

1. Introduction to Universal Algebra: Basic Ingredients

This chapter introduces the concepts of an algebra and its type. The idea is to have an “umbrella” notion which accomodates groups, rings, modules and many other algebraic entities. Thus, the motto is to generalize familiar labels in algebra; e.g., subgroups, submodules, etc. come under the heading “subalgebra”; The term “homomorphism” will come to stand for a feature of maps which preserves whatever algebraic structure we wish to preserve.

An important object coming up is that of the word algebra, which is instrumental in developing free algebras and the concept of an equational class. The chapter concludes with congruences, the common generalization of the notion of a kernel.

The most important references are [C81] and [Gr79]; we also mention [P68], although the latter is hard to get. The relatively recent – and available – [BS80] is also a good introduction.

1.1 Algebras.

Definition 1.1.1. An n -ary operation on a set A is a function from the n -fold cartesian product $A^n = A \times \cdots \times A$ (n times) into A . “Operation” in these notes means “ n -ary operation” for a suitable non-negative integer n . An operation will typically be designated by a small greek letter, say ϕ , and its action on an n -tuple (a_1, a_2, \dots, a_n) by $\phi(a_1, a_2, \dots, a_n)$.

Here are some examples which should be familiar:

Definition & Remarks 1.1.2. (a) Consider the multiplication in any group, multiplication in any ring, or the operation in a lattice which computes the supremum or else the infimum of any two elements. These are examples of 2-ary or *binary* operations.

(b) Consider a group, and the map $a \mapsto a^{-1}$ in the group. This is an example of a 1-ary or *unary* operation. So is the complementation in a boolean algebra. Another example of a unary operation: suppose that M is a left R -module (where R is any ring with identity); pick $r \in R$, and look at the function $x \mapsto rx$ on M .

(c) The case $n = 0$; what is a *zero-ary* or *nullary* operation? Reflect that $|A^0| = 1$, and thus the nullary operations, being maps from a singleton into A , are simply designated elements of the set A . For example, the identity of a multiplicative group G , say e , can be so designated as a nullary operation, when one talks about a group. Likewise, in a boolean algebra one may designate two special elements as nullary operations, the least element 0 and the largest element 1.

Remark 1.1.3. The division in a field is not an operation in the sense we are using the name. The operation $(a, b) \mapsto ab^{-1}$ is defined provided $b \neq 0$. This is an example of a *partial operation*: one which is defined on a *proper* subset of the cartesian product. We shall not deal with partial operations, although there is plenty in the literature on them. (See [Gr79].)

In this context we also exclude infinitary operations. Without getting technical, this means excluding infinite sums, such as of series or sequences or nets, and also infinite suprema and infima.

Definition & Remarks 1.1.4. (a) An *algebra* is a set, together with a certain number of operations (possibly infinitely many). The operations that are agreed upon are referred to as the *type* of the algebra. This is imprecise, for the moment. It is one of the rare times when it is best to proceed intuitively towards a rigorous formulation.

Thus a group $(G; 1, (\)^{-1}, \cdot)$, with one binary operation, one unary operation and one nullary operation can be said to have the *type* $(1, (\)^{-1}, \cdot)$. Often the operations themselves are “identified” with the cardinalities of the sets of n -ary operations under consideration. Thus, the type of group theory can also be viewed as the triple $(1,1,1)$, meaning that a group has 1 nullary, 1 unary and 1 binary operation.

A ring with multiplicative identity affords a choice. One may view it as $(R; 0, 1, -(\), +, \cdot)$, by designating the multiplicative identity, in which case it has the type $(0, 1, -(\), +, \cdot)$, with 2 nullary, 1 unary and 2 binary operations. Depending upon the theory one is interested in, the identity might be ignored, and the ring R viewed as an algebra of type $(0, -(\), +, \cdot)$, or $(1,1,2)$. In commutative algebra, for example, the choice is almost always to treat a ring as an algebra of type $(2,1,2)$. As we will see, this choice forces one to deal only with homomorphisms which preserve the multiplicative identity.

Similarly, a boolean algebra most commonly is regarded as an algebra of type $(0, 1, (\)', \vee, \wedge)$, or $(2,1,2)$, where 0 and 1 denote the least and largest element of the algebra, $(\)'$ the boolean complement, while \vee and \wedge stand for the supremum and infimum, respectively. However, one might wish to “forget” that a boolean algebra is boolean, and just “remember” that it is a distributive lattice, and view it as an algebra of type (\vee, \wedge) , or $(0,0,2)$.

Let us now be more precise.

(b) For each nonnegative integer n , an n -ary symbol is a fixed function θ , the domain of which is the class of all sets, so that, for each set A , θ_A is an n -ary operation on A . A *type* is a set T of n -ary symbols (for various n 's), partitioned by

$$T = T_0 \cup T_1 \cup T_2 \cup \cdots,$$

where T_n is a set of n -ary symbols. (T_n is frequently empty.)

An algebra of type T is a pair, denoted $(A; T)$, where A is a set. T is referred to as the *operator domain* of A . Thus, one should think of the designation $(A; T)$ as associating the operations θ_A (for each $\theta \in T$) with the set A . We are choosing some operations for A , encoding them via the type T .

In practice we shall frequently identify the operator symbols with the actual operations. Indeed, until we introduce homomorphisms, no confusion can ensue from identifying the type with the corresponding sequence of cardinalities of the T_i . (If $T_k = \emptyset$, for all $k > m$, for a suitable m , then instead of listing zeroes in all positions k , for $k > m$, one simply truncates the sequence after the m -ary position.)

The reader should keep in mind that “type” implies a set of operator symbols, each having (usually) a suggestive connotation. Strictly speaking, however, when one views a ring with identity as an algebra of type $(0, 1, -(\), +, \cdot)$, and a boolean algebra as one of type $(0, 1, (\)', \vee, \wedge)$, even though the symbols have particular interpretations in each case, it could be argued that the algebras are of the same type.

Example 1.1.5. If M is a (left) module over the ring R , then the module may be viewed as an algebra of type $(1, 1 + |R|, 1)$, written as

$$(M; 0, -(\), \{ \phi_r : r \in R \}, +).$$

ϕ_r designates the left scalar multiplication by r , $\phi_r a = ra$, which is a unary operation. The reader should take note that here we have an algebra with $1 + |R|$ -many unary operations.

It bears emphasizing that to say that A is an algebra of type (t_0, t_1, \cdots) is a premeditated act. It involves a choice. Given an operator domain T , we denote by $\mathbf{Al}(T)$ the class of all algebras

of the type defined by T . In the sequel, when words such as “subalgebra”, “homomorphism” and “isomorphism” come up, they are always to be understood in the context of $\mathbf{Al}(T)$, that is to say, pertaining to algebras of the same type.

Definition 1.1.6. Let $\{A_i : i \in I\}$ be a family of algebras of type T . Put $A = \prod_{i \in I} A_i$ (the cartesian product over I of the A_i). We make A into an algebra of type T , by coordinatewise operations. Recall that the members of this cartesian product are functions from I into the disjoint union of the A_i , for which $g(i) \in A_i$.

Now, suppose that $\phi \in T_n$. If $n = 0$ we define $\phi_A \in A$ by $\phi_A(i) = \phi_A$, the nullary symbol corresponding to ϕ , in A_i . If $n \geq 1$, and a_1, a_2, \dots, a_n are in A , define

$$\phi_A(a_1, \dots, a_n)(i) = \phi_{A_i}(a_1(i), \dots, a_n(i)).$$

This defines an algebra structure on A of type T . This is the *direct product* of the algebras A_i . When no structure is mentioned explicitly on a cartesian product, one assumes that it is the direct product. We also note that, henceforth, in the context of direct products, we shall suppress subscripts on the operations.

1.2 Subalgebras. We define the working notion of a “subobject”, and begin to consider how a subobject is generated.

Definition 1.2.1. Suppose that $(A; T)$ is an algebra (in $\mathbf{Al}(T)$). A subset B of A is called a *subalgebra* of A , if for each $n = 0, 1, 2, \dots$ and each $\phi \in T_n$, and all $b_1, b_2, \dots, b_n \in B$, we have $\phi(b_1, \dots, b_n) \in B$. The set of all subalgebras of A will be denoted by $\mathbb{S}(A)$. If B is a subalgebra of A we write $B \leq A$.

Remarks 1.2.2. (a) If G is a group (viewed as an algebra of type $(1,1,1)$) then a subalgebra H *must* be nonempty, as it must contain the designated identity e . Moreover, H is closed under multiplication and inversion, and since it inherits the laws that make G a group, then H too is a group. Thus, a subalgebra (as defined here) is a subgroup (in the familiar sense). The reader ought to reflect and become convinced that the reverse is true.

(b) If a ring R with identity 1 is regarded as an algebra of type $(2,1,2)$ (that is, if 1 is designated a nullary operation), then “subalgebra” means “subring which inherits the identity”. If R is regarded as an algebra of type $(1,1,2)$, then “subalgebra” means “subring”, period!

Exercise 1.2.3. (a) Let $(A; T)$ be an algebra. The empty subset of A is a subalgebra if and only if T contains no nullary operations.

(b) The intersection of any number of subalgebras is a subalgebra.

Definition & Remarks 1.2.4. Suppose that $(A; T)$ is an algebra, and consider $\mathbb{S}(A)$. A is a subalgebra of itself, which means that $\mathbb{S}(A)$ is not void. As has already been remarked, \emptyset is a subalgebra, precisely when there are no nullary operations.

Now let $\{B_i : i \in I\}$ be any family of subalgebras containing $X \subseteq A$. By Exercise 1.2.3(b), $B = \cap_i B_i$ is a subalgebra of A , which clearly contains X . Thus, we may speak of the *least* subalgebra containing the subset X , and we call it the subalgebra *generated* by X . It will be denoted by $\langle X \rangle_T$.

In particular, let $X = T_0$, the set of all nullary operations. Since every subalgebra must contain T_0 , it is clear that $\langle T_0 \rangle_T$ is the smallest subalgebra of A .

We then have the following:

Proposition 1.2.5. *The set $\mathbb{S}(A)$ is a complete lattice with largest element A and least element $\langle T_0 \rangle_T$, and in which*

$$\bigwedge_i B_i = \bigcap_i B_i \quad \text{and} \quad \bigvee_i B_i = \langle \bigcup_i B_i \rangle_T,$$

where $\{B_i : i \in I\}$ is a set of subalgebras of A .

Exercise 1.2.6. (a) Suppose that $(A; T)$ is an algebra for which $T_k = \emptyset$, for all $k \geq 2$; (that is, so that T consists of nullary and unary operations only). Then the union of any number of subalgebras of A is a subalgebra of A .

Give an example (say, from group theory) to show that if n -ary operations are allowed, with $n \geq 2$, then the supremum in Proposition 1.2.5 need not be the union.

(b) Suppose that $\{B_i : i \in I\}$ is a set of subalgebras of A , so that I is partially ordered so that $i \leq j$ implies that $B_i \leq B_j$, and for each pair $i, j \in I$, there is a $k \geq i, j$, such that both $B_i, B_j \leq B_k$. Prove that $\bigcup_{i \in I} B_i$ is a subalgebra of A .

(c) If B is a subalgebra of A , and C is a subalgebra of B , then C is a subalgebra of A .

Exercise 1.2.7. Let G be a group with more than one element. Define a binary operation θ as follows:

$$\theta(a, b) = \begin{cases} ab^{-1}, & \text{if one of these is a power of the other;} \\ e, & \text{otherwise.} \end{cases}$$

Prove that, as an algebra of type $\{\theta\}$, in $\mathbb{S}(G)$ the union of subalgebras is a subalgebra.

1.3 The “Algebraic” Nature of Lattices. Suppose that $(A; T)$ is an algebra and that B is a subalgebra of A . We say that B is *finitely generated* if there is a finite subset F of A , for which $B = \langle F \rangle_T$.

Proposition 1.3.1. *Suppose that $(A; T)$ is an algebra and that $B \leq A$. Then B is finitely generated if and only if the following holds:*

for any family of subalgebras of A , $\{C_i : i \in I\}$, so that $B \leq \bigvee_i C_i$, there is a finite subset $I' \subseteq I$, such that $B \leq \bigvee \{C_i : i \in I'\}$.

(Note: this condition is often interpreted as a compactness condition, and it is then said that the finitely generated elements of $\mathbb{S}(A)$ are precisely the compact elements. More on this later.)

Proof. (This is also useful elsewhere.) Suppose that $B \leq A$, and that $\langle X \rangle_T = B$. Define $X_0 = X \cup T_0$; assume now that X_0, X_1, \dots, X_{k-1} have been defined so that $X_i \subseteq X_{i+1}$, for all $i = 0, 1, \dots, k-1$. Define

$$X_k = X_{k-1} \cup \{a \in A : a = \phi(a_1, \dots, a_n), \text{ for some } \phi \in T_n, \text{ and } a_i \in X_{k-1}\}.$$

Then show that $\bigcup_{k=0}^{\infty} X_k = B$. ■

Examples 1.3.2. One of the complicated things to describe, with some grace, is exactly what is contained in a subalgebra generated by a finite set F . Let us look at a few examples. We fix an algebra A and a subset $F = \{x_1, x_2, \dots, x_k\}$ of A . The objective is to give a description of $\langle F \rangle_T$. Remember that every subalgebra must contain every nullary operation.

(a) Suppose that A is an (additive) abelian group, viewed as an algebra of type $(1,1,1)$, written as $(A; 0, -(), +)$. Then

$$\langle F \rangle_T = \{m_1x_1 + m_2x_2 + \dots + m_kx_k : m_i \in \mathbb{Z}\}.$$

(\mathbb{Z} denotes the set of integers.)

(b) Let's complicate matters; now A is a (not-necessarily abelian) group, written multiplicatively. Verify that $\langle F \rangle_T$ consists of all the expressions of the form $g_1g_2 \cdots g_s$, where each g_i is one of the x_j or the inverse of one.

(c) If A is a commutative ring with identity, regarded as an algebra of type $(2,1,2)$, then $\langle F \rangle_T$ is the set of all polynomials in the x_i ; that is, all expressions of the form

$$f(x_1, x_2, \dots, x_k) = m_0 + m_1f_1 + m_2f_2 + \dots + m_sf_s,$$

where each m_i is an integer, and each f_i is a product of the form $x_1^{d_1}x_2^{d_2} \cdots x_k^{d_k}$, where each d_j is a non-negative integer.

Exercise 1.3.3. Work out expressions for the members of $\langle F \rangle_T$ when A is

- (a) a commutative ring of type $(1,1,2)$;
- (b) a not-necessarily commutative ring with identity, regarded as having type $(2,1,2)$;
- (c) a boolean algebra, with type $(2,1,2)$.

One often expresses the content of the following proposition by saying that every member of $\mathbb{S}(A)$ is the supremum of its compact elements.

Proposition 1.3.4. *Let $(A; T)$ be an algebra. In $\mathbb{S}(A)$ each element is the supremum of its finitely generated subalgebras.*

Exercise 1.3.5. Let $(A; T)$ be an algebra. The element $a \in A$ is said to be a *non-generator* if, for any $X \subseteq A$, $\langle X \cup \{a\} \rangle_T = A$ implies that $\langle X \rangle_T = A$. (Think of a non-generator this way: it is an element which can be omitted from any generating set.)

- (a) Let $\text{Fr}(A)$ denote the set of all non-generators of A . Prove that $\text{Fr}(A)$ is a subalgebra of A .
- (b) Show that $\text{Fr}(A)$ is the intersection of all maximal (proper) subalgebras of A . (If A has no maximal subalgebras, then $\text{Fr}(A) = A$; this can happen! Read on.)
- (c) Let p be a prime number, and \mathbb{Z}_{p^∞} be the multiplicative group of all p^n -th complex roots of 1, for every natural number n . (View \mathbb{Z}_{p^∞} as an abelian group, of type $(1, ()^{-1}, \cdot)$.) Show that \mathbb{Z}_{p^∞} has no proper maximal subgroups.

(d) Likewise: The additive group \mathbb{Q} of rational numbers has no proper maximal subgroups. Prove that. (Hint: if H is a proper subgroup of \mathbb{Q} , then, for each $n \in \mathbb{N}$, let

$$H_n = \{x \in \mathbb{Q} : nx \in H\}.$$

Prove that H_n is a subgroup of \mathbb{Q} , and that $H < H_n$, for some $n \geq 2$.)

Exercise 1.3.6. Suppose that $(A; T)$ is a finitely generated algebra. Prove that every proper subalgebra B is contained in a maximal proper subalgebra of A . (Hint: Zorn's Lemma! Any proper subalgebra fails to contain one of the generators.)

These "Noetherian" conditions might look familiar from the theory of commutative rings. As one can see, rings have nothing to do with what makes these conditions equivalent.

Exercise 1.3.7. Let $(A; T)$ be an algebra. Prove that the following are equivalent.

- (a) Every subalgebra of A is finitely generated.
- (b) $\mathbb{S}(A)$ satisfies the *ascending chain condition*; that is, $\mathbb{S}(A)$ has no infinite chains of the form $A_1 < A_2 < \dots$, consisting of subalgebras of A .
- (c) (The Hausdorff Maximal Principle.) Every nonempty family of proper subalgebras of A has a maximal element.

1.4 The Word Algebra.

Definition & Remarks 1.4.1. Let T be an operator domain and X be a set. $W_o(X, T)$ will stand for the set of all finite sequences of elements from $X \cup T$. (Note: the latter union is to be regarded as a disjoint union; that is to say, symbols for operations are not elements of X .)

$W_o(X, T)$ is an algebra of type T ; for if $\phi \in T_n$, and $w_1, w_2, \dots, w_n \in W_o(X, T)$, then define

$$\phi(w_1, w_2, \dots, w_n) \equiv \phi w_1 w_2 \dots w_n,$$

(the sequence obtained by concatenating the symbols). Define the *valency* $v(c)$ of $c \in X \cup T$ as follows:

$$v(c) = \begin{cases} 1, & \text{if } c \in X; \\ -n + 1, & \text{if } c \in T_n. \end{cases}$$

If $w = c_1 c_2 \dots c_k \in W_o(X, T)$, then

$$v(w) \equiv v(c_1) + \dots + v(c_k).$$

If $w = c_1 c_2 \dots c_k \in W_o(X, T)$, with each $c_i \in X \cup T$, we call each sequence $c_i c_{i+1} \dots c_k$ a *right segment* of w .

Now let $W(X, T)$ denote the subset of $W_o(X, T)$, consisting of all the sequences $w = c_1 c_2 \dots c_k$ ($c_i \in X \cup T$) with valency 1, such that each right segment has positive valency. The elements of $W(X, T)$ will be called *words*. We shall refer to X as the *alphabet*, and the members of X as the *letters* of the alphabet.

Then we have the following basic theorem.

Theorem 1.4.2. $W(X, T)$ is the subalgebra of $W_o(X, T)$ generated by X . More generally, $w \in W_o(X, T)$ is a sequence of r words if and only if $v(w) = r$, and every right segment has positive valency.

Proof. We prove the second assertion first, and then show how it implies the first. The proof proceeds by induction on the length of a sequence.

(Sufficiency) If the length of w is 1, then $r = 1$, and $w \in X \cup T$, and therefore a word. So suppose that the sufficiency is valid for all sequences of length less than m , and w is a sequence of length m , valency r , and every right segment has positive valency. Writing $w = cw'$, with $c \in X \cup T$, then either $c \in X \cup T_0$, and then one inducts in the obvious way on w' , or $c \in T_n$, with $n \geq 1$, and then w' satisfies the sufficiency hypotheses, whence w' is a sequence of $r' = r + n - 1$ words, $w' = w_1 \cdots w_{r'}$. Now, let $w^* = cw_1 \cdots w_n$; (note that $r' \geq n$.) Then $v(w^*) = 1$, and all its right segments have positive valency; hence, w^* is a word, and so w is a sequence of $1 + (r' - n) = r$ words.

(Necessity) Again, if w has length 1, and it's a sequence of r words, then it is a word ($r = 1$), and the conclusion is clear. So let us suppose that the implication holds for all strings of length less than that of w . Write $w = w_1 \cdots w_r$, as a sequence of r words. By induction $w_2 \cdots w_r$ has valency $r - 1$, and each of its right segments has positive valency. But then the same is clear of w .

Now to the proof of the first claim. We leave it to the reader to show that $W(X, T)$ is a subalgebra of $W_o(X, T)$. Since it contains X , by definition, we have that $\langle X \rangle_T \leq W(X, T)$.

Suppose then, that B is a subalgebra of $W_o(X, T)$ containing X . If w is a word of length 1, then $w \in X \cup T_0$, and so $w \in B$. So suppose that $w \in W(X, T)$, of length greater than 1. Write $w = c_1 c_2 \cdots c_k$, with each $c_i \in X \cup T$. Then $c_1 \in T_n$, and $n \geq 1$. Now $v(c_2 \cdots c_k) = 1 - (1 - n) = n$, and all the right segments have positive valencies. Thus, by the equivalence just proved,

$$c_2 c_3 \cdots c_k = w_1 w_2 \cdots w_n,$$

where each w_i is a word. Since the length of each w_i is less than that of w , each $w_i \in B$. Now, finally, B is a subalgebra, so that $w = c_1(w_1 \cdots w_n) \in B$. ■

The next proposition essentially establishes that $W(X, T)$ is “free” over the set X . Regarding the use of the term “homomorphism” the reader should either use common sense, or else read ahead.

Proposition 1.4.3. *For each function $f : X \rightarrow A$ and each algebra $(A; T)$ there is a unique homomorphism $f^* : W(X, T) \rightarrow A$, such that $f^*(x) = f(x)$, for each letter x .*

Proof. Define f^* inductively. It is already defined on letters. Define it on nullary operators in the obvious way, $f^*(t) = t$. As for n -ary operations (with $n \geq 1$), assume that f^* has been defined on all words whose lengths are less than that of w , and then use Theorem 1.4.2 to define $f^*(w)$. ■

Remark 1.4.4. *Fastforward.* Looking ahead, $W(X, T)$ will turn out to be crucial in developing the free algebras over a set X in classes of algebras, with appropriate properties (§2.4). In language we shall develop for categories, $W(X, T)$ is the free object over the set X in the class (category) $\mathbf{Al}(T)$. $W(X, T)$ is called the word algebra over the alphabet X .

1.5 Homomorphisms, Kernels and Congruences. Starting now it will become important to think of the type of an algebra as the set of operator symbols itself, rather than the sequence of their cardinalities! This presents a problem, at least in principle: for example, as we may consider a boolean algebra and a ring with identity to be of the same type (2,1,2), we must now “label” the operation in each set T_n of n -aries. In the case at hand we must therefore decide whether we regard \vee and $+$ as being the same operation, or else, say \wedge and $+$. This is determined by context, and as said before, involves a choice or a labelling.

Definition 1.5.1. Suppose that A and B are algebras of type T , and $f : A \longrightarrow B$ is a function. We say that f is a T -homomorphism (or, if the type is understood, a *homomorphism*) if, for each $\phi \in T_n$ ($n \geq 1$) and $a_1, a_2, \dots, a_n \in A$,

$$f(\phi_A(a_1, \dots, a_n)) = \phi_B(f(a_1), \dots, f(a_n));$$

if $\phi \in T_0$, then $f(\phi) = \phi$. A T -isomorphism is a T -homomorphism which is one-to-one and surjective. (Yes, and henceforth we suppress all subscripts on operations.)

