspondence $J \mapsto \overline{J}$ is one-to-one.

(c) If $\overline{J}$ is a two-sided ideal in $\overline{R} = R/I$, its pullback

$$J = \pi^{-1}(\overline{J}) = \{x \in R : \pi(x) \in \overline{J}\}$$

is a two-sided ideal in R such that $J \supseteq I$ and $\pi(J) = \overline{J}$.

(d) The correspondence $J \to \overline{J}$ is a *bijection* $\square$

7.2.4 Example. In the polynomial ring $\mathbb{F}[x]$ let $R = x\mathbb{F}[x]$, the subring of polynomials whose constant term is zero. The polynomial $f(x) = x^2$ is in R, which contains the ideals

$$(f) = f \cdot R \qquad \text{and} \qquad \langle\langle f \rangle\rangle = f \cdot R + \mathbb{Z} \cdot f$$

There is no identity element in R because $\deg(fh) \geq 1 + 1 = 2$ if $f, h \neq 0$ in R. In this example $\langle\langle f \rangle\rangle$ contains but is not equal to the principal ideal (f) because

$$(f) = x^2 \cdot R = x^2\big(x\mathbb{F}[x]\big) = x^3 \cdot \mathbb{F}[x]$$

(polynomials with a zero at the origin of third order or higher), while

$$\langle\langle f \rangle\rangle = x^3\mathbb{F}[x] + \mathbb{Z} \cdot x^2$$

contains element of degree 2. $\square$

In 7.1.27 we showed that the evaluation homomorphisms $\epsilon_a : \mathbb{F}[x] \to \mathbb{F}$ had kernels that were principal ideals

$$\ker(\epsilon_a) = (x - a) \cdot \mathbb{F}[x]$$

We now show that *all* ideals in a polynomial ring $\mathbb{F}[x]$ are singly-generated principal ideals, and that their generators are easily identified. This result has many important generalizations.

7.2.5 Theorem. *If $\mathbb{F}$ is a field and I an ideal in $\mathbb{F}[x]$, then $I = f(x) \cdot \mathbb{F}[x]$ for some $f \in \mathbb{F}[x]$. Furthermore, the generator f is unique up to a nonzero scalar multiple:*

(14) *Elements $f, f' \in \mathbb{F}[x]$ generate the ideal $I \Leftrightarrow$ there is a constant $c \neq 0$ such that $f' = cf(x)$.*

Finally, if I is not the zero ideal it is equal to $g(x)\cdot\mathbb{F}[x]$ for any nonzero element $g(x)$ of lowest degree in I.

PROOF: If $I = (0)$ take $f = 0$. Otherwise $m = \min\{\deg(f) : f \neq 0 \text{ in } I\}$ is a well-defined integer $m \geq 0$ and I contains an element f_0 of this degree.

CASE 1: $\deg(f_0) = 0$. Then $f_0 = c\mathbb{1}$ with $c \neq 0$ and $I \supseteq c\mathbb{1} \cdot \mathbb{F}[x]$ is all of $\mathbb{F}[x]$.

CASE 2: $\deg(f_0) \geq 1$. Then $\deg h \geq \deg f_0$ if $h \neq 0$ in I and we may "long divide" by $f_0(x)$ to get

$$h(x) = f_0(x)q(x) + r(x) \qquad \text{where} \begin{cases} r = 0 \text{ (the zero polynomial), or} \\ 0 \leq \deg r < \deg f_0 \end{cases}$$

But then $r = h - f_0 q$ is in I and has lower degree than f_0, which is impossible unless $r = 0$, in which case we have $h = f_0 q$ and $h \in (f_0)$. Hence $I \subseteq (f_0)$. The reverse inclusion $(f_0) \subseteq I$ is obvious because $f_0 \in I \Rightarrow (f_0) = f_0 \cdot \mathbb{F}[x] \subseteq I$. Therefore every nontrivial ideal in $\mathbb{F}[x]$ is singly generated, by any nonzero element $f_0 \in I$ of minimal degree.

As for uniqueness, the multiplicative units $\mathrm{U}_{\mathbb{F}[x]}$ in $\mathbb{F}[x]$ are precisely the nonzero constants $\{c\mathbb{1} : c \neq 0 \text{ in } \mathbb{F}\}$, so if $(f') = (f) = I$ there exist g_1, g_2 such that $f' = g_1 f$ and $f = g_2 f'$, so $f' = g_1 g_2 \cdot f'$. Since $\mathbb{F}[x]$ has no zero divisors we must have $g_1 g_2 = \mathbb{1}$ so g_1 and g_2 are both units with $g_k = c_k \mathbb{1}$ for suitably chosen constants, proving essential uniqueness of the generators of I. □

All this is summarized by saying: If $\mathbb{F}$ is a field then $\mathbb{F}[x]$ is a **principal ideal domain** – all ideals are singly generated. As shown in Exercise 7.1.25, no such result holds for ideals in the ring $\mathbb{F}[x_1, \ldots, x_n]$ if $n \geq 2$. Nor does it hold in polynomial rings $R[x]$ if the coefficient ring R is not a field – for example $R = \mathbb{Z}[x]$.

7.3 Unique Factorization in Rings.

Here we discuss *Euclidean domains* which embody an abstracted version of the "division with remainder" process in $\mathbb{Z}$ and polynomial rings $\mathbb{F}[x]$, which lay at the heart of our discussion of unique prime factorization for integers $n > 1$ in Chapter 2 of these *Notes*. Most of that discussion remains valid, with little change, in the much larger realm of *Euclidean domains*.

7.3.1 Definition. *Let R be a unital commutative ring without zero divisors (an integral domain). It is a* **Euclidean domain** *if there exists a "degree map" $d : R^\times \to \mathbb{Z}_+$ defined on the nonzero elements in $R^\times \subseteq R$ such that*

(15)
- (i) *$a, b \neq 0$ in $R \Rightarrow d(ab) \geq \max\{d(a), d(b)\}$*
- (ii) *For all $a, b \neq 0$ there exist $r, q \in R$ such that*

$$a = bq + r \qquad \text{with} \qquad \begin{cases} r = 0, \text{ or} \\ r \neq 0 \text{ and } d(r) < d(b) \end{cases}$$

The value $d = 0$ is allowed, but we might have $d(a) > 0$ for all $a \neq 0$.

We do not assume there is an identity element in R, but this turns out to be a consequence of the definition. The elements q, r are not unique: in $R = \mathbb{Z}$ we may write $5 = 3\cdot 2 - 1 = 1\cdot 3 + 2$, and we can't avoid this ambiguity by requiring "$r \geq 0$" because R need not be an ordered ring.

7.3.2 Exercise. Verify that the identity $d(ab) \geq \max\{d(a), d(b)\}$ holds if

(a) $d(ab) = d(a) + d(b)$ or if

(b) $d(ab) = d(a)\cdot d(b)$ and $d(a) \geq 1$ for all $a \neq 0$. □

7.3.3 Lemma. *If R is a Euclidean domain with identity element 1 then*

(i) $d(1) = \min\{d(x) : x \neq 0\}$.

(ii) *For $x \neq 0$, $d(x) = d(1) = \min\{d(x) : x \neq 0\} \Leftrightarrow x$ is a unit in R.*

(iii) *If x is a unit in R and $x \neq 0$ then $d(ux) = d(x)$.*

PROOF: We have $d(x) \geq d(1)$ by (15) because

$$d(x) = d(x \cdot 1) \geq \max\{\, d(a), d(1)\,\} \geq d(1) \qquad \text{for all } x \neq 0$$

That proves (i). For (ii), if $d(x) = d(1)$, we may write $1 = x \cdot q + r$ as in (15). If $r = 0$ then x and q are both units in R; otherwise, $r = 1 - x \cdot q$ would be a nonzero element in R with $d(r) < d(x) = d(1)$ which is impossible by (i). Conversely if u is a unit in R then $uu^{-1} = 1$ and $d(1) \geq d(u)$; but by (i) we have $d(u) \geq d(1)$, hence $d(u) = d(1)$. For (iii), we have

$$d(x) = d(u^{-1}ux) \geq d(ux) \geq d(x) \quad \square$$

Incidentally, the converse of (iii) fails to be true: $d(x) = d(y)$ does *not* imply $y = ux$ for some unit u.

As examples of Euclidean domains we have

- Polynomial rings $\mathbb{F}[x]$ are Euclidean domains taking $d(f) = \deg(f)$ for $f \neq 0$, and the usual algorithm for long division with remainder $f(x) = q(x) \cdot g(x) + r(x)$ for $g \neq 0$ in $\mathbb{F}[x]$.

- $\mathbb{Z}[x]$ is *not* a Euclidean domain despite the fact that $\mathbb{Z}$ is an integral domain. Problems arise in trying to define a degree map $d : \mathbb{Z}[x]^\times \to \mathbb{Z}_+$ and a corresponding division process such that

  $$f = q \cdot g + r \qquad \text{with remainder } r = 0 \text{ or } d(r) < d(g)$$

  for $g \neq 0$ in $\mathbb{Z}[x]$. No such process (or degree map) can be defined for elements $f(x) = x$ and $g(x) = 3x$ because the coefficient ring $\mathbb{Z}$ does not contain an element "$\frac{1}{3}$."

- The integers $\mathbb{Z}$ become a Euclidean domain if we take $d(m) = |m|$.

The next example played an important role in the development of Number Theory.

7.3.4 Example (The Gaussian Integers $\mathbb{Z}[\mathbf{i}]$). The "integral points"

$$\mathbb{Z}[i] = \{m + in : m, n \in \mathbb{Z}\}$$

in the complex plane form a ring under the complex $(+)$ and $(\cdot)$ operations, with identity element $1 = 1 + i0$. We get a Euclidean domain if we define the degree map to be

$$d(z) = |z|^2 = m^2 + n^2 \quad \text{for} \quad z = m + in \in \mathbb{Z}[i] \tag{16}$$

Obviously $d(z)$ is an integer ≥ 1 if $z \neq 0$ and $d(z \cdot w) = d(z) \cdot d(w) \geq \max\{d(z), d(w)\}$. Defining the appropriate division process is a more subtle problem. If $z, w \neq 0$ then $w/z \in \mathbb{C}$ and we may pick an integral point $\left[\frac{w}{z}\right]$ with minimal distance to $\frac{w}{z}$. Generally this "nearest neighbor" in $\mathbb{Z}[i]$ is unique but there could be as many as 4 such points. In any case,

$$\text{The real and imaginary parts of } \frac{w}{z} - \left[\frac{w}{z}\right] \text{ lie in the interval } \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]$$

so that

$$\left| \frac{w}{z} - \left[\frac{w}{z}\right] \right|^2 \leq \left(\tfrac{1}{2}\right)^2 + \left(\tfrac{1}{2}\right)^2 = \tfrac{1}{2}$$

Then we have

$$\begin{aligned} w &= \left(\frac{w}{z}\right)z = \left[\frac{w}{z}\right] \cdot z + \left(\frac{w}{z} - \left[\frac{w}{z}\right]\right) \cdot z \\ &= \left[\frac{w}{z}\right] \cdot z + \left(w - \left[\frac{w}{z}\right] \cdot z\right) = qz + r \end{aligned}$$

where $q \in \mathbb{Z}[i]$ and the remainder r satisfies

$$|r|^2 \leq \left| \frac{w}{z} - \left[\frac{w}{z}\right] \right|^2 \cdot |z|^2 \leq \tfrac{1}{2}|z|^2 < |z|^2$$

as required. Here $r \in \mathbb{Z}[i]$ because $r = w - \left[\frac{w}{z}\right]z$ is in $\mathbb{Z}[i]$. $\square$

7.3.5 Exercise. For the following $z, w \in \mathbb{Z}[i]$ find all possible ways to write $w = qz + r$ with $q \in \mathbb{Z}[i]$ and $r = 0$ or $|r|^2 < |z|^2$.

(a) $z = 2 + i0$, $w = 8 + i9$ (b) $z = 3 + i5$, $w = 8 + i9$ $\square$

7.3.6 Exercise. Show that the multiplicative units in $\mathbb{Z}[i]$ consist of the four points $\{1, +i, -1, -i\}$. We say that $z \neq 0$ is a *prime* in $\mathbb{Z}[i]$ if it is (i) not a unit, and (ii) cannot be factored as a product $z = w_1 \cdot w_2$ of *non-units*.

(a) Is $3 + i4$ a prime in $\mathbb{Z}[i]$?

(b) If $p > 1$ is a prime in the system of integers $\mathbb{Z}$, is $p + i0$ always a prime in the Gaussian integers $\mathbb{Z}[i]$?

(c) The dihedral group D_4 generated by

$$R_{90^\circ} = (90 \text{ degree rotation}) \qquad \text{and} \qquad r_x = (\text{reflection across } x\text{-axis})$$

contains the $\mathbb{R}$-linear operator

$$T(x, y) = (y, x) = R_{90^\circ} \circ r_x \qquad (\text{reflection across the } 45^\circ \text{ line in } \mathbb{R}^2)$$

If we identify $\mathbb{C}$ with the plane $\mathbb{R}^2$, so $T(a + ib) = (b + ia)$, and if $z = a + ib$ is nonzero in $\mathbb{Z}[i]$, prove that

$$a + ib \text{ is a prime in } \mathbb{Z}[i] \;\Leftrightarrow\; T(a + ib) \text{ is a prime}$$

(d) Prove that the reflection r_x across the x-axis also sends primes to primes in $\mathbb{Z}[i]$.

Hints: If $z \in \mathbb{Z}[i]$ and u is a unit show that z is a prime $\Leftrightarrow$ uz is a prime. $\square$

7.3.7 Exercise. Use the observations in Exercise 7.3.6 to prove that all linear operators in the dihedral group D_4 permute the primes in $\mathbb{Z}[i]$.
Hint: Show that the elements T and r_x generate the group D_4.
Note: This simplifies the search for primes in $\mathbb{Z}[i]$: any prime $m + in$ is $A(m' + in')$ for some $A \in D_4$ and some prime in the sector $0 \leq \theta \leq \pi/4$ in $\mathbb{C}$. $\square$

We saw in 7.1.20 that all ideals in $R = \mathbb{Z}$ have the form $I = m \cdot \mathbb{Z}$ for $m = 0, 1, 2, \ldots$ We now show that every ideal in a Euclidean ring R is a principal ideal (single generator).

7.3.8 Theorem. *Every Euclidean domain R is a* **principal ideal domain**: *Every ideal $I \subseteq R$ has the form*

$$I = aR = \{ax : x \in R\}$$

for some $a \in R$. Furthermore,

(i) *A Euclidean ring has an identity element, so every principal ideal* $(a) = Ra$ *is equal to* $\langle\langle a\rangle\rangle$ *and contains the generator* a.

(ii) *If* I *is any ideal in* R*, then* $I = (a)$ *for any element* $a \neq 0$ *in* I *such that* $d(a) = \min\{d(x) : x \neq 0 \text{ in } I\}$.

(iii) *We have* $aR = bR$ *for nonzero* $a, b \in I \Leftrightarrow$ *there exist units* $u, u' \in \mathrm{U}_R$ *such that* $b = ua$ *and* $a = u'b$.

In particular, $x \neq 0$ *generates* $I \Leftrightarrow x$ *has minimal degree* $d(x) = \min\{d(y) : y \neq 0 \text{ in } I\}$ *in* I.