It bears emphasizing that homomorphisms are always considered between algebras of the same type!

If $f : A \longrightarrow B$ is any function, then the relation

$$\ker(f) = \{ (x, y) \in A^2 : f(x) = f(y) \}$$

is an equivalence relation. If f is a T -homomorphism, then $\ker(f)$ has the following additional feature: if $\phi \in T_n$ and $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in \ker(f)$, then, if $x = \phi(x_1, \dots, x_n)$ and $y = \phi(y_1, \dots, y_n)$, it follows that $(x, y) \in \ker(f)$. Any equivalence relation which satisfies this condition is called a T -congruence. If f is a T -homomorphism, $\ker(f)$ is referred to as the *kernel congruence of f* . We shall get to the precise relationship between kernels, congruences and homomorphisms shortly. When speaking of a congruence, we shall refer to the equivalence classes as *congruence classes*.

First, let us make an observation, and then discuss some examples.

Proposition 1.5.2. For an equivalence relation σ on the algebra $(A; T)$ the following are equivalent:

- (a) σ is a T -congruence;
- (b) σ (as a subset of A^2) is a subalgebra of A^2 (taken with the direct product structure).

Examples 1.5.3. (a) Suppose that we consider the type $(0, 1, -(), +, \cdot)$ of rings with identity. A homomorphism $f : A \longrightarrow B$ as defined here is one which preserves addition, multiplication, the additive and multiplicative identities, and additive inverses. As the reader knows well enough, if a function between rings with identity preserves addition and multiplication, then the 0 and additive inverses are automatically preserved as well. This is a consequence of the laws of rings. In general, the multiplicative identity is not preserved, and this feature must be stipulated by spelling out what the type is. Similar comments apply to other algebraic structures.

(b) For the type $(0, 1, ()', \vee, \wedge)$ of boolean algebras, notice that a map $f : A \longrightarrow B$ which preserves \vee and \wedge need not preserve either the least or the largest element. For suppose that $A = B = \mathcal{P}(X)$, the power set of $X = \{a, b\}$, and consider the map $f : A \longrightarrow B$ defined by $f(\emptyset) = f(\{a\}) = \{a\}$, and $f(\{b\}) = f(X) = X$. Then f preserves union and intersection, but is not a homomorphism of the type under consideration.

On the other hand, if $f : A \longrightarrow B$ is a function between two boolean algebras which preserves \vee and complements, then one can show that (because of DeMorgan's Laws), 0, 1 and \wedge are also preserved, so that f is a (boolean) homomorphism.

To summarize, the point of these examples is to caution: identities which govern an algebraic structure of a certain type T , may actually make the algebra one of a larger type T' . On the other hand, one should be precise in defining the domain of discourse; the type determines (in terms of the homomorphisms allowed) the structure under consideration.

Definition & Remarks 1.5.4. Suppose that $(A; T)$ is an algebra, and τ is a congruence on it. Denote by A/τ the set of all the congruence classes. We now define an algebra structure of type T on A/τ . Denote the congruence class of a by $\tau[a]$.

Suppose that $\phi \in T_n$ ($n \geq 1$), and $a_1, a_2, \dots, a_n \in A$. Define

$$\phi(\tau[a_1], \tau[a_2], \dots, \tau[a_n]) = \tau[\phi(a_1, a_2, \dots, a_n)].$$

Let us verify that this n -ary operation is well-defined. Suppose that $\tau[a_i] = \tau[b_i]$, for each $i = 1, 2, \dots, n$. Since each a_i is τ -related to b_i , it follows that $\phi(a_1, a_2, \dots, a_n)$ is τ -related to $\phi(b_1, b_2, \dots, b_n)$. Thus,

$$\tau[\phi(a_1, a_2, \dots, a_n)] = \tau[\phi(b_1, b_2, \dots, b_n)].$$

For the nullary operation $\phi \in T_0$, set

$$\phi_{A/\tau} = \tau[\phi_A].$$

This defines an algebra structure on A/τ of type T . It is called the *factor algebra of of A type T, modulo τ* . Observe that under these definitions, the canonical map $\tau^* : A \rightarrow A/\tau$ by $\tau^*(a) = \tau[a]$ is a T -homomorphism.

Hot Air 1.5.5. About Kernels! The reader is familiar with the concept of “kernel” from the group and ring theories. It is the subset of the domain which a homomorphism sends to the additive identity of the target structure. More precisely, and to the point here, the kernel is one of the equivalence classes in the congruence $\ker(f)$, when $f : A \rightarrow B$ is a homomorphism.

Specifically, suppose that we are dealing with rings, of type $T = (0, -(), +, \cdot)$, and $f : A \rightarrow B$ is a homomorphism. Then the set $K_f \equiv \{ a \in A : (a, 0) \in \ker(f) \}$, is the kernel of f . Note that K_f is an ideal of the ring A . Conversely, suppose that I is an ideal of A ; by defining

$$\sigma_I = \{ (x, y) \in A^2 : x - y \in I \},$$

the result is a congruence, and the factor algebra A/σ_I , as defined in 1.5.4 is none other than the factor ring A/I of lore. Observe as well that the assignment $I \mapsto \sigma_I$ is a one-to-one correspondence between the set of ideals and the set of congruences on A ; it preserves the lattice operations of these two sets.

Similar comments apply to groups, modules and some other structures, where the underlying additive group structure plays a prominent role. We shall revisit this kind of situation later, after a discussion of functors and some category theory. For now, simply observe the connections with familiar settings, but be advised of two important points: first, kernel congruences are not necessarily associated with subsets of the domain (and even when they are the association is less than one would like; see the next exercise); on the other hand, the set of all congruences on an algebra of type T is, by Proposition 1.5.2 identical with the lattice $\mathbb{S}(A^2)$, where A^2 has the direct product structure.

If $(A; T)$ is an algebra, let $\mathcal{C}(A)$ denote the lattice of all congruences on A . It should be said, for the sake of emphasis, that the infimum in $\mathcal{C}(A)$ is set-theoretic intersection, whereas the supremum of a family of congruences $(\sigma_i)_{i \in I}$, $\sigma = \vee_i \sigma_i$ is simply the least congruence on A which contains all the σ_i .

Exercise 1.5.6. This concerns the type $(0, \vee, \wedge)$, with cardinalities $(1, 0, 2)$, and distributive lattices with least element 0. Suppose that A is a distributive lattice (with least element 0). An *ideal* of A is a nonempty subset J of A , which is closed under \vee , and so that if $b \leq a \in J$, then $b \in J$.

Prove the following:

- (a) If σ is a congruence on A , then $\sigma[0]$ is an ideal of A .

- (b) Find a distributive lattice with least element, and a pair of distinct congruences σ_1 and σ_2 , such that $\sigma_1[0] = \sigma_2[0]$. (Which means that one can form distinct factor algebras with the same ideal!)

Recall that a *boolean ring* R is a ring in which every element is idempotent ($x^2 = x$, for each $x \in R$). Recall that if R is a boolean ring, then it is necessarily commutative and of characteristic 2.

Exercise 1.5.7. Suppose that $(A; 0, 1, ()', \vee, \wedge)$ is a boolean algebra. Define

$$a + b = (a \wedge b') \vee (a' \wedge b), \quad \text{and} \quad a \cdot b = a \wedge b, \quad \text{and} \quad -a = a.$$

- (a) Prove that $(A; 0, 1, -(), +, \cdot)$ is a boolean ring with 1.
 (b) Conversely, suppose that $(A; 0, 1, -(), +, \cdot)$ is a boolean ring with identity. Define

$$a \vee b = a + b + ab \quad \text{and} \quad a \wedge b = ab, \quad \text{and} \quad a' = 1 + a.$$

Show that $(A; 0, 1, ()', \vee, \wedge)$ is a boolean algebra.

- (c) Suppose that A is a boolean ring with 1. Show that a subset J of A is an ideal in the ring sense if and only if it is an ideal in the sense of Exercise 1.5.6, relative to the associated boolean algebra structure.

Exercise 1.5.8. If $f : A \rightarrow B$ is any T -homomorphism between algebras of the same type T , then show that the image $f(A)$ of f is a subalgebra of B .

Exercise 1.5.9. Use De Morgan's Laws in boolean algebra to show that a boolean algebra B , with type $(0, 1, ()', \vee, \wedge)$ is isomorphic to itself but with type $(1, 0, ()', \wedge, \vee)$. (The point of this exercise being, to reinforce the notion that the choice of a type is a premeditated act. The concept of isomorphism is defined in §2.1.)

Exercise 1.5.10. If $\{A_i : i \in I\}$ is a set of algebras of type T , show that the direct product A is characterized by being the only algebra structure of type T on the cartesian product, which makes all projection maps T -homomorphisms.

1.6 On Boolean Algebras. In this section we assume the type $(0, 1, ()', \vee, \wedge)$ of boolean algebras. The objective is to prove that every boolean algebra can be embedded as a subalgebra of a power set. We begin with some preliminaries and notation.

Definition & Remarks 1.6.1. As before let $\mathcal{P}(X)$ stand for the power set of the set X ; this is, for the record, the collection of all the subsets of X . On the other hand, we denote by $\mathbf{2}$ the least boolean algebra, $\mathbf{2} = \{0, 1\}$, in which $0 < 1$. Now, if X is any set, then $\mathbf{2}^X$ stands for the set of all functions of X into $\mathbf{2}$. One can also look at $\mathbf{2}^X$ as the cartesian product of X copies of $\mathbf{2}$.

Let us assume that $\mathbf{2}^X$ has the direct product structure of boolean algebras. Thus, we have for each $x \in X$,

$$(f \vee g)(x) = \max\{f(x), g(x)\} \quad \text{and} \quad f'(x) = f(x)'$$

On $\mathcal{P}(X)$ the boolean operations are the familiar ones of set-theoretic union, intersection and complementation.

Finally, if $E \subseteq X$ then $\chi_E \in \mathbf{2}^X$ denotes the *characteristic function* of E , defined by

$$\chi_E(x) = \begin{cases} 1, & \text{if } x \in E; \\ 0, & \text{otherwise.} \end{cases}$$

We are now set up for the following proposition.

Proposition 1.6.2. *The function $E \mapsto \chi_E$, from $\mathcal{P}(X) \longrightarrow \mathbf{2}^X$, is an isomorphism of boolean algebras.*

Proof. Let $\chi(E) \equiv \chi_E$; we verify that χ preserves suprema and complements, and leave the rest for the reader to check. Suppose that $E, F \subseteq X$. Then for each $x \in X$ we have,

$$\chi(E \cup F)(x) = \begin{cases} 0, & \text{if } x \notin E \cup F \\ 1, & \text{otherwise} \end{cases} = \max\{\chi_E(x), \chi_F(x)\} = (\chi_E \vee \chi_F)(x),$$

which shows that $\chi(E \cup F) = \chi(E) \vee \chi(F)$. As for complements, note that

$$(\chi(E))'(x) = \chi(E)(x)' = \begin{cases} 0, & \text{if } x \in E \\ 1, & \text{otherwise} \end{cases} = \chi(X \setminus E)(x).$$

■

Based upon the isomorphism of Proposition 1.6.2, we shall freely identify subsets with their characteristic functions, as indicated.

Definition 1.6.3. A *maximal ideal* of the boolean algebra B is a proper ideal of B , which is not properly contained in any proper ideal of B . (For a definition of ideal of a boolean algebra, see Exercise 1.5.6, and note also Exercise 1.5.7; it tells us that a maximal ideal is one which is a maximal ideal in the ring-theoretic sense.)

Proposition 1.6.4. *Suppose that I is a proper ideal of the boolean algebra B . Then I is maximal if and only if, for each $x \in B$, either $x \in I$ or $x' \in I$, (but not both).*

Proof. If I is maximal and $x \notin I$, then the ideal generated by I and x , denoted (I, x) , properly contains I . Now, verify that

$$(I, x) = \{b \in B : b \leq x \vee a, \text{ for some } a \in I\}.$$

Thus, $1 = x \vee a$, for some $a \in I$, and taking the infimum with x' , one obtains that $x' = 0 \vee (a \wedge x') = a \wedge x' \in I$.

The other half of the proof is left to the reader. ■

The next proposition says, effectively, that the maximal ideals of a boolean algebra are precisely the kernels of homomorphisms into $\mathbf{2}$.

Proposition 1.6.5. *Suppose that B is a boolean algebra. If f is a boolean homomorphism $B \longrightarrow \mathbf{2}$, then $M_f \equiv \{b \in B : f(b) = 0\}$ is a maximal ideal of B . Conversely, if M is a maximal ideal of B , then $g_M : B \longrightarrow \mathbf{2}$ defined by*

$$g_M(b) = \begin{cases} 0, & \text{if } b \in M, \\ 1, & \text{otherwise,} \end{cases}$$

is a homomorphism.

Proof. Use the preceding proposition. ■

Now the crucial, Zornian step, of verifying that there are “enough” maximal ideals in a boolean algebra. The next result is often referred to as the “Boolean Prime Ideal Theorem”.

Proposition 1.6.6. *Suppose that B is a boolean algebra, and $x \neq y$ in B . Then there exists a maximal ideal M , so that either $x \in M$ and $y \notin M$, or the reverse.*

Proof. Apply Zorn’s Lemma. ■

Now, putting the preceding propositions together, we have the main theorem. We say under the circumstances of Theorem 1.6.7 that the algebra B is represented as a *field of sets*.

Theorem 1.6.7. (Stone’s Representation Theorem) *Suppose that B is a boolean algebra. Then there is a set X , and a one-to-one boolean homomorphism $g : B \rightarrow 2^X$.*

Proof. Let $X = \text{Max}(B)$, the set of all maximal ideals of B . By Proposition 1.6.6, $\bigcap \text{Max}(B) = \{0\}$. Now define $g : B \rightarrow 2^X$ as follows: for each maximal ideal M , let g_M be the homomorphism of Proposition 1.6.5; now define

$$g(b)(M) \equiv g_M(b).$$

This way $g(b)$ is well-defined as an element of 2^X , because of Proposition 1.6.5. We leave it to the reader to verify that g is a homomorphism of boolean algebras. Since $\bigcap \text{Max}(B) = \{0\}$, g is one-to-one. ■

Exercise 1.6.8. Suppose that $X = \{0\} \cup \{1/n : n = 1, 2, \dots\}$ is the topological space obtained by restricting the ordinary metric topology on the real line to X . Compute the boolean algebra of clopen (closed-and-open) sets.

Exercise 1.6.9. An element $0 < s \in B$, in the boolean algebra B is said to be an *atom* provided the interval $[0, s]$ consists of 0 and s alone. Prove the following “refinement” of Stone’s Theorem.

There is a representation $g : B \rightarrow 2^X$ of B as a field of subsets of some set X such that the range $g(B)$ contains all finite sets if and only if every $b > 0$ in B exceeds an atom.

In fact, the following is true: in any representation of B as a field of sets the image of an element b is a finite set if and only if b is a finite supremum of atoms.

Boolean algebras with this property are said to be *atomic*.

Here is an example of a boolean algebra without any atoms.

Example 1.6.10. Consider $2^{\mathbb{N}}$, and let B be the subalgebra of all periodic sequences. (A sequence s is *periodic* if there is an integer k such that $s(m+k) = s(m)$, for each m .) The reader will easily verify that B is a subalgebra of A , and that it has no atoms.

Remark 1.6.11. Stone’s Representation Theorem guarantees that every boolean algebra can be viewed as a subalgebra of a power set, where the infimum and supremum are set-theoretic intersection and union, respectively. Theorem 1.6.7 does not say that we can assume that arbitrary suprema and infima agree with unions and intersections. We shall return to this point later, when we discuss Stone Duality, in §9.3.

As to atoms, there is an important family of boolean algebras without atoms, namely the free boolean algebras. The reader may look at Halmos’ very readable introduction to boolean algebra [Ha63]. After a discussion of Stone Duality, then those with some topological background may understand that the fact that free boolean algebras have no atoms comes from the well known theorem in topology which characterizes the Cantor Set as the only compact metric space which is totally disconnected and has no isolated points.

2. Introduction to Universal Algebra: Classes of Algebras

In this chapter we develop the standard isomorphism theorems, and, subsequently, material on closure operators and free algebras, needed to prove Birkhoff's Theorem on equational classes (Theorem 2.5.6).

2.1 The Homomorphism and Isomorphism Theorems.

Definition 2.1.1. A T -homomorphism f from one algebra of type T to another, such that its inverse exists is a T -isomorphism. It is a routine matter to settle that, if f is a T -isomorphism then its inverse f^{-1} is a T -homomorphism as well, and therefore also a T -isomorphism.

Thus the statement " A is isomorphic to B ", meaning that there is a T -isomorphism from the algebra A of type T to B of the same type, is an equivalence relation on $\mathbf{Al}(T)$. We shall frequently use the symbol \cong to denote the isomorphy between two algebras.

We shall use the terms *injective* and *surjective* for one-to-one and onto maps, respectively.

Here is the most basic result about homomorphisms:

Theorem 2.1.2. (Induced Homomorphism Theorem) *Suppose that $f : A \rightarrow B$ and $g : A \rightarrow C$ are T -homomorphisms between algebras of the same type T . If $\ker(f) \subseteq \ker(g)$ and f is onto B , then there is a unique T -homomorphism $g^* : B \rightarrow C$ so that $g^* \cdot f = g$.*

Proof. (Existence) Define $g^*(b) = g(a)$, when $f(a) = b$. Since f is surjective, such an a exists for each b . Now, if $f(a) = f(a') = b$, then $(a, a') \in \ker(f) \subseteq \ker(g)$, whence $g(a) = g(a')$, so that g^* is well-defined. We now verify that g^* is a T -homomorphism, checking for n -ary operations, with $n \geq 1$. We leave the nullary case to the reader. Observe, though, before going further, that $g^* \cdot f = g$ holds, by definition.

So let $\phi \in T_n$ ($n \geq 1$), and $b_1, b_2, \dots, b_n \in B$. Find (for each $i = 1, \dots, n$), an element $a_i \in A$, such that $f(a_i) = b_i$. Now, since $f(\phi(a_1, \dots, a_n)) = \phi(f(a_1), \dots, f(a_n))$, and the corresponding identity holds with g , we get:

$$\begin{aligned} g^*(\phi(b_1, \dots, b_n)) &= g^*(\phi(f(a_1), \dots, f(a_n))) \\ &= g^*(f(\phi(a_1, \dots, a_n))) \\ &= g(\phi(a_1, \dots, a_n)) \\ &= \phi(g(a_1), \dots, g(a_n)) \\ &= \phi(g^*(b_1), \dots, g^*(b_n)). \end{aligned}$$

(Uniqueness) If $h : B \rightarrow C$ is a T -homomorphism, such that $hf = g$, then, for all $b \in B$, and $a \in A$, so that $f(a) = b$, $h(b) = h(f(a)) = g(a)$. Thus, $h = g^*$. ■

Corollary 2.1.3. (First Isomorphism Theorem) *Suppose that $g : A \rightarrow C$ is a surjective homomorphism. Then there is a unique isomorphism $g^* : A/\ker(g) \rightarrow C$, for which $g^* \cdot \tau_g = g$, where τ_g is the canonical homomorphism $A \rightarrow A/\ker(g)$.*

Proof. In Theorem 2.1.2 let $f = \tau_g$. Verify that g^* is one-to-one because the two kernel congruences agree. g^* is surjective because g is. ■

For the next result, let's develop some notation. Suppose that $(A; T)$ is an algebra, and that σ is a congruence on A . Let $\mathcal{C}(A \geq \sigma)$ denote the set of all congruences on A , which contain σ .

For each congruence $\tau \in \mathcal{C}(A \geq \sigma)$, define $\tau/\sigma \in \mathcal{C}(A/\sigma)$ as follows: $(\sigma[a], \sigma[a']) \in \tau/\sigma$ if and only if $(a, a') \in \tau$. Conversely, if $\mu \in \mathcal{C}(A/\sigma)$, then define μ^c by:

$$(x, y) \in \mu^c \Leftrightarrow (\sigma[x], \sigma[y]) \in \mu.$$

Theorem 2.1.4. (Congruence Correspondence Theorem) *With the notation just developed, the maps $\tau \longrightarrow \tau/\sigma$ and $\mu \longrightarrow \mu^c$ are mutually inverse bijections between $\mathcal{C}(A \geq \sigma)$ and $\mathcal{C}(A/\sigma)$, which preserve inclusion. Thus, $\mathcal{C}(A \geq \sigma)$ and $\mathcal{C}(A/\sigma)$ are lattice-isomorphic.*

Proof. The trick is to show that τ/σ is well-defined; this is where the assumption that $\tau \geq \sigma$ comes into play. So suppose that $(a, a') \in \tau$. What must be shown is that, if $(a, b), (a', b') \in \sigma$, then $(b, b') \in \tau$. But this is obvious, since $\sigma \leq \tau$, and τ is both symmetric and transitive.

Now, suppose that $\tau \in \mathcal{C}(A \geq \sigma)$; then $(x, y) \in (\tau/\sigma)^c$ precisely when $(\sigma[x], \sigma[y]) \in \tau/\sigma$, which holds exactly when $(x, y) \in \tau$, proving that $\tau = (\tau/\sigma)^c$.

It remains to verify that $\mu^c/\sigma = \mu$, and that τ/σ and μ^c are congruences, as claimed. We leave this to the reader. That these maps preserve inclusion is obvious. Then, it is also evident that if a bijection between two lattices is order-preserving, it is a lattice-isomorphism; that is to say, it preserves all suprema and infima. ■

Theorem 2.1.5. (Second Isomorphism Theorem) *Let $(A; T)$ be an algebra, and σ and τ be congruences on A , with $\sigma \leq \tau$. Then the map $g : A/\sigma \longrightarrow A/\tau$ by $g(\sigma[x]) = \tau[x]$, is a surjective T -homomorphism, and $\ker(g) = \tau/\sigma$. It follows that*

$$(A/\sigma)/(\tau/\sigma) \cong A/\tau.$$

Proof. Exercise! ■

Theorem 2.1.6. (Third Isomorphism Theorem) *Suppose that $(A; T)$ is an algebra, B a subalgebra of A , and $\sigma \in \mathcal{C}(A)$. Let*

$$B' = \{ \sigma[b] : b \in B \}.$$

Then $B/(\sigma \cap B^2) \cong B'$.

Proof. Note first that $\sigma \cap B^2$ simply denotes the restriction of the relation σ to pairs from B . Observe as well that B' , being the image of a subalgebra of A , is a subalgebra of A/σ .

Now, we sketch the argument: define $g : B \longrightarrow B'$ by $g(b) = \sigma[b]$. Verify that this is a T -homomorphism, which is onto B' (by definition of B'). Its kernel congruence is $\sigma \cap B^2$. Then apply the First Isomorphism Theorem. ■

Remark 2.1.7. A nonidiotic exercise: to reconcile these isomorphism theorems with their special counterparts in group, ring and module theory!!! (Note that in the first two cases "congruence" corresponds with "normal subgroup" and "twosided ideal", respectively.)