PROOF (ii): If $I = (0)$ we may take $a = 0$. Otherwise I contains nonzero elements and $\min\{d(x) : x \neq 0 \text{ in } I\}$ is achieved. If $a \neq 0$ is any element of minimal degree then $I = Ra$, for if $b \neq 0$ in I we may write $b = qa + r$ with $r = 0$ or $d(r) < d(a)$. But $r = b - qa \in I$ since $a \in I$, and by minimality of $d(a)$ we must have $r = 0$. Thus $I = Ra$.

PROOF (iii): Obviously if $u \in \mathrm{U}_R$ we have $aR = (au)R$ since $a = (au)\cdot u^{-1} \in (au)R$. On the other hand if $aR = bR$ we can find $u, u' \in R$ such that $b = au$ and $a = bu'$, so that $b = au = b(u'u)$. By cancellation we get $u'u = 1$ so u and u' are both units.

PROOF (i): Since R itself is an ideal we have $R = aR$ for some $a \neq 0$, and then there must be an element e such that $a = ae$. This is our candidate for the identity element. In fact if $b \in R$ then $b = ax$ for some x, and then $eb = e(ax) = (ea)x = ax = b$ for all b. □

The ring $\mathbb{F}[x]$ of polynomials in one variable is a Euclidean domain, but $\mathbb{F}[x_1, \ldots, x_n]$ is not when $n \geq 2$ and its ideal structure is much more complicated. Likewise, the ring $\mathbb{Z}[x]$ of polynomials with integer coefficients is not a Euclidean domain.

7.3.9 Exercise. Prove that the ideal $I = \{f \in \mathbb{F}[x, y] : f(0, 0) = 0\}$ is *not* a principal ideal in $R = \mathbb{F}[x, y]$. Exhibit *two* elements a, b such that $I = \langle a, b\rangle = aR + bR$.
Hint: Use the degree formula $\deg(f\cdot h) = \deg(f) + \deg(h)$ for nonzero polynomials in two variables. □

7.3.10 Exercise. Prove by counterexample that $R = \mathbb{Z}[x]$ has ideals I that are not principal ideals, and hence cannot be made into a principal ideal domain, no matter how we attempt to define a degree map on R.
Hint: Try $I = \langle\langle 2, x\rangle\rangle = 2\cdot\mathbb{Z}[x] + x\mathbb{Z}[x]$. □

7.3.11 Exercise. We have seen in 7.3.3 that $d(a) = d(b)$ in a Euclidean domain if there is a unit $u \in \mathrm{U}_R$ such that $b = ua$. Provide a counterexample: a familiar Euclidean domain in which the converse

$$d(b) = d(a) \Rightarrow \exists\, u \in \mathrm{U}_R \text{ such that } b = ua$$

is not always true. □

Associated Elements in an Integral Domain. The following remarks apply to an arbitrary integral domain R as well as Euclidean domains. An element $a \neq 0$ in R is a **prime** if it cannot be factored as a product $a = bc$ of *nonunits*. If such nontrivial factorizations exist the factors are nonunique "*up to multiplied units*" – i.e. if $a = b'c'$ there is a unit $u \in \mathrm{U}_R$ such that $b' = bu$ and $c' = u^{-1}c$. This nonuniqueness gets to be annoying, which suggests that we might not want to distinguish nonzero elements $a \in R^\times = R \sim (0)$ that differ by a multiplied unit. Thus in $R = \mathbb{Z}$ we would not distinguish between $\pm f$ and in $R = \mathbb{F}[x]$ we would lump together all scalar multiples to get an equivalence class $[f] = \{c\cdot f : c \neq 0 \text{ in } \mathbb{F}\}$.

It is easy to verify that the following relation on the set $R^\times$ of nonzero elements in an integral domain

$$a \underset{R}{\sim} b \;\Leftrightarrow\; \exists\, u \in \mathrm{U}_R \text{ such that } b = ua \qquad \text{(we say "}b\text{ is an } \textbf{associate} \text{ of } a\text{")} \tag{17}$$

is an RST equivalence relation whose equivalence classes are the "multiplicative cosets" in $(R^\times, \cdot\,)$ of the group $(\mathrm{U}_R, \cdot\,)$,

$$[a] = a\cdot \mathrm{U}_R \qquad \text{for } a \neq 0 \text{ in } R \tag{18}$$

Note that all units $u \in \mathrm{U}_R$ get lumped together in the single equivalence class [1].

Multiplication in R induces a well-defined multiplication operator $(\cdot)$ on the quotient space $R^\times/\mathrm{U}_R = \{[a] : a \in R^\times\}$ of multiplicative cosets if we define

$$[a]\cdot[b] = [ab] \qquad \text{for all } a, b \neq 0 \text{ in } R$$

This makes sense independent of the coset representatives a, b because

$$a' \underset{R}{\sim} a, b' \underset{R}{\sim} b \;\Rightarrow\; a' = ua, b' = u'b \text{ for } u, u' \in \mathrm{U}_R \;\Rightarrow\; a'b' = (ab)\cdot uu'$$

so that

$$[a']\cdot[b'] = (a'b')\mathrm{U}_R = ab\big(uu'\mathrm{U}_R\big) = ab\cdot \mathrm{U}_R = [a]\cdot[b] \quad .$$

The surjective quotient map $\pi(a) = [a]$ from $(R^\times, \cdot\,) \to (R^\times/\mathrm{U}_R, \cdot\,)$ intertwines the product operations $(\cdot)$ in these two systems.

On the other hand the $(+)$ operation in R *does not* induce a well-defined operation on cosets; the system $(R^\times/\mathrm{U}_R, \cdot\,)$ is only an **abelian semigroup** – a system with an associative and commutative multiplication law. Although it contains a multiplicative identity element [1], $R^\times/\mathrm{U}_R$ cannot be made into a group because multiplicative inverses $[a]^{-1}$ don't exist; nor can it be made into a ring, because it lacks a sensible $(+)$ operation.

Nevertheless, many issues about factorization, divisibility, and primality in commutative rings are easier to manage if we consider products of equivalence classes $[a]$ in $R^\times/\mathrm{U}_R$ instead of individual ring elements $a \neq 0$. For instance we say that

> *A class* $[a]$ *is* **prime** *if* $[a]$ *cannot be written as a product* $[b]\cdot[c]$ *of nontrivial classes* $[a], [b] \neq [1]$.

In the multiplicative semigroup $R^\times/\mathrm{U}_R$ there is just one trivial factor, namely [1], while in R itself all units must be regarded as trivial factors.

Finally, because $d(u) = 1$ for any unit in R and $d(au) = d(a)$, the degree map $d : R^\times \to \mathbb{Z}_+$ is constant on classes and induces a well-defined map on the quotient space

$$\tilde{d} : R^\times/\mathrm{U}_R \to \mathbb{Z}_+ \qquad \text{with} \qquad \tilde{d}\big([a]\big) = d(a) \tag{19}$$

which has the useful property

$$\tilde{d}\big([a]\cdot[b]\big) \geq \max\{\, \tilde{d}\big([a]\big), \tilde{d}\big([b]\,\big)\} \quad .$$

7.3.12 Exercise. In an integral domain R we say that "a *divides* b," or "b is a multiple of a," indicated by writing $a|b$, if $b = ac$ for some c. In the semigroup $R^\times/\mathrm{U}_R$ we say that class $[a]$ *divides class* $[b]$ if $[b] = [a]\cdot[c]$ for some class $[c]$.

(a) If $a, b \neq 0$ in R show that $a|b \Leftrightarrow [a]$ divides $[b]$.

(b) If $a \neq 0$ in R show that a is a prime in $R \Leftrightarrow [a]$ is a prime class in $R^\times/\mathrm{U}_R$. □

Prime Factorizations in Euclidean Domains. We now show that every class $[a] \neq [1]$ in a Euclidean ring has a factorization into primes $[a] = [p_1]\cdot[p_2]\cdot\ldots\cdot[p_r]$ that is unique except for the order in which the factors appear. (Remember: by our definitions [1] is not a prime.) We start with the *existence* of such factorizations.

7.3.13 Proposition (Existence of Prime Factorizations). *In any Euclidean domain R every nonzero nonunit a can be factored as a product $a = p_1 \cdot \ldots \cdot p_r$ of primes in R. Likewise any class $[a] \neq [1]$ is a product $[p_1]\cdot[p_2]\cdot\ldots\cdot[p_r]$ of prime classes.*

PROOF: There is nothing to do if a is already a prime. Otherwise we argue by induction on degree $d(a)$, starting from $d(1) = \min\{d(x) : x \neq 0\} \geq 0$. By 7.3.3 the element a is a unit, and not a prime, if $d(a) = d(1)$ so in this case the claim is true by default. Thus we may assume $d(a) > d(1)$ and there are nonunits b, c such that $a = bc$. Then

$$d(1) < d(b),\ d(c) \leq d(bc) = d(a) \quad ,$$

by definition of the degree map.

CASE 1. If both $d(b), d(c) < d(a)$ the Inductive Hypothesis implies that b and c both have prime factorizations, and hence so does a by combining the prime factors of b, c.

CASE 2. One of the factors, say b, has $d(b) = d(a)$. [We make no claim yet about $d(c)$ except that $d(c) \leq d(a)$.] Since b divides a we have $a \in I = bR$. By 7.3.8, an element $x \neq 0$ is a generator of this principal ideal $\Leftrightarrow$

$$d(x) = \min\{d(y) : y \neq 0 \text{ in } I\}$$

Since b is a generator this minimum is equal to $d(b)$, so I is generated by any element $x \neq 0$ such that $d(x) = d(b)$.

Taking $x = a$ we get $aR = bR$, hence by 7.3.8(iii) $b = ua$ for some unit $u \in \mathrm{U}_R$. Then $b = ua = ubc = b\cdot(uc)$ and by cancellation of the nonzero factor b we conclude that $uc = 1$ and $c = u^{-1}$ is a *unit*, contradicting nontriviality of the given factorization $a = bc$. Therefore Case 2 cannot arise, Case 1 prevails, and a has a prime factorization as required to complete the induction step. □

7.3.14 Corollary. *If $a \neq 0$ in a Euclidean domain R, and $a = bc$ with b, c both nonunits, then $d(b) < d(a)$ and $d(c) < d(a)$.*

The remaining issue is uniqueness of such prime factorizations, which is closely related to the notion of *greatest common divisor* $\gcd(a, b)$ of nonzero elements a, b in R. This concept makes sense in any integral domain, not just Euclidean domains.

7.3.15 Definition. *If R is an integral domain, a* **greatest common divisor** (GCD) *of elements $a, b \neq 0$ is any nonzero element c such that* (i) *$c|a$ and $c|b$, and* (ii) *If $c'|a$ and $c'|b$ then $c'|c$. We make no claims about "positivity" of such an element, if it exists, since R need not be an ordered ring.*

If $\gcd(a, b)$ exists it is unique up to a multiplied unit $u \in \mathrm{U}_R$. [If c, c' both satisfy conditions (i)+(ii) then there exist $r, r' \in R$ such that $c' = r'c = (r'r)c'$ and since $c' \neq 0$ we may cancel to get $rr' = 1$, so both are units.] Thus if a GCD exists it determines a unique class $[c]$ in the semigroup. In this sense the class $[\gcd(a, b)]$ is unique even if its representative $\gcd(a, b) \in R$ is not.

7.3.16 Theorem (Existence of GCD). *If R is a Euclidean domain any two elements $a, b \neq 0$ have a greatest common divisor c. This element is unique up to a multiplied unit and lies in the ideal $I = Ra + Rb$. The unique class $[\gcd(a, b)]$ is determined by any nonzero element of minimal degree in I.*

PROOF: If the ideal I reduces to $(a) = Ra$ (or to $(b) = Rb$) then $b \in Ra$ and $a|b$; thus a is a divisor of both a and b, so (i) is satisfied. But if c' is any common divisor we have $c'|a$, so a serves as the $\gcd(a, b)$. In this situation we obviously have

$$d\big(\gcd(a, b)\big) = d(a) = \min\{d(x) : x \neq 0 \text{ in } I\} \quad ,$$

by 7.3.8(ii).

In general, $I \neq (0)$ and there is some nonzero $x \in I$ of minimal degree, so by 7.3.8 we have $I = xR$. Thus $x|a$ and $x|b$. For any other common divisor x' of a and b we have $x'|(ra + sb)$ for all $r, s \in R$. Since x itself is in $I = Ra + Rb$ we get $x'|x$ as required to make x a GCD for a, b. $\square$

7.3.17 Exercise. If $a, b \neq 0$ in a Euclidean ring R show that the class $[\gcd(a, b)]$ determined by any GCD depends only on the classes $[a]$ and $[b]$. $\square$

Thus we may speak of the GCD of two classes $[a], [b]$ in $R^\times/\mathrm{U}_R$.

Existence of a GCD is crucial in proving uniqueness of prime factorizations in Euclidean rings. GCDs exist in rings more general than Euclidean domains; for instance multi-variable polynomial rings $\mathbb{F}[x_1, \ldots, x_n]$ and $\mathbb{Z}[x_1, \ldots, x_n]$ are not Euclidean when $n \geq 2$. Nevertheless every such polynomial has a unique factorization into primes (non-constant irreducible polynomials). But in such rings existence of $\gcd(a, b)$ has to be proved by other methods, and need not have the simple form $d = ra + sb$ for $r, s \in R$. Though the proofs in these settings differ from those given below for Euclidean rings, the following properties of the GCD play a pivotal role in all discussions of unique factorization, as they did for $R = \mathbb{Z}$.

7.3.18 Definition. *If R is a Euclidean ring and $a, b \neq 0$, we write $a \sim b$ if they differ by a multiplied unit, with $a = ub$. They are* **relatively prime** *if the* GCD *is itself a unit, so* $\gcd(a, b) \sim 1$. *An element $a \neq 0$ is* **prime** *if it is a nonunit and*

$$a = bc \implies \text{either } b \sim 1 \text{ or } c \sim 1 \qquad (\textit{no nontrivial factorizations})$$

If $a \neq 0$ is a nonunit we have seen that it has at least one factorization into primes. Note that if a is a unit and $b \neq 0$ we have $\gcd(a, b) \sim 1$. [If d is a GCD then $d|a \Rightarrow a = dx$ for some x, and then $1 = d{\cdot}(xa^{-1}) \Rightarrow d \in \mathrm{U}_R$, so $d \sim 1$.]

7.3.19 Proposition. *Let R be a Euclidean ring, $a \neq 0$, and $p \in R^\times$ a prime. Then*

$$\textit{Either } p|a \textit{ or } p \textit{ is relatively prime to } a.$$

Proof: If $d = \gcd(p, a)$ then $d|p$, $p = xd$ for some x, and at least one of the factors is a unit. If d is a unit we have $\gcd(p, a) \sim 1$. Otherwise x is a unit and we have $p|a$ because

$$d = x^{-1}p \text{ divides } a \implies a = y{\cdot}(x^{-1}p) = (yx^{-1}){\cdot}p \Rightarrow p|a \ . \qquad \square$$

7.3.20 Exercise. If p is a prime in a Euclidean ring R show that

(a) The only divisors of p are u and up where $u \in \mathrm{U}_R$ is any unit.

(b) If p, q are primes in R then $p \sim q$ or $\gcd(p, q) \sim 1$. $\square$

7.3.21 Proposition. *Let R be a Euclidean ring, let $a, b \neq 0$, and let p be a prime. If p divides ab but does not divide a, then p must divide b.*

Proof: By 7.3.19 we must have $\gcd(p, a) \sim 1$ if p does not divide a, and then by 7.3.16 we can find $r, s \in R$ such that $1 = \gcd(p, a) = ra + sp$. It follows that

$$b = b{\cdot}1 = rab + sbp \ .$$

Since $p|ab$ we conclude that p divides b, as claimed. $\square$

7.3.22 Corollary. *If p is a prime in a Euclidean ring R and p divides a product of nonunits $a_1{\cdot}\ldots{\cdot}a_n$, then there is some index i such that $p|a_i$.*

Proof: If $p|a_1$ we are done; otherwise p divides $a_2{\cdot}\ldots{\cdot}a_n$ and we procede inductively. $\square$

7.3.23 Theorem (Uniqueness of Prime Factorization). *In a Euclidean ring R every nonunit $a \neq 0$ has an essentially unique factorization into primes. In terms of classes in $R^\times/\mathrm{U}_R$, if $[a] \neq [1]$ there are distinct prime classes $[p_1], \ldots, [p_r]$ and multiplicities $m_i \in \mathbb{N}$ such that $[a] = \prod_{i=1}^{r} [p_i]^{m_i}$. This factorization is unique except for the order in which the factors appear.*

PROOF: Prime factorizations exist by 7.3.13. If $q_1 \cdot \ldots \cdot q_s$ is another prime factorization we may assume $r \leq s$ by switching roles of the two products. Then argue inductively: if p_1 divides q_1 we have $p_1 \sim q_1$ since both are primes. Otherwise, p_1 does not divide q_1 and by 7.3.21 p_1 must divide $q_2 \cdot \ldots \cdot q_s$. By 7.3.22 there must be an index i_1 such that $p_1 \sim q_{i_1}$.

Dividing both products by p_1 we are left with

$$p_2 \cdot \ldots \cdot p_r \;\sim\; q_1 \cdot \ldots \cdot q_{i_1 - 1} \cdot q_{i_1 + 1} \cdot \ldots \cdot q_s$$

etc. At the end of this inductive process all the p_i will have cancelled; relabeling the surviving q_j we arrive at

$$1 \;\sim\; q_{r+1} \cdot \ldots \cdot q_s$$

which is impossible unless $r = s$. After relabeling the q_j we get $p_i \sim q_i$ for $1 \leq i \leq r$. □

7.3.24 Definition. *If $a \neq 0$ is a nonunit in a Euclidean ring R, the set* $\mathrm{sp}(a)$ *of distinct prime divisors $[p_1], \ldots, [p_r]$ is called the* **spectrum** *of a and the exponents m_i are their* **multiplicities**.

The units in the Euclidean ring $R = \mathbb{F}[x]$ are the nonzero constant polynomials $\mathrm{U}_R = \{c\dagger : c \neq 0 \text{ in } \mathbb{F}\}$, and the primes are the nonconstant polynomials that are **irreducible** (no nontrivial factorizations $f = f_1 f_2$). By degree considerations, every first degree polynomial $f = ax + b$ ($a \neq 0$) is irreducible, but depending on the nature of the coefficient field $\mathbb{F}$ there may be irreducible polynomials of higher degree, for instance $x^2 - 2$ in $\mathbb{Q}[x]$ or $x^2 + 1$ in $\mathbb{R}[x]$. If $\deg(f) \geq 1$ and $f = \prod_{i=1}^{m} f_i$ with nonconstant factors, one could multiply the f_i by various nonzero constants; however if f and the f_i are *monic*, with leading coefficients = 1, then the f_i are unique elements of $\mathbb{F}[x]$. It is often convenient to normalize things this way to deal with nonuniqueness of the factors modulo multiplied units.

The following simple properties of the GCD lead to a fast algorithm for finding the GCD of two elements in a Euclidean ring

7.3.25 Exercise. If R is a Euclidean ring and $a, b \neq 0$ prove that

(a) $\gcd(a, b) \sim \gcd(b, a)$

(b) $\gcd(a, b) \sim \gcd(a, b + ca)$ for any $c \in R$.

(c) $\gcd(a, b) \sim a \Leftrightarrow a$ divides b. □

Property (b) is the basis of the GCD algorithm, described next.

7.3.26 Example. (GCD Algorithm). If $a, b \neq 0$ in a Euclidean ring we may label them so that $d(a) \geq d(b)$. Applying the division algorithm, we write $a = q_1 b + r_1$ with $r_1 = 0$ or $d(r_1) < d(b)$. By 7.3.25(b) we have

$$\gcd(a, b) \sim \gcd(q_1 b + r_1, b) \sim \gcd(b, r_1)$$

If $r_1 = 0$ then $b|a$ and we're done: $\gcd(a, b) \sim b$. If not, we have reduced the degree of the element with the larger degree. Relabeling $a' = b, b' = r_1$ we restore the relation $d(a') \geq d(b')$ and may apply the same process again. This recursive procedure (shown below) must eventually terminate, and in fact does so quite rapidly.

$\gcd(a,b)$	$a = q_1 b + r_1$	Now relabel $a_1 = b, b_1 = r_1$
$\gcd(a_1,b_1)$	$a_1 = q_2 b_1 + r_2$	Now relabel $a_2 = b_1, b_2 = r_2$
$\gcd(a_2,b_2)$	$a_2 = q_3 b_2 + r_3$	Now relabel $a_3 = b_2, b_3 = r_2$
$\vdots$	$\vdots$	
$\gcd(a_k,b_k)$	$a_k = q_{k+1} b_k + r_{k+1}$	With $r_{k+1} = 0$, so $b_k \sim \gcd(a_k,b_k)$

At the first step in which the remainder r_{k+1} is zero, b_k divides a_k and

$$b_k \sim \gcd(a_k,b_k) \sim \ldots \sim \gcd(a,b)$$

Done. □

7.3.27 Example. Find $\gcd(22471, 3266)$.

SOLUTION: We have

$$\begin{aligned} 22471 &= 3266(6) + 2875 \\ 3266 &= 2875(1) + 391 \\ 2875 &= 391(7) + 138 \\ 391 &= 138(2) + 115 \\ 138 &= 115(1) + 23 \\ 115 &= 23(5) + 0 \end{aligned}$$

Therefore $\gcd(a,b) = 23$ □

Note: This algorithm is fast and yields the *gcd* without any need to find prime divisors of a and b. The same procedure works to produce the *gcd* of two nonzero polynomials $f, h \in \mathbb{F}[x]$, or two Gaussian integers $z, w \in \mathbb{Z}[i]$. □

If a, b are nonunits in a Euclidean ring R they may have a certain number of prime divisors in common, so we may write the factorizations as

$$a = p_1 \ldots p_k \cdot a_{k+1} \ldots a_r \qquad \text{and} \qquad b = p_1 \ldots p_k \cdot b_{k+1} \ldots b_s$$

with $r, s \geq k$ and $\gcd(a_i, b_j) \sim 1$ for $i, j > k$.

7.3.28 Exercise. If $a, b \neq 0$ are in a Euclidean ring R, show that $\gcd(a,b) \sim p_1 \ldots p_k$ (product of their common primes), or $\gcd(a,b) \sim 1$ if a, b have no prime divisors in common. □

Euclidean rings fit into a larger hierarchy

$$\left(\begin{array}{c}\text{commutative rings}\\ R \neq (0)\end{array}\right) \supseteq \left(\begin{array}{c}\text{commutative rings } R\\ \text{with identity } 1_R\end{array}\right) \supseteq \left(\begin{array}{c}\text{integral domains:}\\ \exists\, 1_R, \text{no zero divisors}\end{array}\right)$$
$$\supseteq \left(\begin{array}{c}\text{UFD = Unique}\\ \text{Factorization Domains}\end{array}\right) \supseteq \left(\begin{array}{c}\text{PID = Principal Ideal Domains:}\\ \text{Every ideal } I = aR \text{ for some } a\end{array}\right) \supseteq \left(\begin{array}{c}\text{Euclidean}\\ \text{Domains}\end{array}\right)$$

Here $R \in$ UFD means every nonunit $a \neq 0$ has a *unique* factorization into nonunit primes, as in 7.3.23; we have already indicated the meaning of PID's in proving that every Euclidean ring is a Principal Ideal Domain, see 7.3.8. It must of course be proved that (UFD) $\supseteq$ (PID). Further work with commutative rings would yield a crucial result that goes back to the work of Gauss (proof deferred to Part II of these *Notes*).

7.3.29 Theorem (Gauss). *If R is any* UFD*, so is the polynomial ring $R[x]$ with*

coefficients in R.

If $\mathbb{F}$ is a field it is a UFD by default: all nonzero elements are units, and there are no primes; the ring of integers $\mathbb{Z} \in$ UFD because it is Euclidean. From this we get

7.3.30 Corrolary. *The polynomial rings $\mathbb{Z}[x_1, \ldots, x_n]$ and $\mathbb{F}[x_1, \ldots, x_n]$ are unique factorization domains for any n.*

PROOF: If $R = \mathbb{F}$ or $\mathbb{Z}$ then by 7.3.29 we have

$$R \in \text{UFD} \;\Rightarrow\; R[x_1] \in \text{UFD} \;\Rightarrow\; R[x_1, x_2] \cong \big(R[x_1]\big)[x_2] \in \text{UFD} \;\Rightarrow\quad \text{etc} \quad \square$$

We have seen in 7.3.9 that $\mathbb{F}[x_1, x_2]$ is not a PID, so to deal with factorization of polynomials in several unknowns one must expand one's view to include at least the class of UFD rings. On the other hand our work with Euclidean domains is the basis for understanding PID's, and then in turn the class UFD. All this will be covered in Part II of these *Notes*, and is based on the observation that every unital integral domain can be embedded in a unique "*field of fractions*" $\mathbb{F} = \text{Frac}(R)$, which is generated by the elements of R much as the field of rationals $\mathbb{Q}$ is obtained from the ring of integers $\mathbb{Z}$ by forming "fractions" $\frac{m}{n}$ with $n \neq 0$. Even if an integral domain R is not a Euclidean ring, to which the preceding theory would apply, one can learn a lot about factorizations in the polynomial ring $R[x]$ by examining the ring $\mathbb{F}[x]$ with $\mathbb{F} = \text{Frac}(R)$, which in an obvious sense contains $R[x]$ and *is* a UFD. This is the fundamental idea behind the proof of Gauss' theorem.

7.4 The Fraction Field Frac(R) of an Integral Domain.

The reason that the integral domain $\mathbb{Z}[x]$ is not a Euclidean domain is that long division with remainder of certain polynomials, say $(x^2 + 1)/3x$, leads to coefficients with non-trivial denominators that no longer lie in the coefficient ring $\mathbb{Z}$; but they do lie in the field of fractions $\mathbb{Q} = \text{Frac}(\mathbb{Z})$. In the present section we assume R is an integral domain, which by definition has an identity element. We will see that the presence of an identity element in R is essential in constructing the field of fractions $\text{Frac}(R)$.

Given such an R we first observe that if $\mathbb{F}$ is a field and $\phi : R \to \mathbb{F}$ a one-to-one homomorphism then ϕ is automatically *unital*, with $\phi(1_{_R}) = 1_{\mathbb{F}}$. In fact, the set of nonzero elements $\mathbb{F}^\times \subseteq \mathbb{F}$ is an abelian group under multiplication, and the only idempotent element (solution of $x^2 = x$) in a group is its identity element. Since $\phi(1_{_R}) \neq 0$ and $1_{_R}^2 = 1_{_R}$ the element $x = \phi(1_{_R})$ satisfies the idempotent equation in $\mathbb{F}^\times$ and we have $\phi(1_{_R}) = 1_{\mathbb{F}}$.

Second, there is a smallest subfield $\mathbb{K} \subseteq \mathbb{F}$ that contains the unital subring $\phi(R)$. This field is generated by the elements of $\phi(R)$ and it is easy to verify that it consists of the elements

$$\mathbb{K} = \{\phi(a)\phi(b)^{-1} : a, b \in R \text{ and } b \neq 0\} \tag{20}$$

in $\mathbb{F}$. We now show that, up to an isomorphism of fields, $\mathbb{K}$ is *independent of the field $\mathbb{F}$ in which R was embedded.*

7.4.1 Lemma. (Uniqueness of the Generated Field). *Let R be an integral domain. Let $\mathbb{F}_1, \mathbb{F}_2$ be fields and $\phi_k : R \to \mathbb{F}_k$ one-to-one homomorphisms such that $\phi_k(R)$ generates $\mathbb{F}_k$ as a field,as in* (20)*. Then there is a unique isomorphism of fields $\psi : \mathbb{F}_1 \to \mathbb{F}_2$, automatically unital, such that $\psi \circ \phi_1 = \phi_2$ as in the following commutative diagram.*

PROOF: The rings $R_k = \phi_k(R)$ are unital and $\tilde{\psi} = \phi_2 \circ \phi_1^{-1} : R_1 \to R_2$ is a bijective unital homomorphism, hence an isomorphism. If $x = ab^{-1}$ in $\mathbb{F}_1$ with $a, b \in R_1$ and $b \neq 0$, we define $\psi : \mathbb{F}_1 \to \mathbb{F}_2$ letting

$$\psi(x) = \tilde{\psi}(a)\tilde{\psi}(b)^{-1} \quad .$$

The commutative diagram in Lemma 7.4.1.

$$\begin{array}{ccccc} & & \phi_1(R) & \subseteq & \mathbb{F}_1 \\ & \overset{\phi_1}{\nearrow} & & & \\ R & & \downarrow\tilde{\psi} & & \downarrow\psi \\ & \underset{\phi_2}{\searrow} & & & \\ & & \phi_2(R) & \subseteq & \mathbb{F}_2 \end{array}$$

This map is well-defined because if $ab^{-1} = a_1b_1^{-1}$ for a, b, a_1, b_1 in R_1 we have

$$\begin{aligned} ab^{-1} = a_1b_1^{-1} \text{ in } \mathbb{F}_1 &\Leftrightarrow ab_1 = a_1b \\ &\Rightarrow \tilde{\psi}(a)\tilde{\psi}(b_1) = \tilde{\psi}(a_1)\tilde{\psi}(b) \text{ in } R_2 \\ &\Rightarrow \tilde{\psi}(a)\tilde{\psi}(b)^{-1} = \tilde{\psi}(a_1)\tilde{\psi}(b_1)^{-1} \text{ in } \mathbb{F}_2 \end{aligned}$$

It is also a bijection: onto, because $\tilde{\psi}(R_1) = R_2$ and these subrings generate the $\mathbb{F}_k$ as in (20). It is one-to-one because $\tilde{\psi}$ is a bijection, hence

$$\begin{aligned} \psi(ab^{-1}) = \psi(a_1b_1^{-1}) &\Rightarrow \tilde{\psi}(a)\tilde{\psi}(b)^{-1} = \tilde{\psi}(a_1)\tilde{\psi}(b_1)^{-1} \Rightarrow \tilde{\psi}(ab_1) = \tilde{\psi}(a_1b) \\ &\Rightarrow ab_1 = a_1b \text{ in } R \Rightarrow ab^{-1} = a_1b_1^{-1} \text{ in } \mathbb{F}_1 \end{aligned}$$

Finally, ψ is a homomorphism of fields because

$$\begin{aligned} \psi(ab^{-1} \cdot a_1b_1^{-1}) &= \psi\big(aa_1 \cdot (bb_1)^{-1}\big) = \tilde{\psi}(a)\tilde{\psi}(a_1)\big[\tilde{\psi}(b)\tilde{\psi}(b_1)\big]^{-1} \\ &= \tilde{\psi}(a)\tilde{\psi}(a_1)\tilde{\psi}(b_1)^{-1}\tilde{\psi}(b)^{-1} = \psi(ab^{-1}) \cdot \psi(a_1b_1^{-1}) \end{aligned}$$

and similarly for sums. □

7.4.2 Theorem. (Fraction Field Construction). *If R is an integral domain there exist a field $\mathbb{F} = \mathrm{Frac}(R)$ and a one-to-one unital homomorphism $\phi : R \to \mathbb{F}$ such that $\phi(R)$ generates $\mathbb{F}$ as in* (20)*. The pair $(\phi, \mathbb{F})$ is essentially unique in the sense of Lemma* 7.4.1.