2.2 Equational Classes.

Hot Air 2.2.1. *On Equations.* Here we introduce the so-called equational classes, and the closure operators that are related to them. Before developing the formalism, it should be pointed out that the kind of equation under consideration will be one with universal quantifiers:

$$\forall x, y \ \& \ z \ \dots \ f(x, y, z, \dots) = g(x, y, z, \dots),$$

where f and g are expressions in terms of the operations admissible under the type. Here are some examples of classes and equations which this treatment will *exclude*:

- (a) Torsion abelian groups: G is abelian, and for all $g \in G$, there exists a positive integer n such that $ng = 0$.
- (b) Fields: for each $x \neq 0$ there exists a y such that $xy = 1$.
- (c) Nilpotent groups: there exists a positive integer n such that, for all $g_1, g_2, \dots, g_n \in G$, $[[[g_1, g_2] \cdots], g_n] = e$.
- (d) All finite groups. (Think about this one!)
- (e) All p -groups: fix a certain fixed prime p ; then for each $g \in G$, there exists a positive integer n such that $g^{p^n} = e$.

Let us return to the word algebra of §1.4.

Definition 2.2.2. Let $X = \{x_1, x_2, \dots\}$ be a countable alphabet. This alphabet will be referred to as the *standard alphabet*, and denoted X_ω . The discussion we are about to initiate refers to the algebra $W(X_\omega, T)$, where T is a fixed type. The immediate objective is to define the concept of “law” in an algebra of type T . Observe first, that any equation only involves finitely many variables. Thus to define any reasonable concept of law for algebras in $\mathbf{Al}(T)$ it is likely that countably many letters will suffice.

Let $(A; T)$ be an algebra and $w, w' \in W(X_\omega, T)$. We say that A *satisfies the law* $w = w'$ if for each homomorphism $f : W(X_\omega, T) \rightarrow A$, $f(w) = f(w')$. (Note: think of the homomorphisms $f : W(X_\omega, T) \rightarrow A$ as substitutions by elements of A for the letters of the alphabet. Then, to say that the law $w = w'$ is satisfied by A is to say that the equation “ $w = w'$ ” holds for every substitution in A .) Synonymous with A *satisfies the law* $w = w'$ will be the phrase $w = w'$ *holds in* A .

Examples 2.2.3. (a) Consider the type of groups $(e, ()^{-1}, \cdot)$. Then to say that an algebra G of this type is a group, is to say that the following laws are satisfied:

- (i) $x_1(x_2x_3) = (x_1x_2)x_3$;
- (ii) $ex_1 = x_1$ and $x_1e = x_1$;
- (iii) $x_1x_1^{-1} = e$ and $x_1^{-1}x_1 = e$.

(b) Relative to the same type, an abelian group is an algebra which satisfies (i) through (iii) in (a), and also

- (iv) $x_1x_2 = x_2x_1$.

(c) Staying with the type of groups, an elementary p -group (where p is a prime number) is an algebra which satisfies (i) through (iii) in (a) plus: $x_1^p = e$. (A note about exponentiation in the word algebra. In this type, x^n should be defined recursively: first, $x^1 \equiv x$, and $x^{n+1} \equiv x^n x$. This involves a choice. Of course, once associativity is postulated, the law $x_1^p = e$ is “equivalent” to a similar law involving exponents, defined according to some other convention. The meaning of equivalence of laws will be clarified shortly.

(d) Consider the type of rings with 1: $(0, 1, -(), +, \cdot)$. To say that an algebra is a ring is to say that the laws in (b) hold with respect to $(0, -(), +)$ and also

- (v) $x_1(x_2x_3) = (x_1x_2)x_3$;

(vi) $x_1(x_2 + x_3) = x_1x_2 + x_1x_3$ together with its left–right dual.

Armed with (i) through (vi) an algebra which satisfies them is a ring with an extra distinguished element 1, which every subalgebra must contain, and every homomorphism must preserve.

To make it a ring with identity one adds the laws

(vii) $1x_1 = x_1$ and $x_11 = x_1$.

Exercise 2.2.4. (a) With respect to the type of boolean algebras, write down the laws of (i) lattices, (ii) distributive lattices and (iii) boolean algebras.

(b) With respect to the type of groups, $(e, (\)^{-1}, \cdot)$, define a group G to be *metabelian* if the commutator subgroup $[G, G]$ is abelian. (For those who know, these are the *solvable groups of rank 2*. Recall that $[G, G]$ is the subgroup generated by all commutators $[a, b]$.) Find a law that defines the metabelian groups.

(c) Consider the type of groups, and a fixed prime p . Consider the class of all groups which have the property that $[G, G]$ is an elementary p -group. Write a law that characterizes these groups.

Remark 2.2.5. Laws have been defined in terms of the standard alphabet. For technical reasons (which will be better understood later on), it is a good idea to point out that the concepts of “law” and those which derive from it, can be rendered in terms of any infinite alphabet.

Let Y be any infinite alphabet, T a type of algebras, and suppose that w and w' are words in the standard alphabet, so that $w = w'$ holds in the algebra $(A; T)$. Suppose that $h : X_\omega \rightarrow Y$ is any map from the standard alphabet into Y , which is one–to–one on the letters used in w and w' , and let $w(h)$ and $w'(h)$ be the words in $W(Y, T)$ which result from applying h to the letters of X_ω occurring in w and w' . If $f : W(Y, T) \rightarrow A$ is any homomorphism, and $h^* : W(X_\omega, T) \rightarrow W(Y, T)$ is the unique homomorphism induced by h – and this is guaranteed by Proposition 1.4.3 – then, because $w = w'$ holds in A , $f(h^*(w)) = f(h^*(w'))$. This says that $f(w(h)) = f(w'(h))$ and shows that the law $w(h) = w'(h)$ holds in A , relative to the alphabet Y .

All this formalism, only to express the intuitively obvious fact that if all substitutions from A make $w = w'$, when the words are written in the standard alphabet, then the same is true relative to any infinite alphabet Y , and any substitution of the letters from X_ω with distinct letters from Y .

Conversely, if $w = w'$ holds in A , where w and w' are words in the alphabet Y , then $w(g) = w'(g)$ holds in A , where $w(g)$ and $w'(g)$ are the words in $W(X_\omega, T)$ obtained by applying any map $g : Y \rightarrow X_\omega$, which is one–to–one on the letters used in w and w' to these words.

Definition 2.2.6. When we speak of a class of algebras of type T , we shall always mean a subclass of $\mathbf{Al}(T)$ which is *closed under isomorphism*; that is, if A is in the class, and B is isomorphic to A , then B is also included.

We shall say that a class \mathbf{E} of algebras of type T is *equational* if there is a set of laws \mathcal{L} (which may be infinite) so that the algebra $(A; T) \in \mathbf{E}$ if and only if A satisfies all the laws in \mathcal{L} .

There are various synonyms in the literature for an equational class; here are some of them: *equationally definable class*; *primitive class* (in the Russian literature, such as in the books of Kurosh); *variety*. Indeed, the term “variety” has been the most commonly used. However, as that word has a very particular meaning in algebraic geometry, it is probably a good idea to avoid it in this context.

2.3 Closure Operators. We begin with the introduction of a number of conditions which are of interest when considering classes of algebra of a certain type. Throughout, suppose that \mathbf{C} is a class of algebras of type T .

Definition & Remarks 2.3.1. (a) We are interested in when the feature of being in \mathbf{C} is preserved under passage through certain operations.

- (S) *Closure under Subalgebras.* We say that \mathbf{C} is *closed under subalgebras* if, whenever $A \in \mathbf{C}$ and $B \leq A$, then $B \in \mathbf{C}$.
- (P) *Closure under Products.* \mathbf{C} is *closed under products* if, whenever $A = \prod_{i \in I} A_i$, and each $A_i \in \mathbf{C}$, then $A \in \mathbf{C}$.
- (Q) *Closure under Homomorphic Images.* \mathbf{C} is *closed under homomorphic images* if for each surjective homomorphism $f : A \rightarrow B$, whenever $A \in \mathbf{C}$, then also $B \in \mathbf{C}$.

There is one other related closure condition which we shall develop in the sequel. First let us introduce the closure operators which accompany the above.

$$S\mathbf{C} \equiv \{ A \in \mathbf{Al}(T) : A \cong A' \leq B \in \mathbf{C} \}.$$

$$P\mathbf{C} \equiv \{ A \in \mathbf{Al}(T) : A \cong A' = \prod_i A_i \text{ with each } A_i \in \mathbf{C} \}.$$

$$Q\mathbf{C} \equiv \{ A \in \mathbf{Al}(T) : A \cong A' \text{ and there is a surjective } T\text{-homomorphism } B \xrightarrow{f} A' \text{ with } B \in \mathbf{C} \}.$$

(b) \mathbf{C} is closed under subalgebras (resp. products, resp. homomorphic images) if and only if $S\mathbf{C} \subseteq \mathbf{C}$ (resp. $P\mathbf{C} \subseteq \mathbf{C}$, resp. $Q\mathbf{C} \subseteq \mathbf{C}$). We shall abbreviate the label “closure under subalgebras” (resp. “products”, resp. “homomorphic images”) as *S-closure* (resp. *P-closure*, resp. *Q-closure*).

The following result should serve as motivation. It is the easy half of Birkhoff’s Theorem (2.5.6); the hard part is the converse.

Proposition 2.3.2. *Every equational class is S-, P- and Q-closed.*

Proof. We leave it to the reader to verify that every equational class is *S-* and *Q-*closed. It should be pointed out that to show *Q-*closure, the extension properties of Proposition 1.4.3 are required. We verify that an equational class is *P-*closed.

Suppose that \mathbf{E} is an equational class of algebras of type T , and that $\{A_i : i \in I\}$ is a family of algebras in \mathbf{E} , and that $A = \prod_{i \in I} A_i$. Let $w = w'$ be a law satisfied by each A_i . Suppose that $g : W(X, T) \rightarrow A$ is a homomorphism. Consider the composite $\pi_i \cdot g$, where $\pi_i : A \rightarrow A_i$ denotes the i -th projection; as A_i satisfies $w = w'$, $\pi_i(g(w)) = \pi_i(g(w'))$, proving that $g(w) = g(w')$, and so $A \in \mathbf{E}$. ■

The final closure operator often tidies up arguments.

Definition & Remarks 2.3.3. Suppose that A and B_i ($i \in I$) are all algebras of type T , and that there is an injective homomorphism $s : A \rightarrow \prod_{i \in I} B_i$ so that each composite $\pi_i \cdot s$ is surjective. Then we say that A is a *subdirect product* of the B_i . (For an internal characterization, see Exercise 2.3.9.)

With this notation, we say that a class \mathbf{C} is *residually closed* if whenever each $B_i \in \mathbf{C}$, then also $A \in \mathbf{C}$. $R\mathbf{C}$ stands for the class of all subdirect products of algebras in \mathbf{C} . We abbreviate “residually closed” as *R-closed*.

Observe that a class which is *S-* and *P-*closed is necessarily *R-*closed. The converse is false; which is to say that, if \mathbf{C} is *R-*closed, then it is (obviously) *P-*closed, but not necessarily *S-*closed. (Exercise 2.3.7(b) asks for an example of this.)

Examples 2.3.4. (a) Consider the type $(0, -(), +)$. Let G_i be a collection of groups ($i \in I$), and G be their direct sum; that is to say, $G \leq \prod_{i \in I} G_i$, and it consists of those functions that are finitely nonzero: $g(i) = 0$, for all but finitely many indices. (We leave it to the reader to verify that the direct sum is indeed a subgroup of the direct product.) The thing to observe is this: for each $i \in I$ and each $g_i \in G_i$, there is a function $x \in G$ such that $\pi_i(x) = g_i$, namely, let $x(j) = 0$, if $j \neq i$, and $x(i) = g_i$. Thus, G is a subdirect product of the G_i .

(b) Here is an interesting example which is not a subdirect product. In the type $(0, 1, -(), +, \cdot)$ of commutative rings with 1, consider the ring of all real valued sequences $\mathbb{R}^{\mathbb{N}}$. Notice that $\mathbb{R}^{\mathbb{N}}$ is the direct product of countably many copies of the field \mathbb{R} . Then the subring of all sequences which have rational entries is not a subdirect product of the copies of \mathbb{R} .

All of the following are subdirect products of copies of \mathbb{R} :

- (i) The subring of all periodic sequences.
- (ii) The subring of all convergent sequences.
- (iii) The subring of all bounded sequences.

(c) Again in the type of commutative rings with 1: let \mathbb{R}^X be the ring of all real valued functions of the set X into \mathbb{R} ; this is the direct product of copies of \mathbb{R} over X as an index set. Suppose that $|X| = \infty$, and let A be the subring of all functions $f \in \mathbb{R}^X$ for which

$$|\{x \in X : f(x) \neq 0\}| < |X|.$$

Verify that A is a subdirect product of copies of \mathbb{R} .

Hot Air 2.3.5. In preparation for Birkhoff's Theorem, we will presently show that a class \mathbf{C} of algebras of type T is S -, P - and Q -closed if and only if it is R - and Q -closed. The necessity has already been remarked upon. We'll get to the sufficiency in a moment.

First, however, let us look at some examples.

Examples 2.3.6. (a) *Classes which are P - and S -closed, but not Q -closed.* In the type of abelian groups $(0, -(), +)$, consider the class of all torsion free abelian groups. In the type of rings $(0, -(), +, \cdot)$, consider the class of all rings which have no nonzero nilpotent elements. (Recall that an element $x \in R$ is *nilpotent* if $x^k = 0$, for some positive integer k .)

(b) *A class which is P - and Q -closed but not S -closed.* In the type of rings with identity, $(0, 1, -(), +, \cdot)$, consider the class of all commutative rings A with 1, with the property that for each $x \in A$ there is a $y \in A$ such that $x^2y = x$. (Such rings are called *von Neumann regular*; they are, in a sense, generalizations of boolean rings. They play a prominent role in homological algebra. Much more on that later.)

Another example of this occurrence: in the type of abelian groups $(0, -(), +)$, consider the class of all divisible abelian groups. (Recall that a group G is *divisible* if for each $g \in G$ and each positive integer n , there is an $x \in G$, such that $nx = g$.)

Exercise 2.3.7. (a) Give at least three different examples of classes which are S - and Q -closed, but not P -closed.

(b) Find an example of a class which is R - but not S -closed. (Hint: In groups, think of large groups!)

Here is one of the keys to Hall's Theorem (Theorem 2.5.10):

Proposition 2.3.8. *If a class \mathbf{C} of algebras of type T is R - and Q -closed it is also S -closed.*

Proof. Suppose that B is an algebra of type T in \mathbf{C} , and that A is a subalgebra of B . Let

$$K = \{ b \in \prod_{n \in \mathbb{N}} B_n : \text{for some } k, b(n) = b(k) \in A, \text{ for all } n \geq k \},$$

where $B_n = B$, for each positive integer n . It is easily verified that K is a subalgebra of $\prod_{n \in \mathbb{N}} B_n$, and indeed a subdirect product of the copies of B . Thus, by assumption, $K \in \mathbf{C}$. On the other hand, define $g : K \rightarrow A$ by $g(b) = b(k)$, where $b(n) = b(k)$, for all $n \geq k$. We let the reader verify that g is a homomorphism. It is onto A , since, for each $a \in A$, the constant sequence a^* ($a^*(n) = a$, for all n) lies in K , and $g(a^*) = a$. This implies that $A \in \mathbf{C}$. ■

Exercise 2.3.9. Suppose that A and B_i ($i \in I$) are algebras of the same type. Prove that A is a subdirect product of the B_i if and only if there is a family of congruences $\{ \sigma_i : i \in I \}$ on A , with trivial intersection, so that $A/\sigma_i \cong B_i$.

2.4 Free Algebras. Throughout this section, we assume, unless the contrary is specified, that \mathbf{C} is a class of algebras of type T , which is closed under subalgebras, products and homomorphic images. The objective is to prove Birkhoff's Theorem (Theorem 2.5.6); that is to say, to come up with a set of laws \mathcal{L} , so that an algebra $A \in \mathbf{C}$ if and only if A satisfies all the laws of \mathcal{L} . The Proof of Birkhoff's Theorem is concluded in the next section. The existence of free objects in \mathbf{C} is paramount in these proceedings, and this section is devoted to them.

Hot Air 2.4.1. *On Foundations.* We are about to consider the family of all the homomorphic images of a fixed algebra A which happen to lie in \mathbf{C} .

In strict foundational terms this collection isn't even a set. What is a set is the collection of isomorphism classes of homomorphic images of A . The reason? The collection of isomorphism classes, according to the First Isomorphism Theorem (Corollary 2.1.3), is in one-to-one correspondence with the collection of kernel congruences, which is a collection of subalgebras of A^2 , which is contained in the power set of A^2 . Now, if A is a set, then so are A^2 and its power set, by the Zermelo-Fraenkel Axioms of Set Theory.

The point? Well, by our standing assumption, if \mathbf{C} contains one isomorphic copy of an algebra, then it contains them all, and therefore whether a certain homomorphic image of A lies in \mathbf{C} , ought to turn strictly on properties of the associated kernel congruence.

Definition & Remarks 2.4.2. Now suppose that $(A; T)$ is an algebra, and consider the set of all congruences σ on A , for which $A/\sigma \in \mathbf{C}$. Let β_A be the intersection of all such congruences. Form the product $\prod_{\sigma} A/\sigma$, over all the congruences in question, and define $h : A \rightarrow \prod_{\sigma} A/\sigma$ canonically; that is, $h(a)(\sigma) = \sigma[a]$. One has to verify that h is a homomorphism. Also that $\ker(h) = \beta_A$, and that, for each projection map π_{σ} , $\pi_{\sigma} \cdot h : A \rightarrow A/\sigma$ is surjective. Then, by the First Isomorphism Theorem, there is an injective homomorphism

$$\hat{h} : A/\beta_A \rightarrow \prod_{\sigma} A/\sigma,$$

defined by $\hat{h}(\beta_A[a])(\sigma) = \sigma[a]$. Moreover, since the class \mathbf{C} is R -closed, $A/\beta_A \in \mathbf{C}$.

The point is that the collection of congruences σ on A , for which $A/\sigma \in \mathbf{C}$ has a least element, β_A . We call this the *verbal congruence* of A , which is a very suggestive name, the appropriateness of which will soon become apparent. **Observe, however, that, so far, only the residual closure of \mathbf{C} has been used.**

Examples 2.4.3. (a) In the type of group theory $(e, (\)^{-1}, \cdot)$, for any group G , the commutator subgroup $[G, G]$ is the verbal subgroup associated with the class of abelian groups. That is to say, it is the least normal subgroup N of G , for which G/N is abelian. The reader has to make the translation between “congruence” and “normal subgroup” for himself.

(b) Again with respect to groups, let p be a prime, and consider the class \mathbf{C} of all groups in which $g^p = e$ always holds. This is an equational class, and hence R -closed. The normal subgroup of a group G generated by all the powers x^p , is the least normal subgroup N such that $G/N \in \mathbf{C}$.

(c) In the type of rings with identity, $(0, 1, -(\), +, \cdot)$, consider the class of rings of characteristic k . (This is an equational class.) For any ring R with 1, let J_k be the ideal generated by $k1$. Then J_k is the smallest ideal of R , so that R/J_k has characteristic k .

Exercise 2.4.4. (a) (Refer to Exercise 2.2.4(b)) In the type of groups, describe the verbal subgroup associated with a group and the class of metabelian groups.

(b) (Caution: the class mentioned here is not equational!) In the type of commutative rings with 1, $(0, 1, -(\), +, \cdot)$, consider the class of commutative rings which have no nonzero nilpotent elements.

Consider commutative rings with 1, and prove the following:

- (i) The set of all nilpotent elements of A form an ideal $n(A)$.
- (ii) Show that $n(A)$ is the intersection of all the prime ideals of A . (That a prime ideal must contain $n(A)$ is easy; to complete, suppose that x is not nilpotent, and – using Zorn’s Lemma – find a prime ideal P that excludes x .)
- (iii) Show that $n(A)$ is the least ideal I , such that A/I has no nonzero nilpotent elements.
- (iv) Show that A has no nonzero nilpotent elements if and only if it is a subdirect product of integral domains.

Here is the remaining piece of preparation, describing the free objects in \mathbf{C} .

Definition & Remarks 2.4.5. \mathbf{C} is a class of algebras of type T , which is S - and R -closed. Let X be any alphabet; we work now in the word algebra $W(X, T)$. Let $\beta_{\mathbf{C}}$ stand for the verbal congruence of $W(X, T)$, and form

$$F(X, \mathbf{C}) \equiv W(X, T)/\beta_{\mathbf{C}}.$$

Let $u_X : X \rightarrow F(X, \mathbf{C})$ be the restriction of the canonical map to X : $u_X(x) = \beta_{\mathbf{C}}[x]$; (the congruence class of the letter x).

The following theorem establishes the fact that $F(X, \mathbf{C})$ is the free \mathbf{C} -algebra over X . Formally, let $F \in \mathbf{C}$; F is *free in \mathbf{C} over X* if there is a one-to-one map $u : X \rightarrow F$ with the property that, for any map $f : X \rightarrow A$ with $A \in \mathbf{C}$, there is a unique homomorphism $f^* : F \rightarrow A$ such that $f^* \cdot u = f$. We say that u is *universal* for F .

The uniqueness in the definition is of utmost importance: it establishes the uniqueness of a free object, up to isomorphism. (See Exercise 2.4.9, ahead.)

Theorem 2.4.6. *Suppose that \mathbf{C} is a class which is R - and S -closed, and contains at least one nontrivial algebra; (one whose cardinality is > 1 .) Then:*

- (a) *For each alphabet X , the map u_X is one-to-one.*
- (b) *If $f : X \rightarrow A$ is any function into the algebra $A \in \mathbf{C}$, then there is a unique homomorphism $f^* : F(X, \mathbf{C}) \rightarrow A$ so that $f^* \cdot u_X = f$.*

Proof. We prove (b) first, as it is useful in the proof of (a).

If $f : X \rightarrow A$ is a function, then by Proposition 1.4.3, there is a (unique) homomorphism $f' : W(X, T) \rightarrow A$ so that $f'(x) = f(x)$, for each letter $x \in X$. By the First Isomorphism Theorem (Corollary 2.1.3), $W(X, T)/\ker(f')$ is isomorphic to a subalgebra of A , and since \mathbf{C} is S -closed, it follows that $W(X, T)/\ker(f') \in \mathbf{C}$, whence $\beta_{\mathbf{C}} \subseteq \ker(f')$. By Theorem 2.1.2, we have a homomorphism $g : F(X, \mathbf{C}) \rightarrow A$ such that $g(\beta_{\mathbf{C}}[w]) = f'(w)$. Note that

$$g(u_X(x)) = g(\beta_{\mathbf{C}}[x]) = f'(x) = f(x),$$

and so $g = f^*$ is the desired map. The uniqueness is a routine matter, and we leave it to the reader.

The key to establishing (a) is the following claim:

- (#) For each pair of distinct letters $x, y \in X$, there is an algebra $A \in \mathbf{C}$ and a map $g : X \rightarrow A$ such that $g(x) \neq g(y)$.