Proof: We define an RST equivalence relation on the set of ordered pairs $R \times R^\times = \{(a, b) : a, b \in R, b \neq 0\}$, letting

$$(a, b) \sim (a', b') \iff ab' = a'b \quad \text{in } R \tag{21}$$

(an idea shamelessly lifted from the construction of the rationals in Chapter 2). With this in mind, think of pairs (a, b) as defining "fraction symbols" $\frac{a}{b}$, whose equivalence classes are denoted by $\left[\frac{a}{b}\right]$. When the quotient space $\mathbb{F} = (R \times R^\times)/(\sim)$ is equipped with the appropriate algebraic operations $(+), (\cdot)$ it will become the desired fraction field.

These operations are defined by referring to class representatives. We write

$$\mathbf{0} = \left[\frac{0}{1}\right] \qquad \text{and} \qquad \mathbf{1} = \left[\frac{1}{1}\right] = \left[\frac{x}{x}\right] \text{ for any } x \neq 0 \ ,$$

and then define operations on equivalence classes

$$\begin{aligned} \left[\frac{a}{b}\right] + \left[\frac{c}{d}\right] &= \left[\frac{ad + bc}{bd}\right] \\ \left[\frac{a}{b}\right] \cdot \left[\frac{c}{d}\right] &= \left[\frac{ac}{bd}\right] \end{aligned} \tag{22}$$

To see that these make sense independent of the class representatives used to determine the outcomes, consider equivalent pairs $(a', b') \sim (a, b)$ and $(c', d') \sim (c, d)$. Then by (21) we have $ab' = a'b$, $cd' = c'd$, and for the $(+)$ operation we may rewrite

$$adb'd' + bcb'd' = a'bdd' + cd'bb' = bdd'a' + bb'c'd$$

Thus

$$b'd'(ad+bc) = bd(a'd'+b'c') \qquad \text{and} \qquad \frac{ad+bc}{bd} \sim \frac{a'd'+b'c'}{b'd'}$$

as required. The proof for the $(\cdot)$ operation is similar but easier.

Once we know the operations are well-defined, it is immediate that both operations are commutative, associative, and satisfy the usual distributive laws because these hold in R. Furthermore

$$\mathbf{0} \cdot \left[\frac{a}{b}\right] = \mathbf{0} \qquad \mathbf{1} \cdot \left[\frac{a}{b}\right] = \left[\frac{a}{b}\right] \qquad \mathbf{0} + \left[\frac{a}{b}\right] = \left[\frac{a}{b}\right]$$

for all $\left[\frac{a}{b}\right]$, and it is a straightforward matter to check the commutative ring axioms for the system $(\mathbb{F}, +, \cdot)$.

7.4.3 Exercise. Verify the following properties starting from definitions (21) and (22)

(a) The $(+)$ operation makes $(\mathbb{F}, +)$ an additive group with $\mathbf{0}$ as its additive identity element, the additive inverse of $\left[\frac{a}{b}\right]$ being $\left[\frac{-a}{b}\right] = \left[\frac{-1}{1}\right] \cdot \left[\frac{a}{b}\right] = -\mathbf{1} \cdot \left[\frac{a}{b}\right]$.

(b) The multiplication operation $(\cdot)$ is associative and commutative, and the usual distributive laws connecting $(+)$ and $(\cdot)$ hold.

(c) $\mathbb{F}$ has $\mathbf{1}$ as its multiplicative identity element. $\square$

It remains only to show that every nonzero element $\left[\frac{a}{b}\right]$ has a multiplicative inverse. First, the fraction symbol $\frac{b}{a}$ makes sense because

$$\left[\tfrac{a}{b}\right] \neq \mathbf{0} \;\Rightarrow\; \tfrac{a}{b} \not\sim \mathbf{0} \sim \tfrac{0}{1} \;\Rightarrow\; a = 1 \cdot a \neq 0 \cdot b = 0 \quad \text{in } R \;\;,$$

by (21), and then $\left[\frac{a}{b}\right]^{-1} = \left[\frac{b}{a}\right]$ because $\left[\frac{a}{b}\right] \cdot \left[\frac{b}{a}\right] = \left[\frac{ab}{ab}\right] = \left[\frac{1}{1}\right] = \mathbf{1}$. Thus

Any nonzero element $\left[\frac{a}{b}\right]$ *in* $\mathbb{F}$ *has a multiplicative inverse* $\left[\frac{a}{b}\right]^{-1} = \left[\frac{b}{a}\right]$.

Finally, we define the embedding $\phi : R \to \mathbb{F}$ letting $\phi(x) = \left[\frac{x}{1}\right]$. This map is one-to-one because $(x,1) \sim (y,1) \Leftrightarrow x = y$ in R. It is also a unital ring homomorphism because $\phi(1) = \mathbf{1}$ and

$$\begin{aligned} \phi(xy) &= \left[\frac{xy}{1}\right] = \left[\frac{x}{1}\right] \cdot \left[\frac{y}{1}\right] = \phi(x)\phi(y) \\ \phi(x+y) &= \left[\frac{x+y}{1}\right] = \left[\frac{x}{1}\right] + \left[\frac{y}{1}\right] = \phi(x) + \phi(y) \end{aligned}$$

in $\mathbb{F}$. Clearly $\phi(R)$ generates $\mathbb{F}$ because an arbitrary element in $\mathbb{F}$ can be written as

$$\left[\tfrac{a}{b}\right] = \left[\tfrac{a}{1}\right] \cdot \left[\tfrac{1}{b}\right] = \left[\tfrac{a}{1}\right] \cdot \left[\tfrac{b}{1}\right]^{-1} = \phi(a)\phi(b)^{-1}$$

for $a, b \in R$, $b \neq 0$. $\square$

It should be obvious from the discussion at the end of Chapter 2 that $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$.

7.4.4 Exercise. Prove that

$$\left[\tfrac{ac}{bc}\right] = \left[\tfrac{a}{b}\right] \quad \text{in } \mathrm{Frac}(R) \text{ for all } c \neq 0$$

Thus the representative of a nonzero class $\left[\frac{a}{b}\right]$ can be chosen so that $\gcd(a,b) \sim 1$. $\square$

7.4.5 Exercise. In $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$ the identity $\left[\frac{a}{b}\right] = \left[\frac{a'}{b'}\right]$ does not necessarily mean that $a' = ac$ and $b' = bc$ for some $c \neq 0$. Exhibit two fraction symbols $\frac{a}{b} \sim \frac{a'}{b'}$ that are

equivalent, but $\frac{a'}{b'}$ is not of the form $\frac{ac}{bc}$ for any nonzero c.
Hint: Keep it simple. Try various fraction symbols equivalent to $\frac{1}{3}$. $\square$

7.4.6 Exercise. If $a, b \neq 0$ and $\gcd(a,b) \sim 1$, prove that $\left[\frac{a}{b}\right] = \left[\frac{a'}{b'}\right]$ if and only if there exists some $c \neq 0$ such that $b' = bc$ and $a' = ac$. $\square$

7.4.7 Exercise. If an integral domain R is already a field (every nonzero element has a multiplicative inverse), verify that $R \cong \mathrm{Frac}(R)$ as fields. $\square$.

7.4.8 Example (Fields of Rational Functions). Let $\mathbb{F}$ be a field and $R = \mathbb{F}[x]$ an arbitrary polynomial ring – an integral domain whose identity element is the constant polynomial $f = \mathbb{1}$. The fraction field $\mathbb{F}(x) = \mathrm{Frac}\big(\mathbb{F}[x]\big)$ is the **field of rational functions** in the indeterminate x whose elements are determined by quotients of polynomials

$$\frac{P(x)}{Q(x)} \quad \text{where} \quad P(x), Q(x) \in \mathbb{F}[x] \text{ and } Q \not\equiv 0$$

However, elements in $\mathbb{F}(x)$ are actually the equivalence classes of the RST relation

$$\frac{P(x)}{Q(x)} \sim \frac{P'(x)}{Q'(x)} \;\Leftrightarrow\; P(x)Q'(x) = P'(x)Q(x) \qquad \text{in } \mathbb{F}[x]$$

The operations $(+)$ and $(\cdot)$ in $\mathbb{F}(x)$ are determined by the familiar operations on fraction symbols

$$\frac{P}{Q} + \frac{R}{S} = \frac{PS + QR}{QS} \qquad \text{and} \qquad \frac{P}{Q} \cdot \frac{R}{S} = \frac{PR}{QS}$$

Similarly, we can define the fraction field $\mathrm{Frac}(R) = \mathbb{F}(x_1, \ldots, x_n)$ of a multi-variable polynomial ring $R = \mathbb{F}[x_1, \ldots, x_n]$. This consists of equivalence classes of formal quotients $P(x_1, \ldots, x_n)/Q(x_1, \ldots, x_n)$ whose denominators are not the zero polynomial, for instance $(x^2 + xy + y^2)/(x^2 - y^2)$ in $\mathbb{C}(x, y)$.

Notation. In most situations it is customary to ignore the distinction between fraction symbols P/Q and their equivalence classes $[P/Q]$ in the fraction field. The fraction field of a polynomial ring $\mathbb{F}[x]$ is usually denoted $\mathbb{F}(x)$ instead of $\mathrm{Frac}\big(\mathbb{F}[x]\big)$. $\square$

7.5. Fields and Maximal Ideals in Commutative Rings.

An ideal I in a commutative ring is **proper** if $(0) \neq I \neq R$. It is a **maximal ideal** if $I \neq R$ and there are no ideals J such that $I \subsetneq J \subsetneq R$. Maximal ideals play a crucial role in many investigations. For instance they are needed to understand how one might enlarge a field $\mathbb{F}$ by "adjoining" to $\mathbb{F}$ some roots to an irreducible polynomial $f(x) \in \mathbb{F}[x]$, which by definition has no roots in the field $\mathbb{F}$. For instance $x^2 - 2 \in \mathbb{Q}[x]$ has no roots in the rationals, but we will see how to construct (using maximal ideals) a larger field $\mathbb{E} = \mathbb{Q}(\sqrt{2}) \supseteq \mathbb{Q}$ in which a square root of 2 exists.

In algebraic geometry maximal ideals provide a bijective correspondence between the

geometric space $\mathbb{C}^n$ and algebraic objects in the polynomial ring $\mathbb{C}[\mathbf{x}] = \mathbb{C}[x_1, \ldots, x_n]$.

(23) HILBERT NULLSTELLENSATZ. *An ideal $M \subseteq \mathbb{C}[\mathbf{x}]$ is a maximal ideal $\Leftrightarrow$ there exists a point $\mathbf{p} = (p_1, \ldots, p_n)$ in coordinate space $\mathbb{C}^n$ such that $M = \ker(\epsilon_p)$ where ϵ_p is the "evaluation homomorphism"*

$$\epsilon_p\Big(\sum_{\alpha \in \mathbb{Z}_+^n} c_\alpha x^\alpha\Big) = \sum_{\alpha \in \mathbb{Z}_+^n} c_\alpha\, p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n}$$

This is equivalent to saying: M is a maximal ideal in $\mathbb{C}[x] \Leftrightarrow$ it is the ideal

$$M = (x_1 - p_1) \cdot \mathbb{C}[\mathbf{x}] + \ldots + (x_n - p_n) \cdot \mathbb{C}[\mathbf{x}]$$

generated by the first degree polynomials $(x_1 - p_1), \ldots, (x_n - p_n)$ for some point $\mathbf{p} = (p_1, \ldots, p_n)$ in complex coordinate space $\mathbb{C}^n$.

This result, which we won't prove here, establishes a bijective correspondence between geometric and algebraic entities

$$\Big(\text{ maximal ideals in } \mathbb{C}[\mathbf{x}]\Big) \quad \longleftrightarrow \quad \Big(\text{ points } \mathbf{p} = (p_1, \ldots, p_n) \text{ in } \mathbb{C}^n\Big)$$

It is just the first step in building an extensive concordance between algebraic objects in $\mathbb{C}[\mathbf{x}]$ and geometric curves in complex coordinate space $\mathbb{C}^n$.

There are no maximal ideals in a field. In fact if $I \subseteq R$ is an ideal in a field and $I \neq (0)$, then any element $a \neq 0$ is a unit, hence $aR = R$ and $I = R$; in particular a field has no proper ideals. In fact,

7.5.1 Proposition. *If $R \neq (0)$ is an arbitrary commutative ring with identity (existence of zero divisors in R not excluded), then R has no proper ideals $\Leftrightarrow$ R is a field.*

PROOF ($\Rightarrow$): Let $\pi : R \to R/I = \overline{R}$ be the quotient homomorphism, noting that $\overline{R} \neq (0)$ since $I \neq R$. If $\overline{a} \neq \overline{0}$ in $\overline{R}$ and $a \in R$ is any preimage, so that $\pi(a) = \overline{a}$, then $\overline{J} = \overline{a}\overline{R}$ is an ideal in $\overline{R}$ and its pullback $J = \pi^{-1}(\overline{J})$ is an ideal in R that contains both I and a. Since I is maximal we must have $J = R$ and $\overline{J} = \overline{R}$, so there is some $\overline{x} \in \overline{R}$ such that $\overline{a} \cdot \overline{x} = \overline{1} = \pi(1)$. Here, $\overline{1}$ is an identity for $\overline{R}$ because π is surjective, and $\overline{1} \neq \overline{0}$ (otherwise the identity element 1 would lie in I and $I = R$, contrary to the definition of maximality). We conclude that every nonzero element $\overline{a}$ in $\overline{R}$ is a unit, so $\overline{R}$ is a field.

PROOF ($\Leftarrow$): If $I \neq (0)$ is an ideal in a field R any $x \neq 0$ in I has a multiplicative inverse x^{-1}. Then every element $a \in R$ can be written as $a = (ax^{-1}) \cdot x \in (ax^{-1})I \subseteq I$, hence $I = R$. $\square$

As noted above, an ideal I in a commutative ring with identity 1 is all of R if $1 \in I$ or if I contains a unit. In particular, $I = aR$ is equal to $R \Leftrightarrow a \in \mathrm{U}_R$.

Maximal ideals always exist in any commutative ring with identity, although constructing such ideals is another matter. A Zorn's lemma argument (Axiom of Choice from set theory) shows that

(24) *If $I \neq R$ is an ideal in a commutative ring R with identity, then I is contained in a maximal ideal M, so $I \subseteq M \subsetneq R$.*

The maximal ideal M containing J is not necessarily unique.

Maximal ideals in $\mathbb{Z}$ are of considerable interest, and are easily determined.