Before proving (#), let us see that it does the trick: given such x and y , suppose that g is the map, the existence of which is guaranteed by (#). According to (b), there exists a homomorphism $h : F(X, \mathbf{C}) \rightarrow A$, so that $h \cdot u_X = g$. But since g distinguishes x and y , so does u_X . Since x and y were arbitrary, this shows that u_X is one-to-one.

Now the proof of (#): let B be any nontrivial algebra in \mathbf{C} . Pick any two distinct elements b_1 and b_2 of B . If x and y are distinct letters of the alphabet, define $f : X \rightarrow B$, by $f(x) = b_1$, and send all other elements to b_2 . f , so defined, satisfies (#). ■

Remark 2.4.7. Much of this discussion on free algebras can be generalized. Chapter III of [C81] is to be recommended, with caution, as the notation differs from this one. We will return to this topic in the discussion on functors and adjoints.

Definition 2.4.8. Let \mathbf{C} be any class of algebras of type T ; (without having any closure properties, necessarily). We say that the class \mathbf{C} *admits free algebras* if, for each alphabet X , there is an algebra $\Phi(X, \mathbf{C}) \in \mathbf{C}$, which is free over X in \mathbf{C} . Summarizing Theorem 2.4.6, any class of algebras of type T , which is S - and R -closed, and contains a nontrivial algebra, admits free algebras.

The next exercise is important, and especially good practice with the sort of universal conditions that define concepts such as free algebras.

Exercise 2.4.9. Suppose that \mathbf{C} is a class of algebras of type T , which is S -closed and happens to admit free algebras. Show that,

- (i) if $\Phi(X, \mathbf{C})$ is the free algebra over X , then (assuming v_X is the universal map) $v_X(X)$ generates $\Phi(X, \mathbf{C})$;
- (ii) $\Phi(X, \mathbf{C})$ is well-defined up to isomorphism, in the following precise sense: if $\Gamma(X)$ is free in \mathbf{C} over X and $v'_X \rightarrow \Gamma(X)$ is universal for $\Gamma(X)$, then there is an isomorphism $s : \Phi(X, \mathbf{C}) \rightarrow \Gamma(X)$, so that $s \cdot v_X = v'_X$.

Examples 2.4.10. (a) Let \mathbf{Gr} denote the class of groups. In the type of group theory, $(e, ()^{-1}, \cdot)$, let's try to describe the free group on the alphabet X . By Theorem 2.4.6, it is obtained from the word algebra, by factoring out the verbal congruence of groups; that is, by factoring out the congruence generated by the group laws. (We are fastforwarding a bit here, but it's important to consider some examples, even if it is less than rigorously.)

Thus, the “words” of $F(X, \mathbf{Gr})$, modulo the laws of group theory, are (uniquely) of the form

$$y_1^{d_1} y_2^{d_2} \cdots y_k^{d_k},$$

where each y_i is a letter, d_i is a nonzero integer, and $y_i \neq y_{i+1}$, for all $i = 1, \dots, k - 1$.

(b) This concerns 2-groups. Recall that an elementary 2-group is abelian. Stay in the type of groups. We are after the free elementary 2-group on the alphabet X . (Note that the elementary 2-groups form an equational class.)

Since every elementary 2-group is above all an abelian group, the verbal congruence of $W(X, T)$ defining elementary 2-groups contains the one defining abelian groups. Thus, if \mathbf{E}_2 denotes the class of all elementary 2-groups, then $F(X, \mathbf{Gr})$ maps surjectively on $F(X, \mathbf{E}_2)$ (by the Induced Homomorphism Theorem, Theorem 2.1.2).

Allowing for commutativity and the fact that in \mathbf{E}_2 $x^2 = e$ is a law, the typical element of $F(X, \mathbf{E}_2)$ is of the form $y_1 y_2 \cdots y_k$, where each y_i is a letter, and the y_i are distinct. In fact, one can easily verify that $F(X, \mathbf{E}_2)$ is none but the direct sum of X copies of the group \mathbb{Z}_2 of integers modulo 2.

Exercise 2.4.11. In the same fashion as the preceding examples work out the following:

- (a) In the type of rings with identity, $(0, 1, -(), +, \cdot)$, consider the class of all commutative rings with identity $\mathbf{CRn1}$. Show that the free commutative ring with identity on the finite set $\{x_1, x_2, \dots, x_k\}$ is the polynomial ring $\mathbb{Z}[x_1, x_2, \dots, x_k]$.
- (b) Prove that the free (unital) module over a ring R with 1, over the alphabet X , is the direct sum of X copies of the ring R . (Note: A module is *unital* if $1 \cdot m = m$, for each m in the module, where 1 is the ring identity. Refer to the type for R -modules introduced in Example 1.1.5.)

Finally, in this section, some comments about free distributive lattices

Exercise 2.4.12. (a) Prove that every finite boolean algebra B is isomorphic to a finite direct product $\mathbf{2}^k$. (Hint: Use induction: each finite boolean algebra must have an atom. Pick an atom $s \in B$, and show that

$$\{y \in B : y \leq s'\}$$

is a boolean algebra (with top element s') and fewer elements.)

- (b) Consider distributive lattices with top and bottom element, i.e., as algebras in the type $(0, 1, \vee, \wedge)$. Show that $\mathbf{2}^k$ is the free distributive lattice on the alphabet of k letters. (Hint: Let $v : \{x_1, x_2, \dots, x_k\} \rightarrow \mathbf{2}^k$, be the map defined by $v(x_i)(j) = \delta_{ij}$.)
- (c) Show that (as a boolean algebra) $\mathbf{2}^2$ is free on the singleton alphabet, but not free on the alphabet of two letters.

Note: In general, $\mathbf{2}^{2^n}$, is free as a boolean algebra over the alphabet of n letters. We shall prove this in Chapter 9, as a consequence of Stone Duality.

The last one is meant as a challenge. We shall address it and at least sketch a proof in Chapter 9.

Exercise 2.4.13. Show that $\mathbf{2}^{\mathbb{N}}$, the boolean algebra of all 0,1-sequences, is not free.

2.5 Theorems of Birkhoff and Hall. As in the previous section, \mathbf{C} is a class of algebras of type T , which (unless anything is said to the contrary) will be assumed to be S -, P - and Q -closed.

Definition 2.5.1. Suppose that $(A; T)$ is an algebra, and that σ is a congruence on A . We say that σ is *fully invariant*, if for each homomorphism $f : A \rightarrow A$, $(x, y) \in \sigma$ implies that $(f(x), f(y)) \in \sigma$.

Verify that if $\{\sigma_i : i \in I\}$ is any family of fully invariant congruences on A , then $\sigma = \bigcap_i \sigma_i$ is also fully invariant. So if S is any subset of A^2 it makes sense to speak of the least fully invariant congruence on A containing S ; it is the intersection of all the fully invariant congruences which contain S .

The lemma which follows begins to establish a connection between verbal congruences and fully invariant ones.

Lemma 2.5.2. *Relative to the class \mathbf{C} , and any algebra $(A; T)$, the verbal congruence β_A is fully invariant.*

Proof. Fix a homomorphism $f : A \rightarrow A$, and define

$$\delta(f) = \{(x, y) \in A^2 : (f(x), f(y)) \in \beta_A\}.$$

It should be clear that $\delta(f)$ is an congruence relation, because it is the kernel of $\beta_A^* \cdot f$, where β_A^* stands for the canonical map $A \rightarrow A/\beta_A$. By the Induced Homomorphism Theorem, we have a one-to-one homomorphism $g : A/\delta(f) \rightarrow A/\beta_A$, given by $g(\delta(f)[x]) = \beta_A[f(x)]$. Since $A/\beta_A \in \mathbf{C}$, and \mathbf{C} is S -closed, it follows that $\beta_A \leq \delta(f)$, which proves that β_A is fully invariant. ■

Remark 2.5.3. We point out that in Lemma 2.5.2 only S -closure and (implicitly) R -closure are employed.

Hot Air 2.5.4. *Classes vs congruences.* Suppose that \mathbf{C} is any class of algebras of type T – without any closure properties, necessarily – and let \mathbf{C}' be the set of laws in the word algebra $W(X_\omega, T)$ satisfied by all the algebras in \mathbf{C} . Conversely, suppose that θ is a set of laws in the standard alphabet, and let θ' denote the class of all algebras of type T which satisfy all the laws in θ .

We leave it to the reader to verify the following conditions:

- (a) \mathbf{C}' is a fully invariant congruence on $W(X_\omega, T)$, for any class \mathbf{C} .
- (b) θ' is an equational class, for each $\theta \subseteq W(X_\omega, T)^2$.
- (c) If $\mathbf{C}_1 \subseteq \mathbf{C}_2$ then $\mathbf{C}_2' \leq \mathbf{C}_1'$.
- (d) If $\theta_1 \leq \theta_2$ then $\theta_2' \subseteq \theta_1'$.
- (e) $\mathbf{C} \subseteq (\mathbf{C})'$, for each class \mathbf{C} of algebras.
- (f) $\theta \leq (\theta)'$, for each $\theta \subseteq W(X_\omega, T)^2$.

In addition, these properties imply that

- (g) $\theta' = ((\theta)')'$, for each subset θ of $W(X_\omega, T)^2$, while
- (h) $\mathbf{C}' = ((\mathbf{C}')')$, for every class \mathbf{C} of algebras of type T .

Note that (c), (d), (e) and (f) together say that the pair of priming maps define a *Galois connexion* between the class of all classes of algebras and the set of all subsets of $W(X_\omega, T)$. Under this Galois connexion, the two families

$$\text{EQ}(T) = \{ \mathbf{C} : \mathbf{C} = (\mathbf{C}')' \}$$

and

$$\mathcal{FI}(T) = \{ \theta \subseteq W(X_\omega, T)^2 : \theta = (\theta)'\}$$

are in a one-to-one, order-inverting correspondence. (b) tells us that the members of $\text{EQ}(T)$ are equational classes. On the other hand, if \mathbf{E} is an equational class, given by the laws Γ in the standard alphabet (so that $\mathbf{E} = \Gamma'$), then $(\mathbf{E}')' = ((\Gamma')')' = \Gamma' = \mathbf{E}$. Thus, $\text{EQ}(T)$ comprises the equational classes of type T .

On the other hand, every member of $\mathcal{FI}(T)$ is a fully invariant congruence, by (a). After the proof of Birkhoff's Theorem (Theorem 2.5.6), we prove the converse; namely, that every fully invariant congruence on $W(X_\omega, T)$ is in $\mathcal{FI}(T)$.

In the proof of Birkhoff's Theorem, which follows presently, we keep to the notation introduced in 2.5.4. There is one bit of technical nastiness left, having to do with shifting between the standard alphabet and some other infinite alphabet. We sum it up in the following exercise. Refer to the notation in 2.4.5, modified as follows: since we are about to consider different alphabets, let us agree to denote by $\beta_{\mathbf{C}}(Y)$ the verbal congruence of words in Y relative to \mathbf{C} .

Exercise 2.5.5. Let Y be any infinite alphabet, and regard X_ω as embedded in Y in a fixed way. Then, for any class \mathbf{C} :

- (i) $\beta_{\mathbf{C}}(Y) \cap W(X_\omega, T)^2 = \beta_{\mathbf{C}}(X_\omega)$.
- (ii) The equational classes $\beta_{\mathbf{C}}(X_\omega)'$ and $\beta_{\mathbf{C}}(Y)'$ – that is, the class of all algebras of type T for which the laws in $\beta_{\mathbf{C}}(Y)$ hold – coincide.

Finally, here is Birkhoff's Theorem.

Theorem 2.5.6. (Birkhoff's Theorem) *A class of algebras \mathbf{C} is equational if and only if it is S -, P - and Q -closed.*

Proof. The necessity has already been observed. Suppose then that \mathbf{C} has the three featured closure properties. If \mathbf{C} has only trivial algebras, then it is clearly equational, namely, the trivial algebra satisfies all laws! So, we suppose from now on that \mathbf{C} contains a nontrivial algebra. We prove that $\mathbf{C} = \beta_{\mathbf{C}}(X_\omega)'$.

If $A \in \mathbf{C}$ and $g : W(X_\omega, T) \rightarrow A$ is any homomorphism, then, since $W(X_\omega, T)/\ker(g)$ is isomorphic to a subalgebra of A , and \mathbf{C} is S -closed, we get that $\beta_{\mathbf{C}}(X_\omega) \leq \ker(g)$, and so A satisfies all the laws of $\beta_{\mathbf{C}}(X_\omega)$; this means that $A \in \beta_{\mathbf{C}}(X_\omega)'$. (Note that we still have not used Q -closure anywhere!)

Conversely, suppose that B is an algebra of type T , which satisfies all the laws of $\beta_{\mathbf{C}}(X_\omega)$, and therefore (by Exercise 2.5.5) the laws of $\beta_{\mathbf{C}}(Y)$, for any larger alphabet. Choose any alphabet

Y which contains B . Using the freeness of $W(Y, T)$, we see that there is a homomorphism $q : W(Y, T) \longrightarrow B$ (which is necessarily onto B) so that $q(b) = b$.

Since B satisfies all the laws of $\beta_{\mathbf{C}}(Y)$, it follows that $\beta_{\mathbf{C}}(Y) \leq \ker(q)$, which, by the Induced Homomorphism Theorem, gives rise to a surjective homomorphism $q^* : F(Y, \mathbf{C}) \longrightarrow B$. But $F(Y, \mathbf{C})$ is the free algebra in \mathbf{C} over the set Y , and so, since \mathbf{C} is Q -closed (aha!) we obtain that $B \in \mathbf{C}$. ■

Hot Air 2.5.7. We resume the discussion in 2.5.4, with the same notation.

Suppose that θ is a fully invariant congruence on $W(X_\omega, T)$; let $\mathbf{C} = \theta'$. Note the obvious, that \mathbf{C} is an equational class. Because $W(X_\omega, T)/\beta_{\mathbf{C}} \in \mathbf{C}$, we have that $\theta \leq \beta_{\mathbf{C}}$. Now, by the proof of Birkhoff's Theorem, $\mathbf{C} = \beta'_{\mathbf{C}}$, and, as observed in 2.5.4, $\mathbf{C} = (\mathbf{C}')'$. Next, if $(w, w') \in (\beta'_{\mathbf{C}})'$, then the law $w = w'$ is satisfied by every algebra in \mathbf{C} , including $W(X_\omega, T)/\beta_{\mathbf{C}}$, which is *free* in this class, according to Theorem 2.4.6. Thus, $(w, w') \in \beta_{\mathbf{C}}$, and we have that

$$\beta_{\mathbf{C}} = (\beta'_{\mathbf{C}})' = \mathbf{C}'.$$

In particular, $\theta \leq \beta_{\mathbf{C}}$.

To conclude then that θ and $\beta_{\mathbf{C}}$ are actually the same, it suffices to show that $\beta_{\mathbf{C}} \leq \theta$, and to this end it is enough to show that $W(X_\omega, T)/\theta \in \mathbf{C}$. We give a sketch. Suppose that $(w, w') \in \theta$; the goal is to show that, for any homomorphism $f : W(X_\omega, T) \longrightarrow W(X_\omega, T)/\theta$, $f(w) = f(w')$, according to the definition of satisfaction of laws (Definition 2.2.2). Now, define a new function g by letting $g(x)$ be any representative of $f(x)$, for each letter $x \in X_\omega$. Use Proposition 1.4.3 to extend g to a homomorphism g^* on $W(X_\omega, T)$. Since θ is fully invariant, conclude that $(g^*(w), g^*(w')) \in \theta$. And then, on account of the definitions of g and g^* , this means that $f(w) = f(w')$.

Thus, $\theta = \beta_{\mathbf{C}} = (\theta')'$, and this brings the discussion in 2.5.4 full circle, proving that any fully invariant congruence on $W(X_\omega, T)$ lies in $\mathcal{FI}(T)$. Note that we have also demonstrated that a congruence on the word algebra is fully invariant if and only if it's verbal.

We summarize as follows.

Theorem 2.5.8. *The maps $\theta \longrightarrow \theta'$ and $\mathbf{C} \longrightarrow \mathbf{C}'$ form a Galois connexion. Moreover*

- (a) $\mathcal{FI}(T)$ is the set of all fully invariant congruences on $W(X_\omega, T)$; this is also the set of verbal congruences on $W(X_\omega, T)$.
- (b) $\text{EQ}(T)$ is the class of all equational classes of algebras of type T .
- (c) $\theta \longrightarrow \theta'$ and $\mathbf{C} \longrightarrow \mathbf{C}'$ are mutually inverse, order-inverting bijections between $\mathcal{FI}(T)$ and $\text{EQ}(T)$. In particular, $\text{EQ}(T)$ is a set.

How to generate an equational class? Obviously, one wants to close under S , P and Q . But, in what order, and how most efficiently? Theorem 2.5.10 is the best answer. Note that the equational class generated by the class \mathbf{X} is none other than $(\mathbf{X}')'$, from 2.5.4.

We will need the following observation, which is left as an exercise.

Exercise 2.5.9. Suppose that \mathbf{X} is an R -closed class. Then $Q\mathbf{X}$ is also R -closed.

Theorem 2.5.10. (P. Hall's Theorem) *Suppose that \mathbf{X} is a class of algebras and \mathbf{C} is the equational class generated by \mathbf{X} . Then*

$$\mathbf{C} = QR\mathbf{X}.$$

In words, every algebra of \mathbf{C} is a homomorphic image of a subdirect product of algebras in \mathbf{X} .

Proof. If \mathbf{X} consists of the trivial algebra there is nothing to prove. Therefore, we assume that \mathbf{X} contains a nontrivial algebra. Now $QR\mathbf{X}$ is R -closed, by Exercise 2.5.9, also S -closed in view of Proposition 2.3.8, and evidently Q -closed. By Birkhoff's Theorem $QR\mathbf{X}$ is an equational class, which is necessarily the equational class generated by \mathbf{X} . ■

We estimate the size of $\text{EQ}(T)$.

Remarks 2.5.11. The cardinality of $W(X_\omega, T)$ is bounded by that of the (disjoint) union of X_ω and T . If the type T is finite, (i.e., if there are only finitely many operations in the type), then $W(X_\omega, T)$ is a countable set. Since every congruence is a member of the power set of $W(X_\omega, T)^2$, and the latter is also countable, there are at most \mathfrak{c} (the cardinality of the set of real numbers) equational classes.

If T is infinite, then

$$|W(X_\omega, T)| = |W(X_\omega, T)^2| \leq |T|,$$

and so

$$|\text{EQ}(T)| = |\mathcal{FI}(T)| \leq 2^{|T|}.$$

It is known, but it takes some doing to prove, that there are uncountably many – and therefore \mathfrak{c} equational classes of groups; see [Nm67].

This concludes the general discussion of universal algebra, per sé. With this background, further reading in Cohn's book ([C81]) is recommended. Some related themes will come up in the chapters on category theory. The final section of this chapter is devoted to applications of what's gone before.

2.6 Applications. We should like to distill information from the previous sections about free algebras in different settings. The presentation here follows [C81], Chapters III and IV.

For example, if \mathbf{X} is a class of algebras of a certain type T , then what exactly lies in the equational class generated by \mathbf{X} ? We have already seen in Hall's Theorem (Theorem 2.5.10) how to apply the basic closure operators most efficiently. Can we use this kind of information to produce results which predict the size of objects which are free in an equational class? Or how about generating an equational class with a minimal number of algebras? The first question can be answered in a number of ways. Let us begin.

The first result could have been stated earlier.

Proposition 2.6.1. *Suppose that \mathbf{C} is any equational class of algebras of type T . If $A \in \mathbf{C}$, then there is a free algebra F of \mathbf{C} and a surjective homomorphism $f : F \rightarrow A$.*

Proof. Letting X be any generating set of A , we have the extension of the inclusion map on A , which is a surjective homomorphism $g : W(X, T) \rightarrow A$. By definition, the verbal congruence $\beta_{\mathbf{C}}(X) \leq \ker(g)$. Therefore by the Induced Homomorphism Theorem, there is a surjective homomorphism $f : W(X, T)/\beta_{\mathbf{C}}(X) \rightarrow A$, whose domain is free in \mathbf{C} , owing to Theorem 2.4.6. ■

The discussion which follows may seem repetitive and unpleasantly analytic. The reader is urged, nonetheless, urged to consider the details with some care. This is a useful preliminary for the discussion on adjoint functors coming later.

Remark 2.6.2. As is made clear by Theorem 2.4.6, only S - and R -closure are needed to show the existence of free objects in a class \mathbf{C} . Let us go back and analyze that proof anew.

- (i) R -closure defines the verbal congruence $\beta_{\mathbf{C}}(X)$ on $W(X, T)$, and places $W(X, T)/\beta_{\mathbf{C}}(X)$ in \mathbf{C} .

(ii) S -closure makes the universal extension property work; that is, it yields (b) of Theorem 2.4.6.

Let's proceed a little differently. Let \mathbf{C} be an arbitrary class of algebras of type T , which contains a nontrivial algebra. Suppose that for each set X there is an algebra $F(X) \in \mathbf{C}$ and a mapping (not necessarily one-to-one) $v_X : X \rightarrow F(X)$ with the extension feature of Theorem 2.4.6(b) relative to all algebras in \mathbf{C} . Clearly then if B is a nontrivial algebra, and X is a generating set for B , with $|X| = \kappa$, then by extending the inclusion $j : X \subseteq B$ to $F(X)$ to a homomorphism $j^* : F(X) \rightarrow B$, it follows that v_X is one-to-one. Now suppose that Y is any set of cardinality $\geq \kappa$. Then condition (#) of the proof of Theorem 2.4.6 is satisfied for Y and the algebra $F(X)$; the reader will easily verify this. Consequently, applying the argument of that proof, it will follow that v_Y is also one-to-one.

We summarize the import of this discussion.

Proposition 2.6.3. *Suppose that \mathbf{C} is a class of algebras of type T containing a nontrivial algebra. Suppose that for each set X there is an algebra $F(X)$ in \mathbf{C} and a map $v_X : X \rightarrow F(X)$ for which the condition (b) of Theorem 2.4.6 holds. Then there is a cardinal number κ such that for each set Y for which $|Y| \geq \kappa$, $F(Y)$ is free in \mathbf{C} with v_Y universal for $F(Y)$.*

If \mathbf{C} is also S -closed then we may take $\kappa = 2$ in the above discussion, which yields the following corollary.

Corollary 2.6.4. *With the same hypotheses of Proposition 2.6.3, if \mathbf{C} is S -closed then, for each set Y , $F(Y)$ is free in \mathbf{C} with v_Y universal for $F(Y)$.*

Remark 2.6.5. Of course, if in \mathbf{C} is a class of algebras of type T which is R -closed, having a nontrivial algebra then Proposition 2.6.3 applies. If \mathbf{C} is also S -closed, we recover Theorem 2.4.6(b).

Let's revisit the question of how to generate an equational class. We need to introduce one more closure property.

Definition 2.6.6. Suppose that \mathbf{C} is a class of algebras of type T . \mathbf{C} is said to be L -closed if the following holds. Suppose that A is an algebra with an upward directed system of subalgebras $(A_\lambda)_\lambda$, such that $\lambda \leq \mu$ implies that $A_\lambda \leq A_\mu$. If $A = \cup_{\lambda \in \Lambda} A_\lambda$, and each $A_\lambda \in \mathbf{C}$ then $A \in \mathbf{C}$.