7.5.2 Example. The maximal ideals in $\mathbb{Z}$ are the principal ideals $(p) = p \cdot \mathbb{Z}$ for primes $p > 1$. Cosets in the quotient space $\mathbb{Z}/(p)$ are precisely the (mod p) congruence classes in $\mathbb{Z}_p$, so $\mathbb{Z}/(p) = \mathbb{Z}_p$ as sets.

DISCUSSION: As we saw in 7.3.8, every ideal in $\mathbb{Z}$ is a principal ideal $(m) = m \cdot \mathbb{Z}$ for

some $m \geq 0$. If $m = 0$ then $(m) = (0)$ is trivial, and if $m = 1$ (a unit) then $(m) = \mathbb{Z}$ and $\mathbb{Z}/(m)$ is trivial. Neither ideal is maximal in $\mathbb{Z}$. Assuming $m > 1$, we show that

(25) $\qquad (m) = m{\cdot}\mathbb{Z}$ is a maximal ideal in $\mathbb{Z} \;\Leftrightarrow\; m$ is a prime.

PROOF ($\Leftarrow$): We argue by contradiction. If p is a prime $I = (p)$ is maximal, because otherwise there would be some $n > 1$ such that $(p) \subsetneq (n) \subsetneq \mathbb{Z}$. But then $p \in (n)$, $p = nx$ for some $x \neq 0$ in $\mathbb{Z}$, and at least one of the factors is a unit. Since $n > 1$ and $\mathrm{U}_{\mathbb{Z}} = \{\pm 1\}$, n is not a unit so x must be a unit. That would imply $x{\cdot}\mathbb{Z} = \mathbb{Z}$ and hence

$$(n) = n{\cdot}\mathbb{Z} = n{\cdot}x{\cdot}\mathbb{Z} = p{\cdot}\mathbb{Z} \text{ is equal to } (p),$$

contrary to our hypotheses.

PROOF ($\Rightarrow$): Conversely if $m > 1$ is not a prime it would be a product $m = ab$ with neither factor a unit. We show that this would imply $(m) \subsetneq (a) \subsetneq \mathbb{Z}$ contrary to maximality of (m), thereby proving ($\Rightarrow$).

The equality $m = ab$ implies

$$(m) = m{\cdot}\mathbb{Z} = ab{\cdot}\mathbb{Z} \subseteq a{\cdot}\mathbb{Z}$$

so $(m) \subseteq (a)$; we argue by contradiction to show that $(m) \neq (a)$. If $(m) = (a)$ then by 7.3.8(iii) there would be a unit u such that $m = au$. Since $ab = m = au$ and $a \neq 0$, cancellation shows that $b = u$ (a unit), contrary to our hypotheses. Thus $(m) \subsetneq (a)$.

We also have $(a) \neq \mathbb{Z}$, otherwise $1 = ax$ for some x and a would be a unit. Contradiction. Therefore $(m) \subsetneq (a) \subsetneq \mathbb{Z}$ and (m) would not be maximal. $\square$

We now apply these remarks to a polynomial ring $R = \mathbb{F}[x]$ over a field $\mathbb{F}$. The discussion uses the fact that $\mathbb{F}[x]$ is more than a ring, it is a commutative **associative algebra** over $\mathbb{F}$: in addition to the $(+)$ and $(\cdot)$ ring operations there is a natural *scaling operation* $a \mapsto c \cdot a$ for elements $c \in \mathbb{F}$. Scaling has the properties

$$\begin{array}{rclcrcll} c{\cdot}(f+h) & = & (c{\cdot}f) + (c{\cdot}h) & \qquad & (bc){\cdot}f & = & b{\cdot}(c{\cdot}f) & \\ c{\cdot}(fh) & = & (c{\cdot}f)\,h & & 1_{\mathbb{F}} \cdot f & = & f & \text{for all } f \end{array}$$

and $c \cdot \left(\sum_{i=0}^{m} a_i x^i\right) = \sum_{i=0}^{m} (c{\cdot}a_i) x^i$. This makes $\mathbb{F}[x]$ into an infinite dimensional vector space over $\mathbb{F}$ with basis vectors $\mathbb{1}, x, x^2, \ldots, x^n, \ldots$

The ground field $\mathbb{F}$ is itself an associative algebra, with scaling operation $a \mapsto ca$; furthermore, $\mathbb{F}[x]$ contains a faithful copy of the ground field, namely the set of constant polynomials $\tilde{\mathbb{F}} = \{c\mathbb{1} : c \in \mathbb{F}\}$. The embedding map $j : c \mapsto c{\cdot}\mathbb{1}$ from $\mathbb{F}$ to $\tilde{\mathbb{F}} \subseteq \mathbb{F}[x]$ is a unital isomorphism of fields.

7.5.3 Theorem (Maximal Ideals in F[x]). *If $\mathbb{F}$ is a field and $f(x)$ a nonconstant polynomial in $\mathbb{F}[x]$, the principal ideal $I = (f)$ is a maximal ideal $\Leftrightarrow$ $f(x)$ is an irreducible polynomial (no nontrivial factorizations $f = g{\cdot}h$).*

PROOF ($\Leftarrow$): If (f) is not maximal there is some ideal such that $(f) \subsetneq I \subsetneq \mathbb{F}[x]$. But all ideals in this Euclidean ring are principal, so $I = (h)$ for some $h \neq 0$ and we may write $f = gh$. The factor g can't be a unit, since otherwise $(f) = (h) = I$; and h can't be a unit because we would then have $(h) = \mathbb{F}[x]$. Thus f is reducible (not irreducible) when (f) is not maximal, proving ($\Leftarrow$). $\square$

PROOF ($\Rightarrow$): If f is reducible we may write $f = gh$ in which both factors are non-units. Then $(h) \neq \mathbb{F}[x]$, otherwise there would be some $u \in \mathbb{F}[x]$ such that $hu = \mathbb{1}$ and h would be a unit. Contradiction.

Furthermore $(f) \subsetneqq (h)$. We know $(f) \subseteq (h)$ because $f = gh$, but in fact $(f) \subsetneqq (h)$; otherwise $(f) = (h)$ and $\deg(f) = \deg(h)$ since the generators of a principal ideal are precisely the nonzero elements of minimal degree. But $\deg(f) = \deg(gh) = \deg(g) + \deg(h)$ implies $\deg(g) = 0$, so $g(x)$ is constant with $g = c\mathbb{1}$ for some $c \neq 0$, and g is a unit contrary to our hypotheses. Hence $(f) \subsetneqq (h) \subsetneqq \mathbb{F}[x]$ and the ideal (f) is not maximal, proving ($\Rightarrow$). $\square$

7.5.4 Corollary. *If $\mathbb{F}$ is a field and $f \in \mathbb{F}[x]$ is a nonconstant polynomial, then the quotient ring $R = \mathbb{F}[x]/(f)$ is a* FIELD $\Leftrightarrow$ *$f(x)$ is irreducible.*

The fact that this quotient ring is already a field is surprising. You might expect to have to form "fractions " $\overline{a}/\overline{b}$ involving elements $\overline{a}, \overline{b} \in \mathbb{F}[x]/(f)$ in order to obtain a field, but that is not so.

7.5.5 Example. The polynomials $x^2 - 2$ in $\mathbb{Q}[x]$ and $x^2 + 1$ in $\mathbb{R}[x]$ are irreducible over $\mathbb{Q}$ and $\mathbb{R}$ respectively. We shall identify the quotient fields as

$$\begin{aligned} &\mathbb{Q}[x]/(x^2 - 2) \cong \mathbb{Q} + \sqrt{2}\,\mathbb{Q} \qquad \text{(the field constructed in Example 7.1.6)} \\ &\mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C} = \mathbb{R} + i\mathbb{R} \qquad \text{(the field of complex numbers).} \end{aligned} \tag{26}$$

In the first example the effect of the quotient construction is to adjoin a new entity $\sqrt{2}$ to the ground field $\mathbb{Q}$ to obtain a field extension $\mathbb{E} = \mathbb{Q}(\sqrt{2})$ containing roots of $x^2 - 2$, which did not exist in $\mathbb{Q}$. In the second, a new element $\sqrt{-1}$ has been adjoined to $\mathbb{R}$. A detailed explanation of how these new roots arise is given in Section 7.6. $\square$

Irreducible Polynomials. If $\mathbb{F}$ is a field the "primes" in $\mathbb{F}[x]$ are the irreducible nonconstant polynomials, those that cannot be written as a product $f = g \cdot h$ of polynomials with degrees $< \deg(f)$. If $f(x)$ has a root α in $\mathbb{F}$ then f cannot be irreducible because we can long divide by $(x - \alpha)$ to get $f(x) = (x - \alpha)Q(x)$. If $Q(x)$ also has α as a root we may continue this process, arriving at a factorization $f(x) = (x - \alpha)^m Q(x)$ such that $Q(x)$ does not have α as a root. If we can find other roots of $Q(x)$ in $\mathbb{F}$ we can continue peeling off linear factors, finally arriving at a factorization

$$f(x) = \prod_{i=1}^{r} (x - \alpha_i)^{m_i} \cdot Q(x) \qquad \text{with } \alpha_1, \ldots, \alpha_r \text{ in } \mathbb{F}$$

in which $Q(x)$ has *no roots at all* in $\mathbb{F}$. Unfortunately, by itself the absence of any roots in the ground field is not enough to guarantee irreducibility of $Q(x)$ in $\mathbb{F}[x]$, though sometimes it does.

7.5.6 Exercise. For any field $\mathbb{F}$ and nonconstant polynomials $f \in \mathbb{F}[x]$ prove that:

(a) All linear polynomials ($\deg f = 1$) are irreducible.

(b) A quadratic polynomial ($\deg f = 2$) is irreducible $\Leftrightarrow$ it has no roots in $\mathbb{F}$.

(c) A cubic polynomial ($\deg f = 3$) is irreducible $\Leftrightarrow$ it has no roots in $\mathbb{F}$.

Note: Things get more complicated when $\deg f \geq 4$. $\square$

7.5.7 Exercise. Verify irreducibility, or lack thereof, in the following situations.

(a) $f(x) = x^4 + 1$ is irreducible in $\mathbb{R}[x]$, hence also in $\mathbb{Q}[x]$, but is reducible in $\mathbb{C}[x]$.

(b) $f(x) = x^3 - 5$ is irreducible in $\mathbb{Q}[x]$, but is reducible in $\mathbb{R}[x]$.

(c) Is the polynomial $f(x) = x^3 + x + 1$ irreducible in the ring $\mathbb{Z}_5[x]$? Is it irreducible as an element of $\mathbb{Z}_2[x]$? $\square$

7.5.8 Exercise. Prove that:

(a) The polynomial $f(x) = x^3 + x^2 + x + 1$ is reducible in $\mathbb{Q}[x]$, hence also in $\mathbb{R}[x]$.

(b) We can write $f(x) = (x - r) \cdot Q(x)$ where $r \in \mathbb{R}$ and $Q(x)$ is an irreducible quadratic polynomial in $\mathbb{R}[x]$.

(c) What is the unique irreducible factorization of $f(x)$ over $\mathbb{C}$?

Note: Part (b) yields the unique prime factorization of $x^3 + x^2 + x + 1$ in the Euclidean ring $\mathbb{R}[x]$. The same prime factorization also holds in the ring $\mathbb{Q}[x]$. $\square$

The *Fundamental Theorem of Algebra* asserts that the complex number field is **algebraically closed**.

> Fundamental Theorem of Algebra. *Every nonconstant polynomial with coefficients in $\mathbb{C}$ has at least one root in $\mathbb{C}$.*

None of the other fields $\mathbb{Q}, \mathbb{R}, \mathbb{Z}_p$, or the fields of rational functions $\mathbb{F}(x), \mathbb{F}(x_1, \ldots, x_n)$, encountered so far have this property. In $\mathbb{C}[x]$ the process of peeling off one linear factor $(x - \alpha)$ for each root in the ground field $\mathbb{F}$ terminates in a complete splitting of $f(x)$ into linear factors:

$$f(x) = c \cdot \prod_{j=1}^{r} (x - z_j)^{m_j} \qquad (m_1 + \ldots + m_r = n = \deg\ f)$$

where c is the coefficient of the leading term in f. In particular the monic irreducible polynomials in $\mathbb{C}[x]$ all have the form $(x - z)$ with $z \in \mathbb{C}$, and the product above is the prime factorization of $f(x)$.

Interesting things happen when real-coefficient polynomials are regarded as complex polynomials $\mathbb{R}[x] \subseteq \mathbb{C}[x]$ that happen to have real coefficients.

7.5.9 Exercise. If $f(x)$ is a nonconstant polynomial in $\mathbb{R}[x]$ some roots may be real and others non-real complex $z = x + iy$ with $y \neq 0$. Prove that the non-real roots must occur in *conjugate pairs* $z, \overline{z}$ where

$$\overline{z} = x - iy \qquad \text{is the complex conjugate of } z = x + iy$$

Hint: Use the real-coefficient property to show that $f(z) = 0 \Rightarrow f(\overline{z}) = 0$. Recall that $\overline{z + w} = \overline{z} + \overline{w}$, $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$, and $\left(\overline{z}\right)^{-} .$ $\square$

In particular the number of nonreal roots must be even.

7.5.10 Exercise. Use Exercise 7.5.9 and the Fundamental Theorem to prove that;

(a) A quadratic $Ax^2 + Bx + C$ in $\mathbb{R}[x]$ is irreducible $\Leftrightarrow$ its *discriminant* is negative: $B^2 - 4AC < 0$.

(b) Every monic polynomial $f \in \mathbb{R}[x]$ is a product of irreducibles that have the form

- Linear factors: $(x - r)$ with $r \in \mathbb{R}$,
- Irreducible quadratic factors: $x^2 + Bx + C$ with $B, C \in \mathbb{R}$ and $B^2 - 4C < 0$.

Hint: Write out $(x - z)(x - \overline{z})$ for a nonreal complex number $z = a + ib$. Show that a monic real polynomial $x^2 + Bx + C$ is irreducible $\Leftrightarrow$ it is equal to $(x - z)(x - \overline{z})$ for some *nonreal* complex number z. $\square$

7.6. Extension Fields and Adjunction of Roots.

An **extension** of a field $\mathbb{F}$ is any field $\mathbb{E}$ that contains $\mathbb{F}$, for example $\mathbb{R} \supseteq \mathbb{Q}$ and $\mathbb{C} \supseteq \mathbb{R}$ are field extensions. We will exploit the fact that an extension $\mathbb{E} \supseteq \mathbb{F}$ becomes a vector space over $\mathbb{F}$ if we only allow elements of $\mathbb{E}$ to be scaled by elements $\lambda \in \mathbb{F}$. The dimension $\dim_{\mathbb{F}}(\mathbb{E})$ of $\mathbb{E}$ as a vector space over the restricted field of scalars $\mathbb{F}$ could be finite or infinite dimensional. For instance $\dim_{\mathbb{Q}}(\mathbb{R}) = \infty$ ($\mathbb{Q}$-linear combinations of a finite set $x_1, \ldots, x_n$ of real numbers yield a countable set, but $\mathbb{R}$ is uncountable), and $\dim_{\mathbb{R}}(\mathbb{C}) = 2$ since every complex number can be written as $z = x\cdot 1 + y\cdot\sqrt{-1}$, so the complex numbers $\{1, i\}$ are an $\mathbb{R}$-basis when $\mathbb{C}$ is regarded as a vector space over $\mathbb{R}$. The dimension of $\mathbb{E}$ over $\mathbb{F}$ is often denoted by $\dim_{\mathbb{F}}(\mathbb{E}) = [\mathbb{E} : \mathbb{F}]$, and the initial field $\mathbb{F}$ is generally referred to as the *ground field*.