One also says that \mathbf{C} is a *local class*.

We have the following feature of equation classes. The proof is left as an exercise.

Exercise 2.6.7. If \mathbf{C} is Q - and R -closed (i.e., an equation class), then it is also L -closed.

Hint: Suppose that A and the A_λ are as in Definition 2.6.6. Form the product $B = \prod_{\lambda} A_\lambda$, and consider the subdirect product of the A_λ defined by

$$C \equiv \{ f \in B : \exists \mu, f_\lambda = f_\mu, \forall \lambda \geq \mu \}.$$

Obtain A as a homomorphic image of C .

Proposition 2.6.8. *Suppose that \mathbf{C} is a nontrivial equational class. Then any free algebra in \mathbf{C} on an infinite set generates \mathbf{C} . In symbols, if F is any free algebra of \mathbf{C} on an infinite set, then $\mathbf{C} = QR\{F\}$.*

Proof. Evidently $QR\{F\} \subseteq \mathbf{C}$. Now if A is any algebra in \mathbf{C} , it is the direct union of all its finitely generated subalgebras. Each of those is a homomorphic image of F (abusing the proof of Proposition 2.6.1, if necessary). By the preceding exercise it follows that $A \in QR\{F\}$. ■

Remarks 2.6.9. (a) Let's underscore what Proposition 2.6.8 says: Each algebra in \mathbf{C} is a homomorphic image of a subdirect product of copies of F , which is free on an infinite alphabet.

(b) A slight modification of the proof of Proposition 2.6.8 shows that, if \mathbf{C} is a nontrivial equational class, then the collection of algebras in \mathbf{C} which are free on finite alphabets generate \mathbf{C} .

In the spirit of the Proposition 2.6.8 we have the following result.

Proposition 2.6.10. *Suppose that \mathbf{X} generates the equational class \mathbf{C} of algebras of type T . Then $F \in \mathbf{C}$ is free on the alphabet X if and only if every map $f : X \rightarrow A$ can be extended to a homomorphism $f^* : F \rightarrow A$, for each $A \in \mathbf{X}$.*

Proof. By the proof of Proposition 2.6.1, there is a surjective homomorphism $f : W(X, T) \rightarrow F$. If B is any algebra in \mathbf{C} and $g : X \rightarrow B$ is any map, there is a homomorphism $g^* : W(X, T) \rightarrow B$ extending g . We are finished if we can demonstrate that $\ker(f) \leq \ker(g^*)$, on account of the Induced Homomorphism Theorem.

Now let $(w, w') \in \ker(f)$. What we prove is that $w = w'$ is a law for the class \mathbf{X} . If $h : W(X, T) \rightarrow A$ is a homomorphism, with $A \in \mathbf{X}$, we consider the restriction $h|_X$; by assumption it extends to the homomorphism $\bar{h} : F \rightarrow A$. Since $\bar{h}(f(x)) = h(x)$, for each letter $x \in X$, we have by unique extension that $\bar{h} \cdot f = h$, and it follows that $h(w) = h(w')$, as claimed. Thus, $\ker(f)$ is a set of laws for \mathbf{X} , and therefore for \mathbf{C} , which is enough to show that $\ker(f) \leq \ker(g^*)$. ■

Definition 2.6.11. We say that the algebra F of type T is *relatively free* if it is free on some alphabet X , for some equational class. Rephrasing Proposition 2.6.10, we have that F is relatively free (on X in \mathbf{C}) if and only if every map $f : X \rightarrow F$ extends to a homomorphism $f^* : F \rightarrow F$.

In Exercise 2.6.17 the reader will find an example of an algebra which is not relatively free.

When an algebra A of type T generates an equational class \mathbf{C} then Hall's Theorem tells us that each algebra in \mathbf{C} is a homomorphic image of a subdirect product of copies of A . Let's now get some more precise information about the free algebras in $QR\{A\}$. It is a consequence of Proposition 2.6.10; we leave the proof to the reader.

Proposition 2.6.12. *Suppose that A is a nontrivial algebra of type T . Let $\mathbf{C} = QR\{A\}$. For any alphabet X , consider the algebra A^X and, for each $x \in X$, the projection $\pi_x(a) = a(x)$. Let $F(X)$ stand for the subalgebra of A^{A^X} generated by the $(\pi_x)_{x \in X}$. Then $F(X)$ is free in \mathbf{C} on X .*

(Note that X and $(\pi_x)_{x \in X}$ have the same cardinality.)

Since the functions which generate $F(X)$ in the above are onto A , we have the following immediate corollary.

Corollary 2.6.13. *Suppose that A is an algebra of type T , and $\mathbf{C} = QR\{A\}$. The free algebra $F(X)$ on the alphabet X is a subdirect product of copies of A .*

In particular, we have for any equational class:

Corollary 2.6.14. *Let \mathbf{C} be an equational class. Then every free algebra of \mathbf{C} is a subdirect product of the free algebra of \mathbf{C} on the standard alphabet.*

Another easy illustration of Proposition 2.6.12 is the following. The second statement is a consequence of Proposition 2.6.1.

Corollary 2.6.15. *Suppose that A is a finite algebra of type T . In the equational class $\mathbf{C} = QR\{A\}$, every free algebra on a finite alphabet is finite. Consequently, every finitely generated algebra in \mathbf{C} is finite.*

We continue with several illustrations of the recent results.

Remarks 2.6.16. (a) Let's consider **Bool** the class of all boolean algebras (in the type of boolean algebras). By the Stone Representation Theorem (Theorem 1.6.7), the boolean algebra $\mathbf{2}$ generates **Bool**. This implies (Corollary 2.6.15) that every finitely generated boolean algebra is finite.

(b) **Dist** denotes the class of all distributive lattices with top and bottom, of type $(0, 1, \vee, \wedge)$. Exercise 2.4.12 tells us that each free algebra in **Dist** is a finite product of copies of $\mathbf{2}$. Thus, the equational class generated by $\mathbf{2}$ is the same as that generated by the distributive lattices which are free on finite alphabets. Applying the comment in 2.6.9(b), we conclude that $\mathbf{Dist} = QR\{\mathbf{2}\}$.

(c) Recall that **Gr** denotes the class of all groups. Every elementary 2-group (i.e., member of \mathbf{E}_2) is abelian, and the finitely generated elementary 2-groups are direct sums of copies of \mathbb{Z}_2 (by the Fundamental Theorem of Finitely Generated Abelian Groups). Thus, arguing as in (b), it follows that $\mathbf{E}_2 = QR\{\mathbb{Z}_2\}$.

For odd primes one needs more machinery: in §4.1 the free product will be developed. Using that notion it will be shown that every finitely generated p -group is finite.

Exercise 2.6.17. Prove that a finite abelian p -group is relatively free if and only if it is an elementary p -group; that is, a direct sum of copies of \mathbb{Z}_p .

The final major result is a theorem of Jónsson and Tarski, which gives a sufficient condition which insures that two free algebras on finite alphabets of different size are nonisomorphic. Belatedly, and because it is not, in general, a “good” definition, we define the notion of rank of a free object.

Definition & Remarks 2.6.18. If \mathbf{C} is a class of algebras of type T and F is free in \mathbf{C} on the alphabet X then F is said to be \mathbf{C} -free of rank $\kappa = |X|$.

Now if A is any algebra of type T , and $\langle X \rangle_T = A$, while X is a minimal generating set, and X is infinite, then any generating set must have cardinality $\geq |X|$. For suppose that Y is also a generating set of A . For each $y \in Y$ there is a finite subset $X_y \subseteq X$ such that $y \in \langle X_y \rangle_T$. Since $Y \subseteq \cup_{y \in Y} \langle X_y \rangle_T$, it follows that $A = \langle Y \rangle_T \subseteq \langle \cup_{y \in Y} X_y \rangle_T$, that is the union $\cup_{y \in Y} X_y$ also generates A , whence $X = \cup_{y \in Y} X_y$, by minimality. We then have the following cardinal estimates:

$$|X| \leq \sum_{y \in Y} |X_y| \leq \omega|Y| = |Y|.$$

Next, suppose that \mathbf{C} admits free algebras and $Y \subseteq X$. The inclusion j of Y in X has a left inverse $h : X \rightarrow Y$. Both j and h have unique extensions to $j^* : F(Y) \rightarrow F(X)$ and $h^* : F(X) \rightarrow F(Y)$, respectively. The composite $h^* \cdot j^*$ extends the identity on Y , and, by uniqueness, is the identity on $F(Y)$. We draw from this that $F(X)$ contains a copy of $F(Y)$; also, if Y is a proper subset of X then h is not one-to-one, and so neither is h^* , which means that j^* cannot be surjective. The point? Y cannot generate $F(X)$; that is, a free generating set is a minimal generating set.

Putting the above remarks together we have:

Lemma 2.6.19. *Suppose that F is free in a class \mathbf{C} admitting free algebras, on an infinite alphabet of cardinality κ . Then F cannot be free of rank $\neq \kappa$.*

Finally then, here is the theorem of Jónsson and Tarski.

Theorem 2.6.20. (Jónsson & Tarski) *Suppose that \mathbf{C} is a class of algebras of type T admitting free algebras. Assume that \mathbf{C} has a nontrivial finite algebra. Then free algebras in \mathbf{C} of different ranks are nonisomorphic.*

Proof. By Lemma 2.6.19, it suffices to consider free algebras on finite alphabets. We actually show more: every generating set for the free algebra in \mathbf{C} on an alphabet of size $n < \omega$ has at least n generators.

Abbreviate F_n for the free \mathbf{C} -object on an alphabet X of size n . Let Y be any (finite) generating set for F_n . Let B be a finite nontrivial algebra. By definition, every function $X \rightarrow B$ has a unique extension to F_n ; thus, the set H of all homomorphisms $F_n \rightarrow B$ has the same cardinality as B^X . Consider now the action of restricting $h \in H$ to Y . Since $\langle Y \rangle_T = F_n$, it follows that $h \mapsto h|_Y$ is injective. This amounts to a one-to-one map $B^X \rightarrow B^Y$. Thus $|B|^n \leq |B|^{|Y|}$, and since all the cardinalities are finite, it follows that $n \leq |Y|$. ■

Remark 2.6.21. Consider the applications of Jónsson and Tarski's Theorem. \mathbf{Gr} has finite nontrivial groups; so does \mathbf{Abel} , the class of all abelian groups. Ditto with \mathbf{Bool} . And so on . . .

By contrast, here is a curious example, also due to Jónsson and Tarski.

Exercise 2.6.22. Suppose that T is the type $(()^l, ()^r, \cdot)$ of two unaries and a binary operation. We consider the equational class \mathbf{K} of all algebras of type T which satisfy the laws

- (i) $x^l x^r = x$;
- (ii) $(xy)^l = x$ and $(xy)^r = y$.

Prove that if F is free on $X \cup \{a, b\}$ and $X \cap \{a, b\} = \emptyset$, then F is isomorphic to the free algebra on X . Use this to show that all the free algebras of \mathbf{K} of nonzero finite rank are isomorphic.

3. Categories: Introduction to Category Theory

This chapter presents the basic language of category theory, with a minimum of technical definitions. The goal is to acquaint the reader with a multitude of examples of categories and functors. The main reference is the book of Herrlich and Strecker ([HS79]).

3.1 Categories and Morphisms.

Definition 3.1.1. A *category* $\mathbf{C} = (\text{obj}(\mathbf{C}), \text{mor}(\mathbf{C}))$ is an ordered pair of classes $\text{obj}(\mathbf{C})$, whose members are called *objects*, and $\text{mor}(\mathbf{C})$, whose members are called *morphisms*, such that for each $f \in \text{mor}(\mathbf{C})$ there are associated a unique ordered pair of objects A and B , called the *domain* and *codomain*, respectively. We use the suggestive functional notation $f : A \longrightarrow B$ to indicate this. Indeed, one popular view of categories looks at a category as a class of arrows, with a partial composition (as described in the axioms below), and in which the objects get explicitly suppressed.

There is a partial binary operation, called *composition*, defined on $\text{mor}(\mathbf{C})$ with the following stipulations:

- (mor1) $f \cdot g$ is defined if and only if the codomain of g is the domain of f .
- (mor2) If $f \cdot g$ and $(f \cdot g) \cdot h$ are defined, then so are $g \cdot h$ and $f \cdot (g \cdot h)$, and conversely, and in either event $f \cdot (g \cdot h) = (f \cdot g) \cdot h$.
- (mor3) For each object $A \in \mathbf{C}$ there is a morphism $1_A : A \longrightarrow A$ so that $f \cdot 1_A = f$ and $1_A \cdot g = g$, whenever the compositions are defined.
- (mor4) For each pair of objects A and B , the collection of all morphisms with domain A and codomain B is a set. This set is denoted by $\text{Hom}_{\mathbf{C}}(A, B)$. (If it is understood which category one is working in, the subscript is omitted.)

Note: $\text{obj}(\mathbf{C})$ is called the *object class* of \mathbf{C} ; $\text{mor}(\mathbf{C})$ the *morphism class* of \mathbf{C} . 1_A is the *identity on* A . We shall frequently use the terms \mathbf{C} -*object* and \mathbf{C} -*morphism* to refer to an element of $\text{obj}(\mathbf{C})$ and $\text{mor}(\mathbf{C})$, respectively.

It is worth emphasizing that both $\text{obj}(\mathbf{C})$ and $\text{mor}(\mathbf{C})$ are “proper” classes in the ordinary classification of Zermelo–Frankel Set Theory, and not necessarily sets. If, however, $\text{obj}(\mathbf{C})$ is a set, we say that \mathbf{C} is a *small* category. We observe, leaving the verification to the reader, that \mathbf{C} is small if and only if $\text{mor}(\mathbf{C})$ is a set. (Verification hinges upon set–theoretic axiomatics, plus the use of (mor4).)

Suppose that $f : A \longrightarrow B$ and $g : B \longrightarrow A$ are morphisms in \mathbf{C} , so that $g \cdot f = 1_A$ and $f \cdot g = 1_B$; then we call f (and g) an *isomorphism*. Two objects in \mathbf{C} are said to be *isomorphic* if there is an isomorphism from one to the other. If A and B are isomorphic in \mathbf{C} we write $A \cong B$. Obviously, the condition of isomorphy is an equivalence relation between objects.

Examples 3.1.2. The reader should note that this is but a small subset of the list of examples provided in [HS79].

- (A) The category \mathbf{Gr} , of all groups and all group homomorphisms.
- (B) The category \mathbf{Rn} , of all rings and all ring homomorphisms. (Putting it differently, all rings relative to the type $T = (0, -(), +, \cdot)$ and all T -homomorphisms.)

- (C) **Rn1**: all rings with identity, and all homomorphisms preserving the identity. (The reader may rephrase this as the class of rings of a certain type, with all possible homomorphisms of that type.)
- (D) For a fixed ring R with identity, \mathbf{RMod} is the category of all left unital R -modules with all left R -homomorphisms. \mathbf{ModR} denotes the category of all right modules and all right R -homomorphisms.
- (E) For any given type T of algebras, the category $\mathbf{Al}(T)$ of all algebras of type T and all T -homomorphisms.
- (F) The category **Top** of all topological spaces and all continuous maps.
- (G) **Top₂**: all Hausdorff spaces and all continuous maps.
- (H) **Bool**: all boolean algebras, and all morphisms that preserve the operations of boolean algebra (henceforth called *boolean morphisms*).
- (I) Let I be a *quasi-ordered class*; that is to say, a class with a relation \leq , which is reflexive and transitive. We regard I as a category, whose object class is I itself, and so that, for $i, j \in I$, the morphism set $I(i, j)$ is empty, unless $i \leq j$, and if $i \leq j$, then $I(i, j)$ has exactly one element, the “arrow” $i \longrightarrow j$.
- (J) A *monoid* is a set M with a binary operation which is associative, and having a twosided identity. A monoid can be regarded as category with one object M , whose morphisms are its elements.
- (K) If \mathbf{C} is any category, we define \mathbf{C}^{op} , called the *opposite category* of \mathbf{C} , so that

$$\text{obj}(\mathbf{C}) = \text{obj}(\mathbf{C}^{\text{op}}), \quad \text{and} \quad \text{mor}(\mathbf{C}) = \text{mor}(\mathbf{C}^{\text{op}}),$$

and $f \in \text{Hom}_{\mathbf{C}}(A, B)$ if and only if $f \in \text{Hom}_{\mathbf{C}^{\text{op}}}(B, A)$, and $f \cdot_{\text{op}} g = g \cdot f$, whenever defined. This is to be thought of as the category one gets from \mathbf{C} by reversing all the arrows.

Exercise 3.1.3. Suppose that \mathbf{C} is any category. Prove that for each object $A \in \mathbf{C}$, the element 1_A whose existence is postulated in (mor3) is unique.

Definition 3.1.4. A *subcategory* \mathbf{B} of the category \mathbf{C} is a category for which

- (sub1) $\text{obj}(\mathbf{B}) \subseteq \text{obj}(\mathbf{C})$,
- (sub2) $\text{mor}(\mathbf{B}) \subseteq \text{mor}(\mathbf{C})$,
- (sub3) composition in $\text{mor}(\mathbf{B})$ is the restriction of composition in $\text{mor}(\mathbf{C})$, and
- (sub4) every isomorphism of $\text{mor}(\mathbf{C})$ with domain and codomain in $\text{obj}(\mathbf{B})$ is contained in $\text{mor}(\mathbf{B})$.

Included in the axiom (sub4) is the postulate that each identity 1_B of \mathbf{C} , with $B \in \text{obj}(\mathbf{B})$, is in \mathbf{B} . (sub4) does not follow from the other postulates; see Example 3.1.5(D) and also Exercise 3.1.6.

If \mathbf{B} is a subcategory of \mathbf{C} , and in addition, $\text{Hom}_{\mathbf{B}}(B, B') = \text{Hom}_{\mathbf{C}}(B, B')$, for all $B, B' \in \text{obj}(\mathbf{B})$, we say that \mathbf{B} is a *full* subcategory of \mathbf{C} . For a full subcategory (sub4) is superfluous.

The question of whether one category is a subcategory of another can be a tricky business. In the examples which follow, when the morphisms are essentially functions, and the composition is not defined it is assumed that it is ordinary composition of functions.

- Examples 3.1.5.** (A) The category **Abel** of all abelian groups and all homomorphisms is a full subcategory of **Gr**, the category of all groups and all homomorphisms.
- (B) **Top₂** is a full subcategory of **Top**. (Refer to Examples 3.1.2, (F) and (G)).
- (C) Let **DLat** be the category of all distributive lattices and all lattice homomorphisms; (that is, all T -homomorphisms, relative to the type $T = (\vee, \wedge)$.) Let **BLat** be the subcategory of all boolean algebras, with all T -homomorphisms. **BLat** is a full subcategory of **DLat**. However, **Bool**, the category of all boolean algebras and all boolean morphisms, is a subcategory of **DLat** which is not full.
- (D) Let **Set** be the category of all sets and all functions between them. The category **Gr** is not a subcategory of **Set**. Although each group has an underlying set, and each homomorphism is a function, a set may support two nonisomorphic group structures. This is a longish way to say that (sub4) in the definition of “subcategory” fails.
- (E) The category **Rn1** of Example 3.1.2(C) is a subcategory of **Rn**, of all rings and all ring homomorphisms, but it is not full. **Rn** is not a subcategory of **Gr**. To see this, note that two rings with identity, which are isomorphic as rings, are isomorphic as **Rn1**-objects, because a ring isomorphism necessarily preserves the identity. On the other hand, if two rings are additively isomorphic, they need not be isomorphic as rings.

Exercise 3.1.6. (This is 4B in [HS79].) Let **C** is the category with one object A ; there are only two morphisms: $\text{Hom}(A, A) = \{a, b\}$, with $a \cdot a = a$, $a \cdot b = b \cdot a = b$, and $b \cdot b = b$. **B** is the category consisting of one object A , so that $\text{mor}(B) = \{b\}$, with $b \cdot b = b$. Show that **B** is not a subcategory of **C**.

3.2 Functors. We introduce the “mappings” between categories. These mappings will have properties which are similar to those of homomorphisms. They come in two varieties: those which preserve the composition, and those which reverse it.

Definition 3.2.1. Suppose that **C** and **D** are categories. A *covariant functor* $F : \mathbf{C} \longrightarrow \mathbf{D}$ is a function so that for each object $A \in \mathbf{C}$, $F(A) \in \text{obj}(\mathbf{D})$, and for each morphism $g : A \longrightarrow B$ in **C**, $F(g)$ is a morphism in **D** with domain $F(A)$ and codomain $F(B)$, so that

- (fun1) for each $A \in \text{obj}(\mathbf{C})$, $F(1_A) = 1_{F(A)}$;
 (fun2) if $f \cdot g$ is defined in **C**, then $F(f \cdot g) = F(f) \cdot F(g)$.

A *contravariant functor* $F : \mathbf{C} \longrightarrow \mathbf{D}$ is also a function so that $F : \text{obj}(\mathbf{C}) \longrightarrow \text{obj}(\mathbf{D})$ and $F : \text{mor}(\mathbf{C}) \longrightarrow \text{mor}(\mathbf{D})$, so that if $g : A \longrightarrow B$ is a **C**-morphism, then $F(g) : F(B) \longrightarrow F(A)$ is a **D**-morphism so that (fun1) holds and

- (fun2') if $f \cdot g$ is defined in **C**, then $F(f \cdot g) = F(g) \cdot F(f)$.

Observe that a contravariant functor $F : \mathbf{C} \longrightarrow \mathbf{D}$ can be viewed as a covariant functor $F^{\text{op}} : \mathbf{C}^{\text{op}} \longrightarrow \mathbf{D}$.

Examples 3.2.2. There follows a list of examples of functors, which, once again, is but a tiny part from of a wealth of examples to be found in [HS79].

- (A) Let \mathbf{Gr} be the category of groups, \mathbf{Abel} the full subcategory of all abelian groups (both with all homomorphisms). If G is a group, let $[G, G]$ stand for the commutator subgroup, and $A(G) = G/[G, G]$. Note that $A(G)$ is abelian. Now if $f : G \rightarrow H$ is any homomorphism between two groups, then, since the image under f of any commutator is again a commutator, it follows that $f([G, G]) \leq [H, H]$; hence, by the Induced Homomorphism Theorem, there is a unique homomorphism $A(f) : A(G) \rightarrow A(H)$, so that the following diagram commutes:

$$\begin{array}{ccc}
 G & \xrightarrow{\mu_G} & A(G) \\
 \downarrow f & & \downarrow A(f) \\
 H & \xrightarrow{\mu_H} & A(H)
 \end{array}$$

Note: $\mu_G : G \rightarrow A(G)$ denotes the canonical map.

The assignment $A : \mathbf{Gr} \rightarrow \mathbf{Abel}$ is a covariant functor.

- (B) For a fixed natural number n , let \mathbf{Ab}_n be the full subcategory of \mathbf{Abel} consisting of all abelian groups G in which $ng = 0$, for all $g \in G$. For each abelian group G , let nG denote the subgroup $\{ng : g \in G\}$. Letting $G^{(n)} = G/nG$, and μ_G once again denote the canonical map $G \rightarrow G^{(n)}$, we have for each homomorphism $f : G \rightarrow H$, an induced homomorphism $f^{(n)} : G^{(n)} \rightarrow H^{(n)}$, so that the diagram below commutes:

$$\begin{array}{ccc}
 G & \xrightarrow{\mu_G} & G^{(n)} \\
 \downarrow f & & \downarrow f^{(n)} \\
 H & \xrightarrow{\mu_H} & H^{(n)}
 \end{array}$$

The assignment $(\)^{(n)} : \mathbf{Abel} \rightarrow \mathbf{Ab}_n$ is a covariant functor.

- (C) Let \mathbf{TfAb} denote the full subcategory of \mathbf{Abel} consisting of all torsion free abelian groups; (recall $G \in \text{obj}(\mathbf{Abel})$ is *torsion free* if $ng = 0$ implies that $g = 0$, in G .) For each abelian group G , let $T(G)$ stand for the set of all elements of G of finite order. This is a subgroup of G . Let $t_G : G \rightarrow G/T(G)$ denote the canonical homomorphism. Since, under any homomorphism of groups, the image of an element of finite order also has finite order, it follows that, if $f : G \rightarrow H$ is a homomorphism of abelian groups, then $f(T(G)) \leq T(H)$. By the Induced Homomorphism Theorem there is a unique homomorphism $\tau(f) : G/T(G) \rightarrow H/T(H)$, such

that $\tau(f) \cdot t_G = t_H \cdot f$. Letting $\tau(G) = G/T(G)$, we obtain a covariant functor $\tau : \mathbf{Abel} \rightarrow \mathbf{TfAb}$.

We feature the Hom functors, which will be of special importance in the theory of rings and modules. There are two variables, and this furnishes us with a covariant functor in one, and a contravariant one in the other.

Definition & Remarks 3.2.3. Suppose that \mathbf{C} is any category, and $A \in \text{obj}(\mathbf{C})$. The assignment $\text{Hom}(A, _) : \mathbf{C} \rightarrow \mathbf{Set}$ defined by setting, for each morphism $f : B \rightarrow B'$, $\text{Hom}(A, f) : \text{Hom}(A, B) \rightarrow \text{Hom}(A, B')$, defined by

$$\text{Hom}(A, f)(g) = f \cdot g,$$

is a covariant functor.

Check: $\text{Hom}(A, 1_B)(g) = 1_B \cdot g = g$, so that $\text{Hom}(A, 1_B) = 1_{\text{Hom}(A, B)}$. Also: if $f : B \rightarrow B'$ and $g : B' \rightarrow B''$ are \mathbf{C} -morphisms, then

$$\text{Hom}(A, g \cdot f)(h) = (g \cdot f) \cdot h = g \cdot (f \cdot h) = \text{Hom}(A, g)(\text{Hom}(A, f)(h)),$$

which means that $\text{Hom}(A, g \cdot f) = \text{Hom}(A, g) \cdot \text{Hom}(A, f)$.

Dually, for each \mathbf{C} -object A , the function $\text{Hom}(_, A) : \mathbf{C} \rightarrow \mathbf{Set}$ which is defined, for each \mathbf{C} -morphism $f : B \rightarrow B'$, as $\text{Hom}(f, A) : \text{Hom}(B', A) \rightarrow \text{Hom}(B, A)$, as $\text{Hom}(f, A)(h) = h \cdot f$, is a contravariant functor.

This example will be revisited many times and in a number of variations, throughout the course.

If it's pure algebra you're after then the following example is a bit far afield. However, it is an important one, and is typical of other like it, in algebraic geometry, which do interest algebraists a great deal.

Example 3.2.4. *The Ring of Continuous Functions on a Compact Hausdorff Space.* We define a contravariant functor from \mathbf{KTop}_2 , the category of all compact Hausdorff spaces and all continuous maps between them, to $\mathbf{CRn1}$, the category of all commutative rings with identity, with all homomorphisms that preserve the identity.

Let X be a compact Hausdorff space, and $C(X)$ stand for the ring of all continuous real valued functions defined on X . This is a commutative ring with identity, under pointwise addition and multiplication, in which the constant 1 is the multiplicative identity. For each continuous map $f : X \rightarrow Y$ between compact Hausdorff spaces, define $C(f) : C(Y) \rightarrow C(X)$, by $C(f)(g) = g \cdot f$. Since the composite of two continuous functions is continuous, this makes sense. We leave it to the reader to verify that

- (i) $C(f)$ is a ring homomorphism, which preserves the identity, and
- (ii) that $C(_)$ is a contravariant functor.

Notice that this is one of the variations promised in 3.2.3. The next example is yet another one.

Example 3.2.5. Recall that a Hausdorff space is said to be *zero-dimensional* if there is a base for the open sets consisting of clopen (= closed-and-open) sets. Now we define a contravariant functor from the category of compact zero-dimensional spaces and all continuous maps, denoted \mathbf{KZero} , to \mathbf{Bool} . For each $X \in \text{obj}(\mathbf{KZero})$, let $\mathfrak{B}(X)$ stand for the set of all clopen subsets. $\mathfrak{B}(X)$ is a boolean algebra, by defining supremum and infimum to be set-theoretic union and intersection,

respectively, and complements to be set-theoretic complements. The least and largest elements are, respectively, \emptyset and X itself, both of which are clopen.

Since, under a continuous map, the inverse image of any open (resp. closed) set is open (resp. closed), it follows that the following is well defined: for each continuous map $f : X \rightarrow Y$, set $\mathfrak{B}(f) : \mathfrak{B}(Y) \rightarrow \mathfrak{B}(X)$ by $\mathfrak{B}(f)(W) = f^{-1}(W)$.

The reader ought to verify that \mathfrak{B} defines a contravariant functor between the categories indicated here.

Example 3.2.6. Forgetful Functors. Let \mathbf{C} be any category; any covariant functor $G : \mathbf{C} \rightarrow \mathbf{Set}$ such that for all $f, g \in \text{Hom}(A, B)$, $G(f) = G(g)$ implies that $f = g$, will be called an *underlying-set functor*. Examples abound: for any category \mathbf{C} of algebras of a given type T , an algebra $(A; T)$ has a “natural” underlying set $G(A)$, and any T -homomorphism f is, first and foremost, a function $G(f)$. Since composition of homomorphisms is just functional composition, the underlying set assignment is a covariant functor, which clearly has the attribute of injectivity.

Likewise, out of **Top**, one has the underlying set $G(X)$ of a topological space X , and for each continuous function f , the underlying function $G(f)$.

The label “forgetful” is another label frequently used in this context, although that encompasses a larger class of functors. “Grounding” functor is yet another. An underlying-set functor is forgetful in the sense that whatever additional structure was present is ignored. A category which admits an underlying-set functor will be called a *concretizable* category. If \mathbf{C} is already furnished with an underlying-set functor $G : \mathbf{C} \rightarrow \mathbf{Set}$ we say that the pair (\mathbf{C}, G) is a *concrete category*.

As is pointed out in [HS79], the relationship between concrete categories and concretizable categories is roughly like the one between metric spaces (i.e., topological spaces already endowed with a metric) and *metrizable spaces* (topological spaces, on which a metric can be defined).

There are examples of nonconcretizable categories, but they are surprisingly difficult to define. For more on this subject, see Exercise 12L in [HS79].

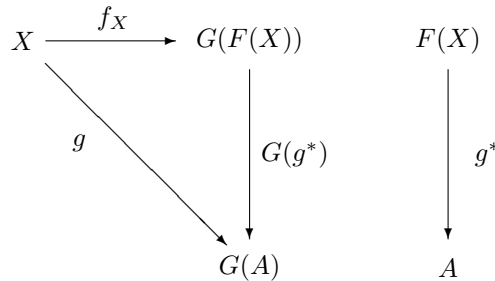
We highlight the injectivity feature of underlying-set functors.

Definition 3.2.7. A functor $F : \mathbf{C} \rightarrow \mathbf{D}$ with the property that, for each pair $f, g \in \text{Hom}_{\mathbf{C}}(A, B)$, $F(f) = F(g)$ implies that $f = g$, will be called a *faithful functor*.

The free constructions of the previous chapter are special examples of free functors.

Example 3.2.8. Free Functors. Suppose that (\mathbf{C}, G) is a fixed concrete category. Suppose that for each set X there is an $F(X) \in \text{obj}(\mathbf{C})$ so that the following properties are satisfied:

- (fr1) For each set X , there is a function $f_X : X \rightarrow G(F(X))$, which is one-to-one, such that
- (fr2) for each $A \in \text{obj}(\mathbf{C})$ and each function $g : X \rightarrow G(A)$, there exists a *unique* \mathbf{C} -morphism $g^* : F(X) \rightarrow A$ such that $g = G(g^*) \cdot f_X$; i.e., such that the diagram below commutes:



These conditions actually define a covariant functor $F : \mathbf{Set} \rightarrow \mathbf{C}$. To see this, let $g : X \rightarrow Y$ be a function, and apply (fr2) to the composite $f_Y \cdot g$. Let $F(g)$ (by definition) be the unique morphism predicted by (fr2). It makes the following diagram commute

$$\begin{array}{ccc}
 X & \xrightarrow{f_X} & G(F(X)) \\
 \downarrow g & & \downarrow G(F(g)) \\
 Y & \xrightarrow{f_Y} & G(F(Y))
 \end{array}
 \quad (*)$$

We leave to the reader the verification that F , thus defined on morphisms, is actually a covariant functor. Such an F is said to be the *free functor for G* . (We note, with emphasis, that a free functor is “paired” with a specific underlying-set functor. Notice as well, that in (fr1) it is required that f_X be injective; this is not standard in the literature. We shall return to this point later.)

It turns out (see Exercise 3.3.9) that if a forgetful functor has a free functor associated with it, then it is essentially unique. This could be proved right away, except that the proper language needed for the formulation of “uniqueness” in this context is that of natural transformations, to be introduced in the next section.

Not every concrete category (\mathbf{C}, G) admits a free functor; see Exercise 3.2.9. We have seen in Theorem 2.4.6 of these notes, that if \mathbf{C} is a class of algebras of a given type T , which is S - and R -closed, then the natural underlying-set functor gives rise to a free functor.

Exercise 3.2.9. Let \mathbf{FGr} be the full subcategory of \mathbf{Gr} consisting of all finite groups. There is no free functor for the natural underlying-set functor of groups; show, in fact, that for any finite set $|X| \geq 1$, the free finite group on X does not exist. (Hint: for a given nontrivial finite set X , show that there are arbitrarily large finite groups having generating sets of the cardinality of X . With this observation, ponder that if a free functor exists, then every group which is generated by X , must be a homomorphic image of the free finite group on X .)

Exercise 3.2.10. Suppose that \mathbf{Dom} is the full subcategory of $\mathbf{CRn1}$ consisting of all integral domains. Does the natural underlying-set functor of \mathbf{Dom} have a free functor? If so identify it.

Definition 3.2.11. Suppose that \mathbf{C} is a subcategory of \mathbf{D} . The functor $I : \mathbf{C} \rightarrow \mathbf{D}$, defined by $I(A) = A$ (for each \mathbf{C} -object A) and $I(f) = f$ (for each \mathbf{C} -morphism f) is covariant, and will be called the *inclusion functor of \mathbf{C} in \mathbf{D}* .

Exercise 3.2.12. Refer to Exercise 2.4.4(b). Let \mathbf{Sp} be the full subcategory of $\mathbf{CRn1}$ consisting of all rings which have no nonzero nilpotent elements. (According to 2.4.4(b), $A \in \text{obj}(\mathbf{Sp})$ if and only if A is a commutative ring with identity, and $n(A) = 0$.) For each $\mathbf{CRn1}$ -object A , let $\hat{A} \equiv A/n(A)$, and $\mu_A : A \rightarrow \hat{A}$ be the canonical homomorphism. Prove that the assignment $(\hat{\quad})$ gives rise to a covariant functor from $\mathbf{CRn1}$ to \mathbf{Sp} .

The final exercise in this section generalizes Exercise 3.2.12.

Exercise 3.2.13. Suppose T is a type of algebras, and \mathbf{C} is a full subcategory of $\mathbf{Al}(T)$, for which the object class is S - and R -closed. For each algebra $(A; T)$ let β_A be the verbal congruence of A , associated with $\text{obj}(\mathbf{C})$. Let $\mu_A : A \rightarrow A^{\mathbf{C}} = A/\beta_A$ denote the canonical homomorphism.

Show the following:

- (i) If $f : A \rightarrow B$ is any T -homomorphism, then $(x, y) \in \beta_A$ implies that $(f(x), f(y)) \in \beta_B$. (Hint: review the proof of Lemma 2.5.2.)
- (ii) Apply the Induced Homomorphism Theorem to show that there is a unique T -homomorphism $f^{\mathbf{C}} : A^{\mathbf{C}} \rightarrow B^{\mathbf{C}}$, so that $f^{\mathbf{C}} \cdot \mu_A = \mu_B \cdot f$.
- (iii) Show that the application $()^{\mathbf{C}}$ is a covariant functor from $\mathbf{Al}(T)$ to \mathbf{C} .

3.3 Natural Transformations. The purpose of this section is to set out conditions under which two categories are regarded as being “the same”. The concept is called *equivalence* of categories. More importantly, one needs to be able to say when two functors act in essentially the same fashion. This is what natural transformations are for.

Definition 3.3.1. Let us suppose that F and G are covariant functors from the category \mathbf{C} to the category \mathbf{D} . A *natural transformation* $\alpha : F \rightarrow G$, is a function from $\text{obj}(\mathbf{C})$ into $\text{mor}(\mathbf{D})$, so that $\alpha_X : F(X) \rightarrow G(X)$ is a \mathbf{D} -morphism for each \mathbf{C} -object X , and for each \mathbf{C} -morphism $g : X \rightarrow Y$, the following diagram commutes:

$$\begin{array}{ccc}
 F(X) & \xrightarrow{\alpha_X} & G(X) \\
 \downarrow F(g) & & \downarrow G(g) \\
 F(Y) & \xrightarrow{\alpha_Y} & G(Y)
 \end{array}$$

We have already seen a fair number of natural transformations, without labeling them as such. Before recalling some of them explicitly, let us issue a warning, which experience teaches is well worth the effort. The label “natural transformation” carries a lot of suggestion; for example, that the mappings involved are always surjective or one-to-one or even isomorphisms. Nothing of the sort is assumed in general! In the notation above the diagram, if $\alpha : F \rightarrow G$ is a natural transformation and, for each \mathbf{C} -object X , α_X is an isomorphism, we say that α is a *natural equivalence of functors*. It is also said that F and G are *naturally equivalent*. The terms *natural isomorphism of functors* and, respectively, *naturally isomorphic* are also used in category theory. If F and G are naturally equivalent functors one writes $F \cong G$.

For a natural transformation $\alpha : F \rightarrow G$ between contravariant functors F and G , do the obvious, by making the vertical arrows in the preceding diagram go up instead of down.

Hot Air 3.3.2. If $F : \mathbf{C} \rightarrow \mathbf{D}$, and $G : \mathbf{D} \rightarrow \mathbf{E}$ are functors between categories, one can define the composition $G \cdot F$, from \mathbf{C} to \mathbf{E} , by setting $(G \cdot F)(A) = G(F(A))$, for each \mathbf{C} -object A , and for each \mathbf{C} -morphism $g : A \rightarrow B$, putting $(G \cdot F)(g) = G(F(g))$. Note that, if either both F and G are covariant or both contravariant, then $G \cdot F$ is covariant, whereas if one of the two is covariant and the other contravariant, then the composite is contravariant.

It may seem strange to have the above remark inserted here. Here is a reason. One important phenomenon in mathematics involves the issue of dualities between categories; that is, having the opposite of one category be equivalent to another. The functors which realize this will be contravariant. On the other hand it is their composites which will be of interest, and those are covariant. So although contravariant functors may be involved in a comparison of categories, the natural transformations through which this “identification” is expressed are between covariant functors.

As in the previous sections, the greatest benefit is derived at this juncture from considering as many examples as possible. We begin.

Examples 3.3.3. We refer freely to the previous lists of examples and use the notation developed there. Let us agree first to denote by $1_{\mathbf{C}}$ the identity functor on the category \mathbf{C} .

- (a) *Free Functors Revisited.* (See 3.2.8.) Let (\mathbf{C}, G) be a concrete category, with underlying-set functor G . Suppose that there is a free functor $F : \mathbf{Set} \rightarrow \mathbf{C}$ associated with G . Then the map $X \rightarrow f_X$ defines a natural transformation $f : 1_{\mathbf{Set}} \rightarrow G \cdot F$; this is illustrated by the diagram (*) in 3.2.8.

Observe that, *in this example*, each f_X is a one-to-one map. It is reasonable to ask to what extent an underlying-set functor G uniquely determines the free functor. More precisely, if F and F' are free functors for G , what is the relationship between F and F' ? Copying the line of reasoning in Exercise 2.4.9, we have for each set X , an isomorphism $\alpha_X : F(X) \rightarrow F'(X)$ so that

$$f'_X = \alpha_X \cdot f_X.$$

(Note: $f'_X : X \rightarrow F'(X)$ is the natural transformation associated to F' .)

We leave it to the reader to check that α is a natural transformation, making the free functors F and F' naturally equivalent. Else, see Exercise 3.3.9.

- (b) Refer to Exercise 3.2.13. Suppose that \mathbf{C} is a subcategory of $\mathbf{Al}(T)$ for which $\text{obj}(\mathbf{C})$ is both S - and R -closed. Let I stand for the inclusion functor of \mathbf{C} in $\mathbf{Al}(T)$. Then the natural homomorphisms $\mu_A : A \rightarrow I(A^{\mathbf{C}})$, define a natural transformation $\mu : 1_{\mathbf{Al}(T)} \rightarrow I \cdot (\)^{\mathbf{C}}$. In this example, each μ_A is, in fact, surjective.

Exercise 3.3.4. With respect to the examples in 3.2.2, and using appropriate inclusion functors, formulate the situations described there in terms of natural transformations between functors.

For the next exercise, refer to Example 3.2.5. It describes “half” of the famous Stone Duality between boolean algebras and compact Hausdorff and zero-dimensional spaces.

Exercise 3.3.5. We consider, once again, the contravariant functor $\mathfrak{B} : \mathbf{KZero} \rightarrow \mathbf{Bool}$ defined in 3.2.5. Suppose now that B is any boolean algebra; let $\text{Max}(B)$ denote the set of all maximal ideals of B . (Note: for the definition of an ideal in a boolean algebra, refer to the discussion in Exercises 1.5.6 and 1.5.7.) We are about to consider the members of $\text{Max}(B)$ as points in a topological space. We will use small German script letters for the points. To define a topology, one must decide which sets are to be regarded as open.

For each $x \in B$, let

$$\Phi(x) \equiv \{ \mathfrak{m} \in \text{Max}(B) : x \notin \mathfrak{m} \}.$$

Now, a certain amount of drudge-work will verify the following:

- (a) Show that, for each $x, y \in B$,
 - (i) $\Phi(x \wedge y) = \Phi(x) \cap \Phi(y)$, and $\Phi(x \vee y) = \Phi(x) \cup \Phi(y)$, and
 - (ii) $\Phi(x') = \text{Max}(B) \setminus \Phi(x)$.

(Hint: For (ii), it helps to invoke Proposition 1.6.4.)

- (b) The sets $(\Phi(x))_{x \in B}$ form a base for a topology on $\text{Max}(B)$, consisting of clopen sets.
- (c) With respect to this topology, $\text{Max}(B)$ is a compact, Hausdorff space.
- (d) If $f : B \rightarrow B'$ is a boolean homomorphism, then the map $\text{Max}(f) : \text{Max}(B') \rightarrow \text{Max}(B)$, defined by

$$\text{Max}(f)(\mathfrak{m}) = f^{-1}(\mathfrak{m}),$$

is a continuous function.

- (e) Show that Max defines a contravariant functor from **Bool** to **KZero**.
- (f) Show that a subset of $\text{Max}(B)$ is clopen if and only if it is of the form $\Phi(x)$, for some $x \in B$.
- (g) For each boolean algebra B , define $\Phi_B : B \rightarrow \mathfrak{B}(\text{Max}(B))$, by setting $\Phi_B(x) = \Phi(x)$. Show that $\Phi_B \in \text{mor}(\mathbf{Bool})$. (Apply (a).) Show, in fact, that Φ_B is an isomorphism.
- (h) Finally, prove that $\Phi : \mathbf{1}_{\mathbf{Bool}} \rightarrow \mathfrak{B} \cdot \text{Max}$ is a natural transformation, and therefore a natural equivalence. (Note that although both \mathfrak{B} and Max are contravariant, their composite is covariant.)

At last, the “natural” notion of identity of categories, accompanied by the notion of dual categories.

Definition & Remarks 3.3.6. Suppose that \mathbf{C} and \mathbf{D} are categories. We say that \mathbf{C} and \mathbf{D} are *equivalent categories* if there exist covariant functors $F : \mathbf{C} \rightarrow \mathbf{D}$ and $G : \mathbf{D} \rightarrow \mathbf{C}$ such that $G \cdot F \cong \mathbf{1}_{\mathbf{C}}$, and $F \cdot G \cong \mathbf{1}_{\mathbf{D}}$.

If the functors in the previous paragraph are contravariant, then we say that the categories are *dual*, and the functors in question are called *dualities*.

Incidentally, in Exercise 3.3.5, the upshot is that $\mathfrak{B} \cdot \text{Max} \cong \mathbf{1}_{\mathbf{Bool}}$, by putting parts (g) and (h) together. It is, in fact, the case that $\text{Max} \cdot \mathfrak{B} \cong \mathbf{1}_{\mathbf{KZero}}$.

Exercise 3.3.7. Construct two inequivalent categories \mathbf{C} and \mathbf{D} , such that $\text{obj}(\mathbf{C}) = \text{obj}(\mathbf{D})$; $\text{mor}(\mathbf{C}) = \text{mor}(\mathbf{D})$; and if f is a morphism in either one, its domain and codomain are the same in both categories. (Hint: you can get by with categories having one object!)

Here is an example involving inclusion of a non-full subcategory.

Exercise 3.3.8. Let \mathbf{Rn} be the category of rings and all ring homomorphisms, and $\mathbf{Rn1}$ be the subcategory of all rings with identity, with all homomorphisms that preserve the identity. (Note that this is not a full subcategory.) Let $I : \mathbf{Rn1} \rightarrow \mathbf{Rn}$ be the inclusion functor.

Now, for each ring R , let $R^* = R \times \mathbb{Z}$, and define addition and multiplication in R^* as follows: the addition is coordinatewise, while

$$(r, m) \cdot (s, n) = (rs + ms + nr, mn).$$

Show that

- (a) R^* is a ring with identity. (But observe that, if R already has an identity, the identity of R^* will be different! Always!)
- (b) The map $\delta_R : R \rightarrow I(R^*)$, defined by $\delta_R(r) = (r, 0)$, is a one-to-one homomorphism.
- (c) For each \mathbf{Rn} -morphism $f : R \rightarrow S$ there is a unique $\mathbf{Rn1}$ -morphism $f^* : R^* \rightarrow S^*$ such that

$$f^* \cdot \delta_R = \delta_S \cdot f.$$

- (d) The application $(\)^* : \mathbf{Rn} \rightarrow \mathbf{Rn1}$ is a covariant functor, and $\delta : \mathbf{1}_{\mathbf{Rn}} \rightarrow I \cdot (\)^*$ is a natural transformation. (Observe that although δ_R is always one-to-one, it is never an isomorphism.)