7.6.1 Exercise. Given a field extension $\mathbb{E} \supseteq \mathbb{F}$, explain why the identity element in $\mathbb{F}$ must agree with that in the larger field: $1_{\mathbb{F}} = 1_{\mathbb{E}}$. □

7.6.2 Exercise. The field $\mathbb{E} = \mathbb{Q}[\sqrt{2}\,] = \mathbb{Q} + \sqrt{2}\,\mathbb{Q}$ defined in Example 7.1.6 is an extension of $\mathbb{Q}$. Explain why $\dim_{\mathbb{Q}}(\mathbb{E}) = 2$. □

7.6.3 Exercise. If $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ are finite dimensional field extensions, prove that

$$[\mathbb{K} : \mathbb{F}] = [\mathbb{K} : \mathbb{E}] \cdot [\mathbb{E} : \mathbb{F}] \tag{27}$$

Hint: Let $\{e_1, \ldots, e_m\}$ and $\{f_1, \ldots, f_n\}$ be bases for $\mathbb{E}$ over $\mathbb{F}$ and for $\mathbb{K}$ over $\mathbb{E}$ respectively. Prove that the products $\{e_i\cdot f_j : 1 \le i \le m, 1 \le j \le n\}$ in $\mathbb{K}$ are a basis for $\mathbb{K}$ over $\mathbb{F}$. □

7.6.4 Notation. *Given an extension $\mathbb{E} \supseteq \mathbb{F}$ and elements $a_1, \ldots, a_r$ in $\mathbb{E}$ there exist:*

(i) *A smallest* SUBRING *$\mathbb{F}[a_1, \ldots, a_r] \subseteq \mathbb{E}$ that contains $\mathbb{F}$ and the elements a_i. It consists of all finite sums $\sum_{\lambda \in \mathbb{Z}_+^r} c_\lambda a^\lambda$ over multi-indices $\lambda = (m_1, \ldots, m_r)$, where $c_\lambda \in \mathbb{F}$ and $a^\lambda = a_1^{\lambda_1}\cdot\ldots\cdot a_r^{\lambda_r}$.*

(ii) *A smallest* SUBFIELD *$\mathbb{F}(a_1, \ldots, a_r) \subseteq \mathbb{E}$ containing $\mathbb{F}$ and the elements a_i. This extension of $\mathbb{F}$ consists of all quotients*

$$\frac{f(\mathbf{a})}{g(\mathbf{a})} = \frac{f(a_1, \ldots, a_r)}{g(a_1, \ldots, a_r)} \qquad \big(\mathbf{a} = (a_1, \ldots, a_r),\ a_i \in \mathbb{F}\big) \tag{28}$$

with $f, g \in \mathbb{F}[x_1, \ldots, x_r]$ and $g(\mathbf{a}) \neq 0$ in $\mathbb{E}$.

7.6.5 Exercise. Writing $h(\mathbf{a}) = h(a_1, \ldots, a_r)$, verify the description of the subfield $\mathbb{F}(a_1, \ldots, a_r)$ given in equation (28). In particular explain why sums

$$\frac{f_1(\mathbf{a})}{g_1(\mathbf{a})} + \ldots + \frac{f_s(\mathbf{a})}{g_s(\mathbf{a})}$$

have the desired form $f(\mathbf{a})/g(\mathbf{a})$. □

Although elements in $\mathbb{F}(a_1, \ldots, a_r)$ can be written as $f(\mathbf{a})/g(\mathbf{a})$ with f, g in $\mathbb{F}[x_1, \ldots, x_n]$ and $g(\mathbf{a}) \neq 0$, different polynomials f', g' may yield the same outcome. In fact if h is any polynomial in $x_1, \ldots, x_n$ with $h(\mathbf{a}) \neq 0$ we can take $f' = f\cdot h$ and $g' = g\cdot h$.

Algebraic Elements in a Field Extension. An element $a \in \mathbb{E}$ in an extension $\mathbb{E} \supseteq \mathbb{F}$ is **algebraic** over $\mathbb{F}$ if $h(a) = 0$ for some nonconstant polynomial $h \in \mathbb{F}[x]$, and is otherwise said to be **transcendental** over $\mathbb{F}$. If the dimension $[\,\mathbb{E} : \mathbb{F}]$ is finite, every $a \in \mathbb{E}$ is algebraic because only finitely many of the powers $1, a, a^2, \ldots$ can be linearly independent over $\mathbb{F}$. We say $\mathbb{E}$ is an **algebraic extension** of $\mathbb{F}$ if every element in $\mathbb{E}$ is

algebraic over $\mathbb{F}$. Every finite dimensional extension $\mathbb{E} \supseteq \mathbb{F}$ is algebraic, but there are algebraic extensions such that $[\,\mathbb{E} : \mathbb{F}] = \infty$.

When $a \in \mathbb{E}$ is algebraic over $\mathbb{F}$ there is a nonconstant polynomial $h(x) \in \mathbb{F}[x]$ of lowest degree such that $h(a) = 0$ in $\mathbb{E}$. If $a \in \mathbb{F}$ we may take $h(x) = x - a$; otherwise $\deg\ h(x) \geq 2$. This **minimal polynomial** for a is *unique* if we require that it have leading coefficient $= 1$. [In fact, $I_a = \{h \in \mathbb{F}[x] : h(a) = 0\}$ is an ideal in $\mathbb{F}[x]$, hence a principal ideal, and by 7.3.8 is generated by any nonzero $f \in I_a$ of minimal degree. The generator f is unique up to a multiplied unit $u = c\mathfrak{1}$, $c \neq 0$ in $\mathbb{F}$.] Since there is a polynomial that kills a there is an $n \in \mathbb{N}$ such that the powers $\{\mathfrak{1}, a, a^2, \ldots, a^{n-1}\}$ are linearly independent over $\mathbb{F}$, while all higher powers lie in $\mathbb{F}$-span$\{\mathfrak{1}, a, \ldots, a^{n-1}\}$. In particular there is a linear dependence

$$c_0 + c_1 a + c_2 a^2 + \ldots + c_{n-1} a^{n-1} + a^n = 0 \quad \text{in } \mathbb{E} \qquad (c_k \in \mathbb{F}) \tag{29}$$

and the monic minimal polynomial for a is $h(x) = x^n + \sum_{k=0}^{n} c_k x^k$.

7.6.6 Lemma. *If λ is an algebraic element in a field extension $\mathbb{E} \supseteq \mathbb{F}$, its minimal polynomial $f(x) \in \mathbb{F}[x]$ is always irreducible.*

Proof: If not, we would have $h(x) = f(x) \cdot g(x)$ with $f, g \in \mathbb{F}[x]$ and $\deg(f), \deg(g)$ are both $< n = \deg(h)$. Since $0 = h(a) = f(a) \cdot g(a)$ at least one of the factors is zero, say $f(a) = 0$. Then $f \in I_a$, but $\deg(f)$ is lower than the minimum degree of elements in $I_a = (h)$. Contradiction. $\square$

7.6.7 Exercise. Show that $\mathbb{C} \supseteq \mathbb{Q}$ and $\mathbb{R} \supseteq \mathbb{Q}$ cannot be algebraic extensions of $\mathbb{Q}$ by exhibiting explicit elements that are not algebraic over $\mathbb{Q}$. Explain why $\mathbb{C}$ *is* an algebraic extension of $\mathbb{R}$. $\square$

7.6.8 Exercise. Explain why the fraction field $\mathbb{E} = \mathrm{Frac}\big(\mathbb{F}[x]\big) \supseteq \mathbb{F}[x] \supseteq \mathbb{F}$ is generated as a field by the elements of $\mathbb{F}$ and the single element $\lambda = \left[\frac{x}{1}\right]$ in $\mathbb{E}$, so $\mathbb{E} = \mathbb{F}(\lambda)$ in the notation set forth in 7.6.4. Explain why there cannot be a nontrivial polynomial relation

$$0 = \sum_{i=0}^{m} c_i \lambda^i \quad \text{in } \mathbb{E}$$

with $m < \infty$, coefficients $c_i \in \mathbb{F}$, and $c_m \neq 0$. $\square$

Thus the generator $\lambda \in \mathbb{E}$ is transcendental over $\mathbb{F}$ and $[\,\mathbb{E} : \mathbb{F}] = [\,\mathbb{F}(\lambda) : \mathbb{F}] = \infty$. By abuse of notation the fraction field $\mathbb{E}$ is generally denoted $\mathbb{F}(x)$ instead of $\mathbb{F}(\lambda)$.

Adjoining New Roots of a Polynomial to the Ground Field. Let $f(x)$ be a nonconstant irreducible polynomial in $\mathbb{F}[x]$, let $(f) = f(x) \cdot \mathbb{F}[x]$ be the corresponding principal ideal, and let $\mathbb{E} = \mathbb{F}[x]/(f)$ be the quotient ring, which is a field by 7.5.4. We first observe that in a natural sense $\mathbb{E}$ "contains" the original field $\mathbb{F}$, and so can be regarded as a field extension $\mathbb{E} \supseteq \mathbb{F}$.

The quotient map $\pi : \mathbb{F}[x] \to \mathbb{E}$ is a surjective homomorphism that sends the identity $\mathfrak{1} \in \mathbb{F}[x]$ to the coset $\overline{1} = \pi(\mathfrak{1}) = \mathfrak{1} + (f)$ in the quotient ring; this is precisely the identity element $1_{\mathbb{E}}$ in the quotient field. [Obviously $(\overline{1})^2 = \overline{1}$, and the multiplicative identity in $\mathbb{E}$ is the only idempotent element in $\mathbb{E}^\times$.] The maps $c \xrightarrow{\ j\ } c\mathfrak{1} \xrightarrow{\ \pi\ } \overline{c} = c\mathfrak{1} + (f)$ are bijective unital ring isomorphisms

$$\mathbb{F} \xrightarrow{\ j\ } \tilde{\mathbb{F}} = \mathbb{F} \cdot \mathfrak{1} \xrightarrow{\ \pi\ } \overline{\mathbb{F}} = \pi(\mathbb{F} \cdot \mathfrak{1})$$

that identify the ground field $\mathbb{F}$ with a subfield $\overline{\mathbb{F}}$ contained in the quotient field $\mathbb{E} = \mathbb{F}[x]/(f)$; the identity elements match up too, with $1_{\mathbb{F}} \to \mathfrak{1} \to \overline{1} = 1_{\mathbb{E}}$.

In what follows we will abuse notation and regard $\mathbb{F}$ as an actual subfield $\mathbb{F} \subseteq \mathbb{E}$, ignoring the distinctions between $\mathbb{F}, \tilde{\mathbb{F}} = \mathbb{F} \cdot \mathfrak{1}$, and $\overline{\mathbb{F}}$. Once these identifications are made

we may regard $\mathbb{F}[x]$ as a subring $\mathbb{F}[x] \subseteq \mathbb{E}[x]$, with a common identity element $\mathbb{1}$, just as we may regard real coefficient polynomials $h \in \mathbb{R}[x]$ as complex polynomials $h \in \mathbb{C}[x]$ that happen to have real coefficients. In particular, if $h \in \mathbb{F}[x]$ its roots in $\mathbb{F}$ are a subset of its roots in $\mathbb{E}$ when we regard h as an element of $\mathbb{E}[x]$. Additional roots often arise in $\mathbb{E}$ when we make this shift in our point of view, as they did in Example 7.5.5.

7.6.9 Theorem. *Let $\mathbb{F}$ be a field, $f = \sum_{k=0}^{n} c_k x^k$ a nonconstant irreducible polynomial in $\mathbb{F}[x]$, and $\mathbb{E} = \mathbb{F}[x]/(f)$ the quotient field with quotient map $\pi : \mathbb{F}[x] \to \mathbb{E}$. Let $\lambda = \pi(x)$, the image in $\mathbb{E}$ of the polynomial $h(x) = x$. The element λ is a root in $\mathbb{E}$ of $f(x)$, which has no roots in the subfield $\mathbb{F} \subseteq \mathbb{E}$. Furthermore,*

(i) *Every element $y \in \mathbb{E}$ can be written uniquely as a linear combination*

$$y = a_0 + a_1 \cdot \lambda + \ldots + a_{n-1} \cdot \lambda^{n-1}$$

with $n = \deg(f)$ and coefficients $a_i \in \mathbb{F}$.

(ii) *Viewing $\mathbb{E}$ as a vector space over $\mathbb{F}$, $[\,\mathbb{E} : \mathbb{F}\,] = \dim_{\mathbb{F}}(\mathbb{E})$ is equal to the degree $n = \deg(f)$.*

The extension $\mathbb{E}$ is generated as a field by λ and the elements of $\mathbb{F}$, so $\mathbb{E} = \mathbb{F}(\lambda)$.

PROOF: (i) $\Rightarrow$ (ii). Uniqueness of the expansion $y = \sum_{i=0}^{n-1} a_i \lambda^i$ means that $1, \lambda, \ldots, \lambda^{n-1}$ are a vector basis for $\mathbb{E}$ over the ground field $\mathbb{F}$. Furthermore, $\lambda = \pi(x)$ is a root of $f(x)$ in $\mathbb{E}$ because the quotient map $\pi : \mathbb{F}[x] \to \mathbb{E}$ is a homomorphism of rings that kills all elements in the ideal $I = (f)$ and $\pi(c \cdot \mathbb{1}) = c \cdot 1$ for $c \in \mathbb{F}$. Thus

$$f(\lambda) \;=\; \sum_{k=0}^{n-1} c_k \cdot (\pi(x))^k \;=\; \sum_{k=0}^{n-1} \pi\big(c_k \cdot x^k\big) \;=\; \pi\Big(\sum_{k=0}^{n-1} c_k \cdot x^k\Big) \;=\; \pi\big(f(x)\big) \;=\; \overline{0} \;\;,$$

proving (ii).

To prove (i) we may as well multiply $f(x)$ by a unit to make it a monic polynomial $f(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_0$ with $c_k \in \mathbb{F}$, so that

$$\overline{0} = f(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + +c_1\lambda + c_0\mathbb{1} \qquad \text{in } \mathbb{E} \tag{30}$$

Since the quotient map is surjective $\mathbb{E}$ consists of finite linear combinations of the powers λ^k, but $\mathbb{E}$ is actually equal to $M = \mathbb{F}\text{-span}\{1, \lambda, \ldots, \lambda^{n-1}\}$. In fact, arguing recursively from (30) we show that $\lambda^{n+k} \in M$ for $k = 0, 1, 2 \ldots$. When $k = 0$ this follows directly from (30):

$$\lambda^n = -\big(c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0\big) \in M$$

At the next step, $\pi(f) = \overline{0}$ implies $\pi\big(x \cdot f(x)\big) = \overline{0}$, so that

$$\begin{aligned} \overline{0} \;=\; \pi(x \cdot f(x)) \;\;&=\;\; \lambda \cdot (\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_0) \\ &=\;\; \lambda^{n+1} + c_{n-1}\lambda^n + (\text{terms in } M) \end{aligned}$$

Thus $\lambda^{n+1} \in -c_{n-1}\lambda^n + M = M$, and so on inductively.