Exercise 3.3.9. Suppose that (\mathbf{C}, G) is a concrete category, and that F_1 and F_2 are free functors for G . Prove that F_1 is naturally equivalent to F_2 .

4. Categories: Limits and Colimits

This chapter explores the general principle of “universality” through the concepts of limits and colimits. Among the limits are the product, the inverse limit and the pullback; colimits generalize ubiquitous notions, like the free product, the direct limit and the pushout. We have seen these examples already in the chapters on universal algebra. Here we will also highlight results involving limits and colimits which make theorems in homological algebra easier to understand.

4.1 Products and Coproducts. We begin with a particular pair of universal concepts. In due course the concept of a categorical diagram will make its formal appearance. In the case of products and coproducts the diagrams are simplest; indeed, void.

Definition & Remarks 4.1.1. Suppose $\{A_i : i \in I\}$ is a set of algebras of the same type T . Let A be their direct product, and (for each $i \in I$) let $\pi_i : A \rightarrow A_i$ be the projection upon the i -th coordinate. Note (again) that each π_i is a T -homomorphism.

Now suppose that $(B; T)$ is an algebra of type T , and $\{g_i : B \rightarrow A_i : i \in I\}$ is a set of T -homomorphisms. Then there is a unique T -homomorphism $g : B \rightarrow A$ such that $\pi_i \cdot g = g_i$, for each $i \in I$. (Just define g as follows:

$$g(b)(i) = g_i(b),$$

for each index i . One shows that g is a T -homomorphism – the proof is left as an exercise – and the uniqueness should be obvious.

Observe the following, as a consequence of the uniqueness: suppose that $(A'; T)$ is an algebra of type T , furnished with a set of T -homomorphisms $(\pi'_i : A' \rightarrow A_i)_{i \in I}$ so that, for each algebra $(B; T)$, and each family of homomorphisms $(g_i : B \rightarrow A_i)_{i \in I}$, there is a unique homomorphism $g' : B \rightarrow A'$, for which $\pi'_i \cdot g' = g_i$, for each index i . Then A' is isomorphic to the direct product A .

Proof. There is a homomorphism $g : A' \rightarrow A$ such that $\pi_i \cdot g = \pi'_i$, for each $i \in I$. By the same token there is a homomorphism $h : A \rightarrow A'$ so that $\pi'_i \cdot h = \pi_i$, for each index i . Composing we get:

$$\pi_i \cdot g \cdot h = \pi_i = \pi_i \cdot 1_A,$$

and

$$\pi'_i \cdot h \cdot g = \pi'_i = \pi'_i \cdot 1_{A'},$$

for each $i \in I$. By the uniqueness provision in the above, we conclude that $g \cdot h = 1_A$ and $h \cdot g = 1_{A'}$. ■

The thing to stress here, and which will be pointed out again in the examples to come, is the “universality” of the direct product; that is to say, the notion that the direct product of algebras is completely determined, up to isomorphism, by a condition which is expressed in terms of objects and morphisms only. **NO ELEMENTS ARE MENTIONED!** This either endears category theory to one’s soul, or disgust springs forth instantly!

In any event, the above discussion motivates the following definition.

Definition & Remarks 4.1.2. Suppose that \mathbf{C} is any category. Let $\{A_i : i \in I\}$ be a set of objects in \mathbf{C} . The *product* of the A_i is an object A , together with a family of morphisms $p_i : A \rightarrow A_i$ such that, for each set of morphisms $\{g_i : B \rightarrow A_i : i \in I\}$, there is a unique morphism $g : B \rightarrow A$, such that $p_i \cdot g = g_i$, for each index i . (There is no reason, a priori, that the product should exist!) If this exists, then the argument in 4.1.1 demonstrates that it is unique up to isomorphism.

If every set of \mathbf{C} -objects has a product we shall say that the category *has products*. In a product such as the one described in the preceding paragraph, the morphisms p_i are called *projections*.

In many of the concrete categories we love the product exists and is none other than the cartesian product of the underlying sets, with some canonical structure defined on it. This is what we have already discovered in categories of algebras.

For topological spaces one can define the product by defining the topology on the cartesian product to be the smallest one (with fewest open sets) making all the projections continuous. The topology described in this manner has, as a base of closed sets, the following subsets of the cartesian product: suppose that $Y = \prod_{i \in I} Y_i$; pick a finite subset F of I , and for each $i \in F$, an open set V_i of Y_i ; then consider

$$V(F) = \{f \in Y : f(i) \in V_i, \forall i \in F\}.$$

These $V(F)$ (over all finite $F \subseteq I$) form a base for the so-called *Tychonoff* topology on Y .

It is not hard to show that, with this topology, and the coordinate projection maps, Y is the product, in **Top**, of the Y_i . For further details we refer the reader to [Wi70].

Next, we illustrate the dual of products, in a very particular circumstance, the category of abelian groups **Abel**. Dualizing refers to turning all the arrows around in the definition of product.

Exercise 4.1.3. Suppose that $\{G_i : i \in I\}$ is a set of abelian groups. Let G be their *direct sum*:

$$G = \{g \in \prod_{i \in I} G_i : g(i) = 0, \forall \text{ except finitely many } i\}.$$

Let $\delta_i : G_i \rightarrow G$ be the map defined by

$$\delta_i(g)(j) = \begin{cases} 0, & \text{if } j \neq i \\ g, & \text{if } j = i. \end{cases}$$

Then prove the following:

- (a) Each δ_i is a homomorphism.
- (b) If $\{f_i : G_i \rightarrow H : i \in I\}$ is a set of homomorphisms, then there is a unique homomorphism $f : G \rightarrow H$ so that $f \cdot \delta_i = f_i$, for each index i .
- (c) Prove that if A is an abelian group, with homomorphisms $\delta'_i : G_i \rightarrow A$ ($i \in I$), satisfying the universal condition in (b), then $A \cong G$.

(So notice once again that the direct sum is characterized completely by conditions involving ONLY objects and morphisms of the category.)

Definition 4.1.4. Suppose that \mathbf{C} is a category, and $\{A_i : i \in I\}$ is a set of objects. Their *coproduct* is an object A together with a set of morphisms $(d_i : A_i \rightarrow A)_{i \in I}$, such that, for each family of morphisms $h_i : A_i \rightarrow B$, there is a unique morphism $h : A \rightarrow B$ so that $h \cdot d_i = h_i$, for each $i \in I$.

If for each set of objects in \mathbf{C} the coproduct exists, then we say that \mathbf{C} *has coproducts*. Notice that the definition of the coproduct is exactly that of the product, with all the arrows reversed. Another way of saying this is that a coproduct in \mathbf{C} is a product in \mathbf{C}^{op} . In the coproduct of the previous paragraph, the morphisms d_i are called the *coprojections*. As with products, the definition of coproduct makes it unique up to an isomorphism. We denote the coproduct of A_i (with coprojections d_i) by $\coprod_{i \in I} A_i$.

We should observe here that, unlike the case with the product, the coproduct is badly behaved in most familiar category. **Abel** and the categories \mathbf{RMod} and \mathbf{ModR} of all left and, respectively, right unital modules over the ring R with identity, are notable exceptions. (In the two categories of R -modules the coproduct is again the direct sum.)

Even in \mathbf{Gr} , which has coproducts, it is not easy to describe what the coproduct is. It is *not* the direct sum; refer to the exercise which follows. In general, the free product of groups G and H is denoted by $G * H$; each element ($\neq e$) can be written uniquely as $x_1 x_2 \cdots x_n$, where $x_i \in G \cup H$, but no two adjacent x_i lie in G or else in H . The product is defined, inductively:

$$(x_1 x_2 \cdots x_n) \cdot (y_1 y_2 \cdots y_m) = \begin{cases} x_1 x_2 \cdots x_n y_1 y_2 \cdots y_m, & \text{if } x_n \in G, y_1 \in H, \text{ or } x_n \in H, y_1 \in G; \\ x_1 x_2 \cdots x_{n-1} z y_2 \cdots y_m, & \text{if } x_n, y_1 \in G, \text{ or } x_n, y_1 \in H, \\ & \text{with } z = x_n \cdot y_1 \neq e; \\ (x_1 \cdots x_{n-1})(y_2 \cdots y_m), & \text{otherwise.} \end{cases}$$

In \mathbf{Gr} the coproduct is often referred to as the *free product*.

Exercise 4.1.5. In \mathbf{Gr} argue the following: Let Z_1 and Z_2 be two copies of the additive group of integers. Show that the coproduct $Z_1 \coprod Z_2$ is the free group on two generators. Then conclude that the coproduct cannot be the direct sum of the two groups.

We also give a topological example.

Exercise 4.1.6. This takes place in **Top**. Let $\{X_i : I \in I\}$ be a set of topological spaces. Let X be their disjoint union. (In X , therefore, every subset A is itself a disjoint union of subsets $A_i \subseteq X_i$.) Call a subset $A = \cup_{i \in I} A_i$ open in X if each A_i is open in X_i . Let $u_i : X_i \rightarrow X$ denote the inclusion map.

- Prove that, with these coprojections, X is the coproduct of the spaces X_i .
- Prove that the topology defined on X is the largest – with the most open sets – making all the u_i continuous.

The final exercise of this section ties in with Exercise 4.1.5, above.

Exercise 4.1.7. If (\mathbf{C}, G) is a concrete category, which has coproducts and also a free functor F for the underlying-set functor G , then show that, for each set X , $F(X)$ is the coproduct of X copies of the free object $F(1)$ on the set of one element.

4.2 Direct Limits and Equalizers. We initiate the discussion on limits and colimits with an example.

Example 4.2.1. In the category \mathbf{RMod} of left (unital) modules over the rings R with identity, suppose that M is a module over R , and $\{K_i : i \in I\}$ is a family of submodules of M , with I bearing a partial ordering \leq which is *upper directed*: if $i, j \in I$, there exists a $k \in I$ such that $i \leq k$ and $j \leq k$. Assume further that if $i \leq j$ then $K_i \leq K_j$. Denote the inclusion of K_i in K_j ($i \leq j$) by u_{ij} .

Suppose now that A is a left R -module and take a family $(f_i : K_i \rightarrow A)$ of \mathbf{RMod} -morphisms so that, for all $i \leq j$, $f_j \cdot u_{ij} = f_i$; in effect, so that f_j agree with f_i when restricted to K_i . Then, letting $u_i : K_i \rightarrow K = \cup_{i \in I} K_i$ denote the inclusion of K_i in the union, there is a unique \mathbf{RMod} -morphism $f : K \rightarrow A$ such that $f \cdot u_i = f_i$ ($i \in I$).

To establish this claim, let $x \in K$, and define $f(x) = f_i(x)$, whenever $x \in K_i$. The compatibility among the f_i insures that f is well defined. It is trivial that f is an R -homomorphism and unique as asserted.

This is an example of a *direct limit*. It is an example of a universally defined concept, with an interactive condition between the objects; note: $u_{ii} = 1$ (on K_i), and if $i \leq j \leq k$, then $u_{jk}u_{ij} = u_{ik}$. Notice, in the discussion of the preceding paragraphs, that the condition defining f , from the f_i and u_{ij} , is element-free.

With the union as a model we embark on a description of direct and inverse limits, in general.

Definition & Remarks 4.2.2. Let \mathbf{C} be any category. Suppose that I is an upper directed partially ordered set – henceforth abbreviated *poset* – and assume that we are given a family of \mathbf{C} -morphisms $\{\tau_{ij} : B_i \rightarrow B_j : i \leq j \in I\}$, so that

(dir1) $\tau_{ii} = 1_{B_i}$, for each $i \in I$;

(dir2) if $i \leq j \leq k$, then

$$\tau_{jk} \cdot \tau_{ij} = \tau_{ik}.$$

Such a system of morphisms is called a *direct system*. We denote it, in brief, by (B_i, τ_{ij}) . We shall refer to the defining conditions (dir1) and (dir2) as *bonding conditions*. The morphisms τ_{ij} themselves are called *bonding morphisms*.

Now, suppose that (B_i, τ_{ij}) is a direct system. We say that the *direct limit* of the system exists, if there is a \mathbf{C} -object B , and \mathbf{C} -morphisms $\tau_i : B_i \rightarrow B$ (for each $i \in I$), such that the diagram below commutes:

$$\begin{array}{ccc} B_i & \xrightarrow{\tau_{ij}} & B_j \\ & \searrow \tau_i & \downarrow \tau_j \\ & & B \end{array}$$

and such that, whenever $f_i : B_i \rightarrow X$ are \mathbf{C} -morphisms for which the previous diagram commutes, with X replacing B and f_i in place of τ_i – that is, we assume that, for each $i \leq j$, $f_j \cdot \tau_{ij} = f_i$ – then there is a unique \mathbf{C} -morphism $f : B \rightarrow X$ such that $f \cdot \tau_i = f_i$, for each index i . For brevity the direct limit is denoted

$$B = \varinjlim (B_i, \tau_{ij}).$$

In the exercise which follows it is asserted that if such a direct limit exists, then it is unique up to isomorphism.

If every direct system in \mathbf{C} has a direct limit, we say that the category *has direct limits*. Once again, it is worth pointing out, that this is a concept which is defined (up to isomorphism) in terms of objects and morphisms. The direct limit is a cousin to the coproduct, in the sense that the arrows go from the members of the system into the limit object, and that the limit object is universal relative to any other one which is compatible with the system.

Exercise 4.2.3. Verify that if a direct system has a limit, then it is unique up to isomorphism.

The following theorem serves as a model for many constructions which occur everywhere in algebra, when proving the existence of what we shall soon define as a colimit. The idea of the proof is simple: one constructs a free object, and then proceeds to factor out the minimum that forces bonding. It's the details that are annoying.

It should also be mentioned that we shall greatly improve on Theorem 4.2.4 in the next section; see Theorem 4.3.14.

Theorem 4.2.4. *Suppose that \mathbf{E} is a category of algebras of type T , so that $\text{obj}(\mathbf{E})$ is an equational class. Then \mathbf{E} has direct limits.*

Proof. Suppose that $\{A_i : i \in I\}$ is a family of \mathbf{E} -objects, with I upper directed, and (A_i, τ_{ij}) is a direct system. Let X be the disjoint union of the A_i ; technically speaking, $X = \cup_{i \in I} G(A_i)$, where G is the natural underlying-set functor. Let F be the free \mathbf{E} -algebra on the set X , and $u_X : X \rightarrow G(F)$ denote the universal embedding.

Let θ be the smallest congruence which contains all the following pairs:

(a) $(u_X(x), u_X(G(\tau_{ij}(x))))$, for all $x \in A_i$, and all $j \geq i$.

(b) For each n -ary operation ϕ , with $n \geq 1$, and all $a_1, \dots, a_n \in A_i$,

$$(\phi_F(u_X(a_1), \dots, u_X(a_n)), u_X(\phi_{A_i}(a_1, \dots, a_n))),$$

where the first composition is carried out in F , the second in A_i .

(c) If ϕ is a nullary operation, the pair $(\phi_F, u_X(\phi_{A_i}))$.

(Note: ϕ_B denotes the corresponding nullary element in the algebra B .)

Put $A = F/\theta$, and let $\tau_i : A_i \rightarrow A$, the map $\tau_i(x) = \theta[u_X(x)]$. Note that by definition, $\tau_j \cdot \tau_{ij} = \tau_i$, for each $i \leq j$. (This depends on the freeness of F and the fact that the underlying-set functor is faithful.)

Let us first verify that each τ_i is a homomorphism. If ϕ is a nullary operation, then, by virtue of (c),

$$\tau_i(\phi_{A_i}) = \theta[u_X(\phi_{A_i})] = \theta[\phi_F] = \phi_A.$$

If ϕ is n -ary ($n \geq 1$), and $a_1, \dots, a_n \in A_i$, then, using (b),

$$\begin{aligned} \tau_i(\phi(a_1, \dots, a_n)) &= \theta[\phi(u_X(a_1), \dots, u_X(a_n))] \\ &= \phi(\theta[u_X(a_1)], \dots, \theta[u_X(a_n)]) \\ &= \phi(\tau_i(a_1), \dots, \tau_i(a_n)). \end{aligned}$$

Now, suppose that $g_i : A_i \rightarrow B$ ($B \in \text{obj}(\mathbf{E})$) is a family of homomorphisms, so that $g_j \tau_{ij} = g_i$, whenever $i \leq j$. Let $h : F \rightarrow B$ be the unique homomorphism for which $h(u_X x) = g_i(x)$, for each $x \in A_i$. Observe that

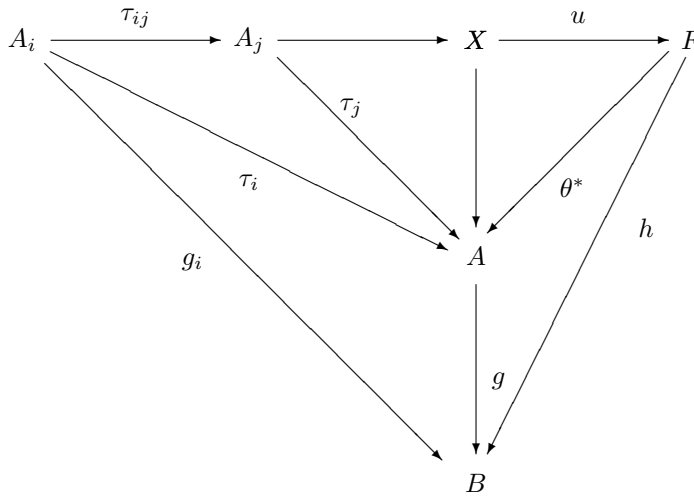
$$h(u_X x) = g_i(x) = g_j(\tau_{ij}(x)) = h(u_X(\tau_{ij}(x))),$$

whenever $x \in A_i$, whence $\theta \leq \ker(h)$. The reader may then also easily verify that the pairs in (b) and (c) are identified by h . Consequently, $\theta \leq \ker(h)$, and by the Induced Homomorphism Theorem, there is a unique homomorphism $g : A = F/\theta \longrightarrow B$, for which $g(\theta[z]) = h(z)$, for each $z \in F$. This means that if $x \in A_i$,

$$g(\tau_i(x)) = g(\theta[u_X(x)]) = h(u_X(x)) = g_i(x).$$

Thus: $g \cdot \tau_i = g_i$. Except for the demonstration that g is uniquely determined by the g_i (as opposed to “by h ”), which we leave to the reader, this completes the proof. ■

Remark 4.2.5. To complement the above proof, we display a diagram which might help guide the reader through the labyrinth of maps. $\theta^* : F \longrightarrow A = F/\theta$ is the canonical homomorphism. Although not said in that way in the foregoing, $g \cdot \theta^* = h$. Technically speaking, the underlying-set functor ought to be applied to all objects from \mathbf{E} appearing in the diagram. We have also suppressed the subscript X on the universal map u_X .



Hot Air 4.2.6. Intuition? Keep the following in mind: in the existence arguments involving these limits and colimits say very little about the structure of the object that’s produced. This has already been hinted at in the remarks about coproducts. In most concrete settings it is frequently the case that products, pullbacks and the like – what we will call limits – are relatively easy to describe, whereas coproducts, direct limits, pushouts, etc. – the soon-to-be-christened colimits – are often shrouded in mystery. This apparent imbalance in the state of affairs is curious. It is usually explained away by saying that when one constructs something akin to the direct limit, by factoring out a chunk out of a free object, it is not clear what (exactly) is being identified. This observation, while right on the money, is not very encouraging, and does very little to solve the individual mysteries.

For example, returning to Example 4.2.1, in \mathbf{RMod} , and with a direct system for which the bonding maps are inclusions – i.e., one-to-one – the direct limit turns out to be the set-theoretic union. If, however, even in \mathbf{RMod} , the maps are more complicated, then the direct limit is much more difficult to describe.

Exercise 4.2.7. Let p be a fixed prime number. In **Abel** consider the groups \mathbb{Z}_{p^i} ($i = 1, 2, \dots$). Let $u : \mathbb{Z}_{p^i} \rightarrow \mathbb{Z}_{p^{i+1}}$ denote the homomorphism $u(k) = pk$. This yields a direct system $u^j : \mathbb{Z}_{p^i} \rightarrow \mathbb{Z}_{p^{i+j}}$, by iteration of u . What is the direct limit in **Abel** of this system? (Hint: Have a look back at Exercise 1.3.5(c), and think multiplicatively.)

Let's reverse arrows in a direct limit, first considering matters for groups. (b) in the exercise which follows gives a "construction" of inverse limits of groups.

Exercise 4.2.8. Work in **Gr**. Suppose that I is an upper directed set, and $\{G_i : i \in I\}$ is a set of groups, endowed with homomorphisms $r_{ij} : G_i \rightarrow G_j$ (for all $i \geq j$), so that

- (inv1) $r_{ii} = 1_{G_i}$, for each $i \in I$;
- (inv2) if $i \geq j \geq k$, then $r_{jk} \cdot r_{ij} = r_{ik}$.

Such a system (G_i, r_{ij}) ($i \geq j$) is called an *inverse system*. An *inverse limit* $(G, r_i)_{i \in I}$ is a group G with a family of homomorphisms $r_i : G \rightarrow G_i$, so that if $i \geq j$, $r_{ij} \cdot r_i = r_j$, and so that if H is any group together with homomorphisms $f_i : H \rightarrow G_i$, for which $r_{ij} \cdot f_i = f_j$, whenever $i \geq j$, then there is a unique homomorphism $f : H \rightarrow G$, so that $r_i \cdot f = f_i$, for each index i . We write

$$G = \varprojlim (G_i, r_{ij}).$$

Prove the following:

- (a) If $(G_i, r_{ij})_{i \geq j}$ is an inverse system, so that all the G_i are subgroups of a fixed group H , and each r_{ij} is the inclusion map, then the inverse limit is the intersection of the subgroups G_i .
- (b) If $(G_i, r_{ij})_{i \geq j}$ is any inverse system, then the inverse limit G can be described as follows:

$$G = \{g \in \prod_{i \in I} G_i : r_{ij}(g_i) = g_j, \forall i \geq j\}.$$

- (c) Consider the inverse system $(\mathbb{Z}_n, d_{nm})_{n \geq m}$, where $n \geq m$ means: m divides n , and $d_{nm} : \mathbb{Z}_n \rightarrow \mathbb{Z}_m$ is defined by $d_{nm}(k) = k \bmod m$. Show that the inverse limit $\hat{\mathbb{Z}}$ contains a copy of \mathbb{Z} , but is not isomorphic to \mathbb{Z} (by showing that $\hat{\mathbb{Z}}$ is not cyclic).

Exercise 4.2.9. In **Abel** show that every torsion free abelian group is a direct limit of finitely generated free groups.

The notion of the equalizer and its dual, the coequalizer, are presented here, not so much as motivators, but as illustrations which, in the next section, turn out to be of some technical importance.