The elements $1, \lambda, \ldots, \lambda^{n-1}$ are also independent, hence a basis for $\mathbb{E}$ over $\mathbb{F}$. For if there exist coefficients $a_k \in \mathbb{F}$ such that $\overline{0} = a_{n-1}\lambda^{n-1} + \ldots + a_0$, then the polynomial $g(x) = \sum_{k=0}^{n-1} a_k x^k$ maps to $\overline{0}$ under the quotient map, hence $g \in \ker \pi = I = (f)$. But by 7.3.8 $\deg(f) = n > n - 1 = \deg(g)$ is the lowest degree of any nonzero element in this principal ideal, so we get a contradiction unless $g = 0$ in $\mathbb{F}[x]$. That proves independence, so $1, \lambda, \lambda^2, \ldots, \lambda^{n-1}$ is a basis and $\dim_{\mathbb{F}}(\mathbb{E}) = n$. $\square$

7.6.10 Example. To illustrate, consider the polynomial $f(x) = x^2 - 2$ in $\mathbb{Z}_5[x]$. It is easy to check that f has no roots in $\mathbb{F} = \mathbb{Z}_5$, so $f(x)$ is irreducible over $\mathbb{Z}_5$ and the quotient

ring $\mathbb{E} = \mathbb{Z}_5[x]/(x^2-2)$ is an extension of $\mathbb{Z}_5$ with dimension $[\,\mathbb{E} : \mathbb{Z}_5] = \deg(f) = 2$. The vectors $\{1, \lambda\} = \{1, \pi(x)\}$ are a basis for $\mathbb{E}$ over $\mathbb{Z}_5$ so $\mathbb{E} = \mathbb{Z}_5 \cdot 1 + \mathbb{Z}_5 \cdot \lambda$ and $\mathbb{E}$ is a finite field with $|\mathbb{E}| = 5 \cdot 5 = 25$. This is not one of the familiar finite fields $\mathbb{Z}_p$ (p a prime).

The relation $\overline{0} = \pi(x^2-2) = \lambda^2 - 2$ can be used to compute products and sums of elements in $\mathbb{E}$.

$$\begin{aligned} (a + b\cdot\lambda) + (c + d\cdot\lambda) &= (a+c) + (b+d)\cdot\lambda \\ (a + b\cdot\lambda) \cdot (c + d\cdot\lambda) &= (ac + 2bd) + (ad + bc)\cdot\lambda \end{aligned}$$

The multiplicative inverse of an element $(a + b\alpha) \neq 0$ can be computed by clearing denominators.

$$(a+b\lambda)^{-1} = \frac{1}{a+b\lambda} \cdot \frac{a-b\lambda}{a-b\lambda} = \Big(\frac{a}{a^2-2b^2}\Big) + \Big(\frac{-b}{a^2-2b^2}\Big) \cdot \lambda$$

The denominators $a^2 - 2b^2$ are nonzero in $\mathbb{Z}_5$. This is clear if $b = 0$ and otherwise we get

$$a^2 = 2b^2 \implies \Big(\frac{a}{b}\Big)^2 = 2 \ ,$$

which is impossible because there is no "$\sqrt{2}$" in $\mathbb{Z}_5$. But $\lambda^2 = 2$ and $\mathbb{E}$ is generated by λ and the elements of $\mathbb{Z}_5$, so $\mathbb{E} = \mathbb{Z}_5(\lambda)$. We can also write $\mathbb{E} = \mathbb{Z}_5(\sqrt{2})$ to indicate that we get $\mathbb{E}$ by adjoining a square root of 2 to $\mathbb{Z}_5$. $\square$

7.6.11 Example. Revisiting Example 7.5.5 we show that $\mathbb{E} = \mathbb{R}[x]/(x^2+1)$ is isomorphic to the field $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$ of complex numbers.

Discussion. The element $\lambda = \pi(x)$ in the quotient field satisfies the relation

$$\overline{0} = \pi(x^2+1) = \lambda^2 + 1$$

so it serves as a "$\sqrt{-1}$" in that field; furthermore, $[\mathbb{E} : \mathbb{R}] = \deg(x^2+1) = 2$ so $\{1, \lambda\}$ is a basis for $\mathbb{E} = \mathbb{R}\cdot 1 + \mathbb{R}\cdot\lambda$ over $\mathbb{R}$. Since $\mathbb{C}$ has $\{1, i\}$ as an $\mathbb{R}$-basis, we get an isomorphism of vector spaces over $\mathbb{R}$ via the correspondence

$$\phi : \mathbb{E} \to \mathbb{C} \qquad \text{with} \qquad \phi(a + b\lambda) = a + bi \text{ for } a, b \in \mathbb{R}$$

It remains only to check that the $(+)$ and $(\cdot)$ operations match up under the bijection ϕ. This is clear for $(+)$, and for $(\cdot)$ we have

$$\begin{aligned} \phi\Big((a+b\lambda)\cdot(c+d\lambda)\Big) &= \phi\Big((ac + \lambda^2 bd) + (ad+bc)\lambda\Big) \\ &= \phi\Big((ac-bd) + (ad+bc)\lambda\Big) \qquad (\text{since } \lambda^2 = -1) \\ &= (ac-bd) + i(ad+bc) \ = \ (a+ib)\cdot(c+id) \\ &= \phi(a+b\lambda)\cdot\phi(c+d\lambda) \end{aligned}$$

Done.

Note that ϕ sends elements of the ground field $\mathbb{R} \subseteq \mathbb{E}$ to elements of the ground field $\mathbb{R} = \mathbb{R} + i0 \subseteq \mathbb{C}$, with $\phi(x + 0\cdot\lambda) = (x + i0)$, and once these identifications are made ϕ acts as the identity map between these subfields of $\mathbb{E}$ and $\mathbb{C}$. $\square$

7.6.12 Exercise. Prove that the quotient field $\mathbb{E} = \mathbb{Q}[x]/(x^2-2) \supseteq \mathbb{Q}$ is isomorphic to the field

$$\mathbb{K} = \mathbb{Q} + \sqrt{2}\,\mathbb{Q} = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$$

equipped with the algebraic operations $(+), (\cdot)$ defined in Example 7.1.6.
Note: Since $\mathbb{E}$ is generated by the elements of $\mathbb{Q} \subseteq \mathbb{E}$ and the single element λ we have

$\mathbb{E} = \mathbb{Q}(\lambda) = \mathbb{Q}(\sqrt{2})$. If we identify the ground fields $\mathbb{Q}$ in $\mathbb{E}$ and $\mathbb{K}$, the isomorphism $\phi : \mathbb{E} \to \mathbb{K}$ becomes the identity map when restricted to $\mathbb{Q}$. □

The following useful variant of Theorem 7.6.9 describes the field $\mathbb{F}(\lambda)$ generated by an algebraic element in a field extension $\mathbb{E} \supset \mathbb{F}$.

7.6.13 Theorem. *Let $\mathbb{E} \supseteq \mathbb{F}$ be a field extension, let $\lambda \in \mathbb{E}$ be an algebraic element in $\mathbb{E}$, and let $f \in \mathbb{F}[x]$ be the minimal polynomial for λ. Then*

(a) *The minimal polynomial $f(x)$ is irreducible in $\mathbb{F}[x]$, so by* 7.5.4 *the quotient ring $\mathbb{F}[x]/(f)$ is a field.*

(b) *The subfield $\mathbb{F}(\lambda)$ generated by λ and $\mathbb{F}$ is isomorphic to the quotient field $\mathbb{F}[x]/(f)$, so that $\dim_{\mathbb{F}}(\mathbb{E}) = [\,\mathbb{F}(\lambda) : \mathbb{F}\,]$ is equal to $d = \deg(f)$.*

(c) *The elements $1, \lambda, \lambda^2, \ldots, \lambda^{d-1}$ are an $\mathbb{F}$-basis for $\mathbb{F}(\lambda)$. Furthermore, the unital* SUBRING *generated by λ,*

$$\mathbb{F}[\lambda] = \left\{ \sum_{i=0}^{m} c_i \lambda^i : c_i \in \mathbb{F},\ m < \infty \right\}$$

coincides with the subfield $\mathbb{F}(\lambda)$ in this situtation.

PROOF: Let $\epsilon_\lambda : \mathbb{F}[x] \to \mathbb{E}$ be the evaluation map, $\epsilon_\lambda(h) = h(\lambda) \in \mathbb{E}$ for $h \in \mathbb{F}[x]$. Its kernel

$$I = \ker(\epsilon_\lambda) = \{h \in \mathbb{F}[x] : h(\lambda) = 0 \text{ in } \mathbb{E}\}$$

is a principal ideal $I = (f)$ for some $f(x) \in \mathbb{F}[x]$ such that $f(\lambda) = 0$. By 7.3.8 the generator f has the minimum possible degree among nonzero elements in I, which means it is the minimal polynomial for λ. Since f is irreducible the quotient ring $\mathbb{F}[x]/(f)$ is a *field*, by 7.6.6.

The subring $\mathbb{F}[\lambda] \subseteq \mathbb{E}$ generated by λ is the $\mathbb{F}$-linear span of the powers λ^i, but because λ is algebraic over $\mathbb{F}$ only finitely many powers $1, \lambda, \ldots, \lambda^{d-1}$ can be linearly independent over $\mathbb{F}$. Then λ^d and all higher powers are $\mathbb{F}$-linear combinations of λ^i, $i \le d-1$, so the generated ring $\mathbb{F}[\lambda]$ is precisely the linear span of these powers. The (monic) minimal polynomial for λ has the form $f(x) = x^d + c_{d-1}x^{d-1} + \ldots + c_0$.

We claim that the quotient field $\mathbb{F}[x]/(f)$ is isomorphic to the generated subfield $\mathbb{F}(\lambda)$. Since $I = (f)$ is equal to $\ker(\epsilon_\lambda)$, the First Isomorphism Theorem 7.1.1 implies that $\mathbb{F}[x]/(f) = \mathbb{F}[x]/\ker(\epsilon_\lambda)$ is isomorphic to the range $\epsilon_\lambda\big(\mathbb{F}[x]\big) = \mathbb{F}[\lambda]$ of the homomorphism $\epsilon_\lambda : \mathbb{F}[x] \to \mathbb{E}$,

$$\mathbb{F}[x]/(f) \;=\; \mathbb{F}[x]/\ker(\epsilon_\lambda) \;\cong\; \epsilon_\lambda\big(\mathbb{F}[x]\big) \;=\; \mathbb{F}[\lambda] \quad , \tag{31}$$

so $\mathbb{F}[\lambda] \cong \mathbb{F}[x]/(f)$.

Finally we observe that $\mathbb{F}(\lambda) = \mathbb{F}[\lambda]$ as subsets of $\mathbb{E}$, hence no quotients of elements in $\mathbb{F}[\lambda]$ need be taken to obtain the generated subfield. In fact by (31) the subring $\mathbb{F}[\lambda]$ is already a *subfield* of $\mathbb{E}$ that contains $\lambda = \epsilon_\lambda(x)$ so $\mathbb{F}[\lambda] \supseteq \mathbb{F}(\lambda)$, while the reverse inclusion is obvious. Hence $\mathbb{F}[\lambda] = \mathbb{F}(\lambda)$, and our proof is complete. □

7.6.14 Exercise. If $\mathbb{F} \subseteq \mathbb{E}$ and $\lambda \in \mathbb{E}$ show that λ is algebraic over $\mathbb{F} \Leftrightarrow$ there is a subfield $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ that contains λ and is finite dimensional over $\mathbb{F}$. □

7.6.15 Exercise. Show that there is no $\sqrt{-1}$ in $\mathbb{Z}_7$. Then

(a) In the manner of Example 7.6.11, decribe the multiplication law in $\mathbb{E} = \mathbb{Z}_7(\lambda) = \mathbb{Z}_7(\sqrt{-1})$, with adjoined element such that $\lambda^2 = -1$.

(b) Decide whether the element $3 = 3 + \lambda \cdot 0 \in \mathbb{E}$ has a square root in $\mathbb{E}$, and list all such square roots if any exist. □

$$\begin{array}{ccc} \mathbb{F}[x] & \xrightarrow{\ \phi\ } & A \subseteq \mathbb{K}_1 = \mathbb{F}(\alpha_1) \\ \pi\downarrow & \nearrow_{\tilde{\phi}} & \\ \mathbb{F}(\lambda) = \mathbb{F}[x]/(f) = \mathbb{E} & & \end{array}$$

Figure 7.4. The maps in the proof of 7.7.1, with $\lambda = \pi(x)$. By definition, $f(\alpha_1) = 0$.

7.6.16 Exercise. Using the result of Exercise 7.6.15, prove that the generator $\lambda = \sqrt{-1}$ in $\mathbb{E} = \mathbb{Z}_7(\lambda)$ does not itself have a square root $\mu = \sqrt{\lambda}$ in $\mathbb{E}$. $\square$.

7.7 Splitting Fields for Polynomials in F[x].

Returning to our study of root adjunctions, if a polynomial $f \in \mathbb{F}[x]$ is irreducible it has no roots in $\mathbb{F}$. But the quotient construction yields a field extension $\mathbb{E} = \mathbb{F}[x]/(f) \supseteq \mathbb{F}$ containing a root $\lambda = \pi(x)$ such that $\mathbb{E} = \mathbb{F}(\lambda)$. A natural question arises: $f(x)$ might have several roots in this extension, say $\lambda = \alpha_1, \alpha_2, \ldots, \alpha_s$. By definition $\mathbb{E} = \mathbb{F}(\lambda)$; what would happen if we adjoined to $\mathbb{F}$ some other root $\alpha_i \in \mathbb{E}$ of $f(x)$? How are the generated subfields $\mathbb{F}(\alpha_i) \subseteq \mathbb{E}$ related?

7.7.1 Theorem. *Let $f \in \mathbb{F}[x]$ be a nonconstant irreducible polynomial and for $i = 1, 2$ let $\mathbb{K}_i = \mathbb{F}(\alpha_i) \supseteq \mathbb{F}$ be extension fields generated by $\mathbb{F}$ and a single new element α_i that is a root of f in $\mathbb{K}_i$, when we regard f as an element in $\mathbb{K}_i[x] \supseteq \mathbb{F}[x]$. Then there is a unique isomorphism of fields $\phi : \mathbb{K}_1 \to \mathbb{K}_2$ such that*

$$\text{(i) } \phi|\mathbb{F} = \mathrm{id}_{\mathbb{F}} \qquad \textit{and} \qquad \text{(ii) } \phi(\alpha_1) = \phi(\alpha_2)$$

(*see Figure* 7.4).

In 7.7.1 we do not assume the extensions $\mathbb{K}_i$ lie within some common larger field $\mathbb{K}$. The following variant resolves the question "*Which root got adjoined in forming* $\mathbb{E} = \mathbb{F}[x]/(f)$?"

7.7.2 Corollary. *If $f \in \mathbb{F}[x]$ is a nonconstant irreducible polynomial, $\mathbb{E} = \mathbb{F}[x]/(f)$, and α any root of f in $\mathbb{E}$, then $\mathbb{E} = \mathbb{F}(\alpha)$.*

To prove the corollary just compare dimensions $[\mathbb{E} : \mathbb{F}]$ and $[\mathbb{F}(\alpha) : \mathbb{F}]$ with 7.7.1 in mind.

Proof (7.7.1): It suffices to show there is an isomorphism $\tilde{\phi} : \mathbb{F}[x]/(f) \to \mathbb{K}_1$ such that $\tilde{\phi}|\mathbb{F} = \mathrm{id}_{\mathbb{F}}$ and $\tilde{\phi}\big(\pi(x)\big) = \alpha_1$, where π is the quotient map $\mathbb{F}[x] \to \mathbb{F}[x]/(f)$.

The range of the substitution map $\phi : \mathbb{F}[x] \to \mathbb{K}_1$ such that

$$\phi(x) = \alpha_1 \qquad \text{and} \qquad \phi\big(h(x)\big) = h(\alpha_1) \in \mathbb{K}_1$$

is a subring $A \subseteq \mathbb{K}_1$. Then $\ker(\phi) = \{h \in \mathbb{F}[x] : h(\alpha_1) = 0\}$ while $\ker(\pi) = (f)$. Since $f(\alpha_1) = 0$ by definition we have $\ker(\pi) \subseteq \ker(\phi)$; but in fact the kernels coincide because f is irreducible over $\mathbb{F}$, $\ker(\pi) = (f)$ is a maximal ideal by 7.5.3, and $\ker(\phi)$ is not all of $\mathbb{F}[x]$.

By the First Isomorphism Theorem there is a one-to-one homomorphism of rings $\tilde{\phi}$ from $\mathbb{F}[x]/(f)$ onto A such that $\phi = \tilde{\phi} \circ \pi$, and since the quotient is a field so is its isomorphic image A, see Figure 7.4. Furthermore, if $\lambda = \pi(x)$ in the quotient we have

$$\tilde{\phi}(\lambda) = \tilde{\phi} \circ \pi(x) = \phi(x) = \alpha_1 \qquad \text{and} \qquad \tilde{\phi}|\mathbb{F} = \mathrm{id}_{\mathbb{F}}$$

when we make the natural identification $\mathbb{F} \subseteq \mathbb{K}_1$. By hypothesis $\mathbb{K}_1 = \mathbb{F}(\alpha_1)$ so $A = \mathbb{K}_1$, and $\tilde{\phi} : \mathbb{F}[x]/(f) \to \mathbb{K}_1$ is the desired isomorphism of fields.

The same discussion applies to α_2 and $\mathbb{K}_2$. Uniqueness of the induced map $\psi = \tilde{\phi}_2 \circ \tilde{\phi}_1^{-1}$ from $\mathbb{K}_1$ to $\mathbb{K}_2$ follows from properties (i) and (ii) because a homomorphism is determined by its action on generators. $\square$

Repeated adjunction of roots to the various irreducible factors of an arbitrary nonconstant polynomial $f \in \mathbb{F}[x]$ yields an extension field big enough to include all possible roots of $f(x)$.

7.7.3 Definition. *If f is a nonconstant polynomial in $\mathbb{F}[x]$, a* **splitting field** *for f is any extension $\mathbb{E} \supseteq \mathbb{F}$ such that*

(i) *$f(x)$ splits as a product $c \cdot \prod_{i=1}^{r}(x-\alpha_i)^{m_i}$ of linear factors in $\mathbb{E}[x]$, with $\alpha_i \neq \alpha_j$ if $i \neq j$ and $\sum_{i=1}^{r} m_i = \deg(f)$.*

(ii) *$\mathbb{E}$ is generated by $\mathbb{F}$ and the distinct roots of f in $\mathbb{E}$, so $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_r)$.*

7.7.4 Example. If $\mathbb{F} = \mathbb{Q}$ and $f \in \mathbb{Q}[x]$ is a nonconstant polynomial, we may regard $\mathbb{Q} \subseteq \mathbb{C}$ and obtain a splitting field by taking the *distinct* roots $\alpha_1, \ldots, \alpha_r$ of f in $\mathbb{C}$, letting $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_r)$ be the subfield of $\mathbb{C}$ generated by these roots. $\square$

7.7.5 Exercise. Show that $\mathbb{F}(a_1, a_2) = \big[\mathbb{F}(a_1)\big](a_2)$ if $\mathbb{F} \subseteq \mathbb{K}$ and $a_1, a_2 \in \mathbb{K}$. $\square$

7.7.6 Theorem (Existence of a Splitting Field). *Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x]$ a nonconstant polynomial. Then a splitting field $\mathbb{E} \supseteq \mathbb{F}$ exists and its dimension $[\mathbb{E} : \mathbb{F}]$ is finite.*

Proof: We argue by induction on $n = \deg(f)$, the result being trivial taking $\mathbb{E} = \mathbb{F}$ if $n = 1$. So, assume the result holds for all $\mathbb{F}$ and polynomials of degree $\leq n-1$. If there is a root α in $\mathbb{F}$ then $f = (x-\alpha)\cdot g(x)$ where $g \in \mathbb{F}[x]$ has lower degree, and a splitting field for $g(x)$ is obviously a splitting field for $f(x)$.

In the remaining case $f(x)$ has no roots in $\mathbb{F}$. Writing f as a product of irreducible polynomials $\prod_{i=1}^{m} f_i$, let $\mathbb{K}_1 = \mathbb{F}[x]/(f_1) \supseteq \mathbb{F}$. Then $f_1(x)$ has at least one root $\alpha_1 = \pi(x)$ in $\mathbb{K}_1 \sim \mathbb{F}$ and $[\mathbb{K}_1 : \mathbb{F}] = \deg(f_1)$. If $\alpha_1, \ldots, \alpha_r$ are the distinct roots of $f_1(x)$ in $\mathbb{K}_1$ we have $\mathbb{K}_1 = \mathbb{F}(\alpha_1, \ldots, \alpha_r)$ and

$$f_1(x) = \prod_{i=1}^{r}(x-\alpha_i)^{m_i} \cdot g(x) \qquad \text{in } \mathbb{K}_1[x]$$

where $g(x) \in \mathbb{K}_1[x]$ has no roots in $\mathbb{K}_1$. If $h(x) = \prod_{j=2}^{m} f_j(x)$ and $G(x) = g(x)h(x)$ we get a factorization of $f(x)$

$$f(x) = \prod_{i=1}^{r}(x-\alpha_i)^{m_i} \cdot G(x) \qquad \text{in } \mathbb{K}_1[x], \text{ with } \deg G < n$$

By the induction hypothesis there is a splitting field $\mathbb{E} \supseteq \mathbb{K}_1$ for $G(x)$. If $\beta_1, \ldots, \beta_\ell$ are the distinct roots of G in $\mathbb{E}$ then $\mathbb{E} = \mathbb{K}_1(\beta_1, \ldots, \beta_\ell)$ and $f(x)$ is a product of linear factors $(x-\alpha_i), (x-\beta_j)$ in $\mathbb{E}[x]$. By 7.7.5 we conclude that $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_\ell)$ so $\mathbb{E}$ is a splitting field for $f(x)$. $\square$

By 7.3.6 we also see that $[\mathbb{E} : \mathbb{F}] = [\mathbb{E} : \mathbb{K}_1] \cdot [\mathbb{K}_1 : \mathbb{F}] = [\mathbb{E} : \mathbb{K}_1] \cdot \deg(f_1)$.

Splitting Fields and the Galois Group Gal($\mathbb{E}/\mathbb{F}$). By adjoining to $\mathbb{F}$ one root λ of an irreducible polynomial in $\mathbb{F}[x]$ we obtained an essentially unique field extension $\mathbb{E} = \mathbb{F}(\lambda)$, as in Theorem 7.7.1. A similar result holds for the splitting field $\mathbb{E}_f$ of any nonconstant polynomial, but the sense in which $\mathbb{E}$ is unique is not completely straightforward, and the uniqueness proof has more moving parts than that of 7.7.1. We defer this discussion until the next Chapter, concluding this chapter with a precise description of what is

meant by "uniqueness of the splitting field" – which involves more than existence of a field automorphism between any two splitting fields of $f \in \mathbb{F}[x]$. We also indicate where all this discussion is headed.

7.7.7 Theorem (Uniqueness of Splitting Fields). *Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x]$ a nonconstant polynomial. If $\mathbb{E}_1$ and $\mathbb{E}_2$ are two splitting fields containing $\mathbb{F}$, so that*

(i) *$f(x)$ splits into linear factors in $\mathbb{E}_i[x]$ for $i = 1, 2 \ldots$*

(ii) *$\mathbb{E}_i = \mathbb{F}\big(\alpha_1^{(i)}, \ldots, \alpha_{m_i}^{(i)}\big)$ where the $\alpha_k^{(i)}$ are the distinct roots of $f(x)$ in $\mathbb{E}_i$*

then $m_1 = m_2$ and there is a UNIQUE *field isomorphism $\phi : \mathbb{E}_1 \to \mathbb{E}_2$ such that*

$$\phi|\mathbb{F} = \mathrm{id}_{\mathbb{F}} \qquad \textit{and} \qquad \phi\big(\alpha_k^{(1)}\big) = \alpha_k^{(2)} \quad \textit{for all } k \tag{32}$$

The whole point is to prove existence of the desired isomorphism ϕ in (32); uniqueness follows immediately because each $\mathbb{K}_i$ is generated by $\mathbb{F}$ and the roots $\alpha_k^{(i)}$. Note bthat we obtain s different isomorphism ϕ if we list the roots $\{\alpha_1^{(2)}, \ldots, \alpha_1^{(m)}\}$ in a different order, and likewise for the roots $\{\alpha_i^{(1)}\}$.

Observe that if $\mathbb{E}$ is a particular splitting field for $f \in \mathbb{F}[x]$ and Δ_f is the set of distinct roots in $\mathbb{E}$, the property $\phi|\mathbb{F} = \mathrm{id}_{\mathbb{F}}$ for an automorphism $\phi \in \mathrm{Aut}(\mathbb{E})$ already implies that ϕ permutes the roots in Δ_f.

7.7.8 Lemma. *Let $\mathbb{E} \supseteq \mathbb{F}$ be any splitting field for a nonconstant polynomial $f \in \mathbb{F}[x]$. If $\phi \in \mathrm{Aut}(\mathbb{E})$ fixes all points in the ground field, so $\phi|\mathbb{F} = \mathrm{id}_{\mathbb{F}}$, then $\phi(\Delta_f) = \Delta_f$.*

PROOF: If $f = \sum_i c_i x^i \in \mathbb{F}[x]$ and $f(\alpha) = 0$ for some $\alpha \in \mathbb{E}$, then

$$f\big(\phi(x)\big) = \sum_i c_i \big(\phi(\alpha)\big)^i = \sum_i \phi(c_i)\phi(\alpha^i) = \phi\Big(\sum_i c_i \alpha^i\Big) = \phi(f(\alpha)) = 0$$

in $\mathbb{E}$ so $\phi(\Delta_f) \subseteq \Delta_f$, and vice-versa for ϕ^{-1}. $\square$

Once we know the splitting field $\mathbb{E}$ of a polynomial $f \in \mathbb{F}[x]$ is essentially unique, our attention is directed to a certain group of automorphisms associated with $f(x)$, the **Galois group**

$$G_f = \mathrm{Gal}(\mathbb{E}/\mathbb{F}) = \{\phi \in \mathrm{Aut}(\mathbb{E}) : \phi|F = \mathrm{id}_{\mathbb{F}}\} \tag{33}$$

This is obviously a subgroup of $\mathrm{Aut}(\mathbb{E})$ under composition of mappings. Lemma 7.7.8 shows that each $\phi \in G_f$ permutes the set of roots $\Delta_f \subseteq \mathbb{E}$, and is completely determined by its action on those roots since $\mathbb{E} = \mathbb{F}(\alpha_1, \ldots, \alpha_r)$. [Some roots may lie within $\mathbb{F}$ since we are not assuming $f(x)$ is irreducible; $\phi(\alpha) = \alpha$ for those roots.] Therefore we obtain a natural one-to-one embedding

$$\Psi : \mathrm{Gal}(\mathbb{E}/\mathbb{F}) \to \mathrm{Per}(\Delta_f) = \begin{pmatrix} \text{all permutations on} \\ \text{the set of roots } \Delta_f \end{pmatrix} \tag{34}$$

One consquence is the fact that all Galois groups are finite since $|G_f| \leq m!$ where $m = |\Delta_f| \leq \deg f$.

Not every permutation of points in Δ_f arises in this manner: certain clusters of roots may in fact be invariant under all Galois automorphisms, for instance the roots in $\Delta_f \cap \mathbb{F}$. To put it another way, a permutation σ of the roots is the restriction of some $\phi \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ if and only if the map $\sigma : \Delta_f \to \Delta_f$ extends to an automorphism of $\mathbb{E}$ that fixes all points in $\mathbb{F}$.

Other important properties of the "Galois correspondence" (34) are less apparent and will be discussed in the next chapter. These include

- If $\mathbb{E}', \mathbb{E}$ are splitting fields for a nonconstant polynomial $f \in \mathbb{F}[x]$ there is a natural group isomorphism $\mathrm{Gal}(\mathbb{E}'/\mathbb{F}) \cong \mathrm{Gal}(\mathbb{E}/\mathbb{F})$

Although $\mathrm{Gal}(\mathbb{E}'/\mathbb{F})$ and $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ are not the same object, all such Galois groups attached to a polynomial $f(x) \in \mathbb{F}[x]$ are isomorphic as groups, and in this sense do not depend on the particular splitting field used to construct them.

- If $\Delta_f = \{\alpha_1, \ldots, \alpha_r\}$ are the distinct roots in a splitting fiels $\mathbb{E} \supseteq \mathbb{F}$ of some polynomial $f \in \mathbb{F}[x]$, identifying those permutations of Δ_f that correspond to elements of $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ can be a challenge.

We will gradually develop the tools needed for systematic calculation of Galois groups.

The Galois group turns out to be the key for deciding when a polynomial $f \in \mathbb{F}[x]$ can be "solved by radicals" – that is, its roots can be produced by a finite succession of operations involving only

$$\text{sums, products, quotients, and radicals } \sqrt[k]{\cdots}$$

of numbers constructed in previous steps, starting from the field $\mathbb{Q}$ of rational numbers at the first step. We defer a full discussion of "Galois theory," and for the moment merely note that *solvability* of the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$, in the sense of commutator subgroups (see Section 6.4), is directly related to the issue of solvability by radicals.

> THEOREM (GALOIS). *A nonconstant polynomial $f \in \mathbb{Q}[x]$ is solvable by radicals if and only if the Galois group $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ associated with a splitting field $\mathbb{E}_f$ is a solvable group in the sense of Definition 6.4.5.*

We close this chapter by noting that

> THEOREM. EVERY *nonconstant polynomial $f \in \mathbb{Q}[x]$ of degree $\deg f \leq 4$ is solvable by radicals.*

For $\deg f = 2$ we have the quadratic formula: if $a \neq 0$ then

$$f = ax^2 + bx + c = 0 \;\Rightarrow\; \text{roots } = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

For $\deg f = 3$ there is a finite algorithm that includes "IF ... THEN ..., ELSE ..." branching statements, and similarly if $\deg f = 4$. *But if* $\deg f \geq 5$ *there are polynomials over $\mathbb{Q}$ that cannot be solved by radicals.* This is a consequence of the fact that the permutation group $S_5 = \mathrm{Per}\{1, 2, \ldots, 5\}$ is *not* a solvable group. Once you recall the definition of "solvable group" this follows easily from the fact that the only proper normal subgroup of S_5 is the group of even permutations A_5, which is a simple group whose commutator subgroup is $[A_5, A_5] = A_5$ – see Examples 5.6.8-5.6.9 and 6.4.14.