Definition 4.2.10. Suppose that \mathbf{C} is any category. Let $f, g : A \rightarrow B$ be a pair of morphisms. The *equalizer of f and g* , labelled $\text{Eq}(f, g)$, is a pair $(\text{Eq}(f, g), e)$, where $e : \text{Eq}(f, g) \rightarrow A$, so that $f \cdot e = g \cdot e$, and so that, whenever $h : X \rightarrow A$ is a \mathbf{C} -morphism for which $f \cdot h = g \cdot h$, it follows that there is a unique morphism $h^* : X \rightarrow \text{Eq}(f, g)$ for which $e \cdot h^* = h$.

If any pair of morphisms, with common domain and codomain have an equalizer, we say that the category \mathbf{C} *has equalizers*.

The *coequalizer of f and g* , labelled $\text{Coeq}(f, g)$ is defined dually, by reversing all the arrows in the definition just given for the equalizer. (Do it for practice.) Likewise, the phrase \mathbf{C} *has coequalizers* means that each pair of morphisms with common domain and codomain have a coequalizer.

The next bunch of exercises concern the equalizer and coequalizer in a variety of settings.

Exercise 4.2.11. Prove that the equalizer, when it exists, is unique up to isomorphism.

Exercise 4.2.12. In \mathbf{Gr} , let $f : G \rightarrow H$ be any homomorphism; denote by $z : G \rightarrow H$ the homomorphism $z(g) = e$, for all $g \in G$. Show that if K is the kernel (subgroup) of f , and $i : K \rightarrow G$ the inclusion map, then (K, i) is the equalizer of f and z .

Exercise 4.2.13. Prove that any full subcategory \mathbf{C} of algebras of type T , for which the object class is S -closed has equalizers, and describe the equalizer. Do the same for coequalizers in any full subcategory of algebras whose object class is Q -closed. (Thus, any equational class has both equalizers and coequalizers.)

Exercise 4.2.14. Let \mathbf{C} be the full subcategory of \mathbf{Set} whose objects have *at least* two elements. Then \mathbf{C} has neither equalizers nor coequalizers. Show this.

Exercise 4.2.15. Suppose that $(\text{Eq}(f, g), e)$ is the equalizer of f and g . Show that $f = g$ if and only if e is an isomorphism.

4.3 Limits and Completeness. Finally, we get to the general notion of limits in a category. The keys to an understanding of limits, beyond the examples of the previous two sections, are the concepts of diagram, source and sink.

Definition 4.3.1. Let \mathbf{C} be a category, and I be a small category; that is, one so that $\text{obj}(I)$ is a set. A *diagram in \mathbf{C}* is a covariant functor $D : I \rightarrow \mathbf{C}$.

For example, to obtain the diagram that is used to define the product one chooses a set I , and treats it as a category in which the only morphisms are the identities on each object $i \in I$. For the equalizer $I = \{i, j\}$, with identities on i and j , and exactly two morphisms $i \rightarrow j$.

The appropriate dualized diagrams are used for the coproduct and coequalizer.

Definition 4.3.2. Suppose that $D : I \rightarrow \mathbf{C}$ is a diagram. A *source for D* is an object $A \in \text{obj}(\mathbf{C})$ together with morphisms $\alpha_i : A \rightarrow D(i)$, which are *compatible with D* , meaning, that if $\tau : i \rightarrow j$ is a morphism, then

$$D(\tau) \cdot \alpha_i = \alpha_j.$$

A *sink for D* is the dual concept; that is, a \mathbf{C} -object B with morphisms $\beta_i : D(i) \rightarrow B$, such that, for each $\tau : i \rightarrow j$, $\beta_i = \beta_j \cdot D(\tau)$.

We now have all the preliminaries in place to define the general concept of a limit in a category.

Definition 4.3.3. Suppose that $D : I \rightarrow \mathbf{C}$ is a diagram. We say that D *has a limit* if there is a *universal source for D* ; that is to say, there is a source $(\delta_i : L \rightarrow D(i))_{i \in I}$, with the property that, for any source $(f_i : A \rightarrow D(i))_{i \in I}$ there is a *unique \mathbf{C} -morphism $f : A \rightarrow L$* such that, for each $i \in I$, $\delta_i \cdot f = f_i$. The universality of sources proves that the limit of D is unique, when it exists. The argument is similar to the one needed in Exercise 3.3.9, for example.

D is said to *have a colimit* if there is a *universal sink for D* . (We leave the fleshing out of “universal sink” to the reader.)

A category \mathbf{C} is *complete* if every diagram in \mathbf{C} has a limit. We say that \mathbf{C} is *cocomplete* if every diagram in \mathbf{C} has a colimit.

Remarks 4.3.4. *Looking Back.* Here is a brief review of the universal concepts introduced in the preceding two sections, in terms of the language presented here.

- (a) A product in a category \mathbf{C} is the limit of a diagram $D : I \longrightarrow \mathbf{C}$, where the only morphisms in I are the identities. A coproduct is the colimit of a similar diagram.
- (b) A direct limit is the colimit of a diagram $D : I \longrightarrow \mathbf{C}$, in which I is an upper directed poset. As a category, I is to be thought of as the set of objects with exactly one morphism $i \longrightarrow j$ if and only if $i \leq j$. An inverse limit is the limit of a diagram $D : I^{\text{op}} \longrightarrow \mathbf{C}$, with I again an upper directed poset.
- (c) An equalizer (resp. coequalizer) is a limit (resp. colimit) of a diagram $D : I \longrightarrow \mathbf{C}$, where $I = \{i, j\}$, with the identities on i and j , and exactly two morphisms $i \longrightarrow j$.

Technically speaking, a source should consist of two morphisms in \mathbf{C} , $e_i : A \longrightarrow D(i)$ and $e_j : A \longrightarrow D(j)$, such that

$$e_j = D(m) \cdot e_i = D(m') \cdot e_i,$$

where m and m' are the two morphisms of the diagram. The reader should ponder that this yields a concept of equalizer that is equivalent to the one introduced in the previous section.

Definition 4.3.5. More modestly, if I is a small category, we say that \mathbf{C} is I -complete if every diagram in \mathbf{C} out of I has a limit. I -cocompleteness is defined dually.

Thus, to say that \mathbf{C} has products is to say that \mathbf{C} is I -complete, where I is the category described in 4.3.4(a). If I is an upper directed poset, then “ I -cocomplete” is synonymous with “having direct limits”.

Before proceeding with additional examples of limits and colimits, let us establish an important uniqueness result, already telegraphed by several examples. Since a limit in \mathbf{C} is a colimit in \mathbf{C}^{op} , and vice-versa, the proposition which follows has an obvious “colimit” analogue.

Proposition 4.3.6. *Suppose that \mathbf{C} is a category, and $D : I \longrightarrow \mathbf{C}$ is a diagram. If D has limit sources*

$$(\delta_i : L \longrightarrow D(i))_{i \in I} \quad \text{and} \quad (d_i : L' \longrightarrow D(i))_{i \in I}$$

then there is a unique isomorphism $d : L' \longrightarrow L$ such that $\delta_i \cdot d = d_i$, for each $i \in I$.

Proof. The existence and uniqueness of $d : L' \longrightarrow L$ with the desired property follows since the first source is universal. Likewise, there is a morphism $g : L \longrightarrow L'$ such that $d_i \cdot g = \delta_i$, for each index i . Composing, we have (for each i)

$$\delta_i \cdot (d \cdot g) = \delta_i = \delta_i \cdot 1_L,$$

and

$$d_i \cdot (g \cdot d) = d_i \cdot 1_{L'}.$$

By the uniqueness proviso in the definition of the limit, it follows that $d \cdot g = 1_L$ and $g \cdot d = 1_{L'}$. ■

Examples 4.3.7. Some additional important limits and colimits follow:

- (A) *The Pullback.* This is the limit of a diagram $D : I \rightarrow \mathbf{C}$ where $I = \{a, b, c\}$, partially ordered by $a < c$ and $b < c$. (Thus, the only morphisms in I are the identities, plus exactly two more; one $m : a \rightarrow c$ and another $m' : b \rightarrow c$. As was pointed out in 4.3.4(c) in reference to equalizers, if $D(m) : D(a) \rightarrow D(c)$ and $D(m') : D(b) \rightarrow D(c)$ is a diagram, then a source consists of three morphisms $(f_i : A \rightarrow D(i))_{i \in \{a, b, c\}}$ such that $f(c) = D(m) \cdot f_a = D(m') \cdot f_b$. In practice, however, mention of $f(c)$ is dropped.

Here is a diagram highlighting a pullback; it depicts a typical source. That is, the picture in question is a commutative square.

$$\begin{array}{ccc}
 P & \xrightarrow{p_a} & D(a) \\
 \downarrow p_b & & \downarrow D(m) \\
 D(b) & \xrightarrow{D(m')} & D(c)
 \end{array}$$

- (B) *The Pushout.* This is the colimit of a diagram $D : I^{\text{op}} \rightarrow \mathbf{C}$, where I is as in (A); that is to say, the pushout is the dual of the pullback.
- (C) *Multiple Equalizer.* The limit of a diagram $D : I \rightarrow \mathbf{C}$, where $I = \{i, j\}$, having the identities, and with *no* morphisms $j \rightarrow i$. (The number of morphisms $i \rightarrow j$ is arbitrary.) The dual notion is the *multiple coequalizer*.

One needs a result which tells us that in order to tell that a category is complete it suffices to test a relatively small number and only certain types of limits. The main theorem on this theme follows. The Proof is accomplished through two lemmas, Lemmas 4.3.10 and 4.3.12.

Theorem 4.3.8. *A category \mathbf{C} is complete if and only if it has products and equalizers. Dually, \mathbf{C} is cocomplete if and only if it has coproducts and coequalizers.*

As products and equalizers are special cases of limits, the necessity is obvious. Likewise, since \mathbf{C} is complete if and only if \mathbf{C}^{op} is cocomplete, each one of the claims implies the other.

To prove the sufficiency, it is useful to bring in pullbacks, and formulate matters this way; we leave it to the reader to state the dual theorem, involving pushouts.

Theorem 4.3.9. *Let \mathbf{C} be a category; the following are equivalent:*

- (a) \mathbf{C} is complete.
- (b) \mathbf{C} has products and equalizers.
- (c) \mathbf{C} has products and pullbacks.

We will show that (b) implies (c), and that (c) implies (a).

First, let us say that \mathbf{C} has *finite products* if the product of any finite set exists. For example, the category \mathbf{FGr} of finite groups has finite products, but, obviously, not arbitrary products.

Lemma 4.3.10. *If \mathbf{C} has finite products and equalizers, then it also has pullbacks.*

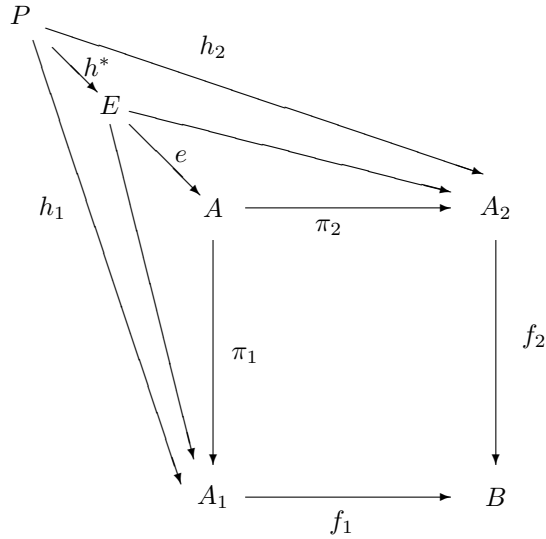
Proof. The Proof consists of constructing the pullback as an equalizer of two morphism out of a product of two objects.

Consider morphisms $f_1 : A_1 \rightarrow B$ and $f_2 : A_2 \rightarrow B$. Let A be the product of A_1 and A_2 , with the respective projections π_1 and π_2 . Now form $f_1 \cdot \pi_1$ and $f_2 \cdot \pi_2$, which have A as domain and B as codomain. Let (E, e) be the equalizer of the $f_i \cdot \pi_i$ ($i = 1, 2$).

Before proceeding any further, consider the diagram which follows, for reference. Let

$$\{h_1 : P \rightarrow A_1, h_2 : P \rightarrow A_2\}$$

be a source for the diagram made up of f_1 and f_2 . We claim that $\pi_1 \cdot e : E \rightarrow A_1$ and $\pi_2 \cdot e : E \rightarrow A_2$ is the universal source.



(Note: the squares bounded by P, A_1, A_2 and B , and by E, A_1, A_2 and B commute. So do the triangles formed by P, E and A_1 and by P, E and A_2 . The square bounded by A, A_1, A_2 and B , however, does not necessarily commute, which is the point, right?)

Continuing now, by definition of the product, there is a morphism $\tilde{h} : P \rightarrow A$ for which $\pi_i \cdot \tilde{h} = h_i$ ($i = 1, 2$). Observe next that

$$(f_1 \cdot \pi_1) \cdot \tilde{h} = (f_2 \cdot \pi_2) \cdot \tilde{h}.$$

By definition of equalizer there is a unique morphism $h^* : P \rightarrow E$ such that $e \cdot h^* = \tilde{h}$. But then $(\pi_1 \cdot e) \cdot h^* = \pi_1 \cdot \tilde{h}$ and $\pi_2 \cdot \tilde{h} = (\pi_2 \cdot e) \cdot h^*$, proving that the pair $\{\pi_1 \cdot e, \pi_2 \cdot e\}$ is indeed a universal source for the pullback diagram. ■

Before stating the next lemma some preliminaries are in order.

Definition & Remarks 4.3.11. Let $D : I \longrightarrow \mathbf{C}$ be a diagram in \mathbf{C} ; put $P = \prod_{i \in I} D(i)$, with projections π_i . Let M be the set of morphisms in I , and P^M denote the product of M copies of P ; suppose that q_m stands for the projection on the m -th component, with $m \in M$. Continuing, let $\Delta : P \longrightarrow P^M$ be the *diagonal morphism*; that is, the one defined by $q_m \cdot \Delta = 1_P$, for every $m \in M$, and whose existence is guaranteed if \mathbf{C} has products.

Finally, let $\mu : P \longrightarrow P^M$ denote the following morphism. To define such a morphism, it suffices to define $q_m \cdot \mu$, for each $m \in M$. If $m : j \longrightarrow j'$, then, for $i \in I$, $i \neq j'$, $\pi_i \cdot (q_m \cdot \mu) = \pi_i$, while

$$\pi_{j'} \cdot (q_m \cdot \mu) = D(m) \cdot \pi_j.$$

Lemma 4.3.12. (Freyd) *With the notation established in the above preamble, assume that \mathbf{C} is a category having products and pullbacks. Form the pullback of Δ and μ :*

$$\begin{array}{ccc} L & \xrightarrow{\mu'} & P \\ \Delta' \downarrow & & \downarrow \Delta \\ P & \xrightarrow{\mu} & P^M \end{array}$$

Assuming that no object of I is the codomain of every morphism in M , then $(\pi_i \cdot \Delta' : L \longrightarrow D(i))_{i \in I}$ is a universal source for the diagram D .

Proof. First, let's show that the family of morphisms is a source for D . In fact, it turns out that $\Delta' = \mu'$; first, for any $n \in I$ not having i as codomain,

$$\pi_i \cdot \Delta' = \pi_i \cdot q_n \cdot \mu \cdot \Delta' = \pi_i \cdot q_n \cdot \Delta \cdot \mu' = \pi_i \cdot \mu'.$$

(And we have used the rather peculiar property of I already!) By the uniqueness in the definition of the product, $\Delta' = \mu'$. (Note: the above diagram is, after all, an equalizer in disguise!)

Suppose now that $m : i \longrightarrow j$ is a morphism in I . Then,

$$D(m) \cdot \pi_i \cdot \Delta' = \pi_j \cdot q_m \cdot \mu \cdot \Delta' = \pi_j \cdot q_m \cdot \Delta \cdot \mu' = \pi_j \cdot \mu' = \pi_j \cdot \Delta',$$

and so the set of morphisms $(\pi_i \cdot \Delta' : L \longrightarrow D(i))_{i \in I}$ is a source.

Next, suppose that $(g_i : A \longrightarrow D(i))_i$ is a source for D . Let $\hat{g} : A \longrightarrow P$ be the unique morphism for which $\pi_i \cdot \hat{g} = g_i$, for every index i . Let's compare $\mu \cdot \hat{g}$ and $\Delta \cdot \hat{g}$. For each $m : i \longrightarrow j$, and $k \in I$, $k \neq j$,

$$\pi_k \cdot q_m \cdot \mu \cdot \hat{g} = \pi_k \cdot \hat{g} = g_k,$$

while

$$\pi_j \cdot q_m \cdot \mu \cdot \hat{g} = D(m) \cdot \pi_i \cdot \hat{g} = D(m) \cdot g_i = g_j.$$

On the other hand,

$$\pi_k \cdot q_m \cdot \Delta \cdot \hat{g} = \pi_k \cdot \hat{g} = g_k,$$

for every index k . Thus, $\mu \cdot \hat{g} = \Delta \cdot \hat{g}$, whence, by definition of pullbacks, it follows that for some $h : A \rightarrow L$, $\Delta' \cdot h = \hat{g}$; from that we get (for each $i \in I$) that $\pi_i \cdot \Delta' \cdot h = g_i$.

There remains to show, in the above proceedings, that the morphism h is (all the way) uniquely determined by the g_i ; we leave this bit of work to the reader. ■

But wait!!! ...

Hot Air 4.3.13. *Are We Done?* What about the peculiar assumption in Lemma 4.3.12? The reader should realize that one may always assume, without loss of generality, that the small category I may be taken with the following property: for each $i \in I$, there is a morphism of which i is not the codomain. The reason is that if I does not have this property, then one can enlarge I to a small category I' which does have this feature, and extend any diagram $D : I \rightarrow \mathbf{C}$ to $D' : I' \rightarrow \mathbf{C}$, so that D has a limit if and only if D' does, and the limits agree. Here's how.

If the object $i \in I$ is the codomain of every morphism in I , then enlarge I as follows: adjoin one new object v , the identity on v , and exactly two more morphisms $\alpha : i \rightarrow v$ and $\beta : v \rightarrow i$. Extend D by setting $D'(v) = D(i)$, and $D'(\alpha) = D'(\beta) = 1_{D(i)}$. The reader can then easily verify that D and D' yield the same limits; (one has a universal source precisely when the other does, and they agree.) Lemma 4.3.12 now applies to D' .

Lemma 4.3.12 then, evidently, shows that (c) implies (a) in Theorem 4.3.9; Lemma 4.3.10 says that (b) implies (c). The Proof of Theorem 4.3.9 is therefore done, which also settles Theorem 4.3.8.

Let us apply the foregoing to algebras. What follows is true, in particular, for equational classes.

Theorem 4.3.14. *Suppose that \mathbf{C} is a full subcategory of $\mathbf{Al}(T)$ (for some type T), so that $\text{obj}(\mathbf{C})$ is both S - and R -closed. Then \mathbf{C} is complete and cocomplete.*

Proof. That \mathbf{C} is complete follows, because it clearly has products, and (by Exercise 4.2.13) equalizers.

We outline the proof of the claim that \mathbf{C} has coproducts, and leave the rest as exercise. Compare what follows with the proof of Theorem 4.2.4.

Suppose that $\{A_i : i \in I\}$ is a family of algebras in \mathbf{C} . Let X be the disjoint union of the A_i , and consider $W = W(X, T)$. Denote the universal embedding of X in W by w . Now, let θ be the least congruence containing all the pairs of the form

- (a) $(w(\phi_{A_i}), \phi_W)$, for each nullary operator $\phi \in T$;
- (b) for each $n \geq 1$, and each n -ary operation $\phi \in T$, and each choice $x_1, x_2, \dots, x_n \in A_i$, all pairs $(w(\phi(x_1, \dots, x_n)), \phi(w(x_1), \dots, w(x_n)))$.

Let $A = W/\theta$ and, for each $i \in I$, set $\delta_i : A_i \rightarrow A$ to be $\delta_i(a) = \theta[w(a)]$. The reader can easily verify that $(\delta_i : A_i \rightarrow A)_{i \in I}$ is a universal sink, making it the coproduct in $\mathbf{Al}(T)$.

However, we want a coproduct in \mathbf{C} . So proceed as follows. By R -closure, we may form the verbal congruence: let β_A be the least congruence of A , for which $\tilde{A} \equiv A/\beta_A \in \text{obj}(\mathbf{C})$. Define (for each $i \in I$) $\hat{\delta}_i : A_i \rightarrow \tilde{A}$ to be the composite of δ_i and the canonical homomorphism $A \rightarrow \tilde{A}$. Then $(\hat{\delta}_i : A_i \rightarrow \tilde{A})_{i \in I}$ is the coproduct sink in \mathbf{C} . Let's check this.

Suppose that $(g_i : A_i \rightarrow B)_{i \in I}$ is a sink in \mathbf{C} . There is a unique T -homomorphism $g : A \rightarrow B$ such that $g \cdot \delta_i = g_i$, for each index i . Furthermore, by the Induced Homomorphism Theorem, $A/\ker(g)$ is isomorphic to a subalgebra of B (namely $g(A)$), whence $\beta_A \leq \ker(g)$, owing to the

minimality of β_A . Appealing once more to the Induced Homomorphism Theorem – verify! – we obtain a unique homomorphism $\tilde{g} : \tilde{A} \rightarrow B$ such that $\tilde{g}(\beta_A[a]) = g(a)$. Finally, observe that

$$\tilde{g} \cdot \hat{\delta}_i = g \cdot \delta_i = g_i,$$

for each index i , which finishes the argument. ■

Observe that Theorem 4.3.14 supersedes Theorem 4.2.4, proving that direct limits exist under the milder hypotheses of S - and R -closure.

Now a finitistic version of Theorems 4.3.8 and 4.3.9. Say that the category \mathbf{C} is *finitely complete* if for each finite category I , every diagram $D : I \rightarrow \mathbf{C}$ has a limit. The dual concept, *finitely cocomplete*, is defined similarly. Likewise, if for each finite set I , and every set $\{A_i : i \in I\}$ of \mathbf{C} -objects, their product exists, we say that \mathbf{C} *has finite products*. (The dual concept is: \mathbf{C} *has finite coproducts*.)

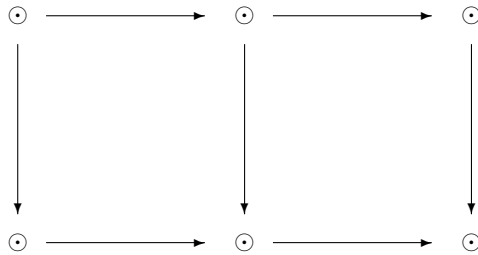
Finally, an object X in \mathbf{C} is *terminal* if, for each \mathbf{C} -object A , $|\text{Hom}(A, X)| = 1$. Observe that any two terminal objects are isomorphic.

Exercise 4.3.15. For a category \mathbf{C} the following are equivalent.

- (a) \mathbf{C} is finitely complete.
- (b) \mathbf{C} has finite products and equalizers.
- (c) \mathbf{C} has finite products and pullbacks.
- (d) \mathbf{C} has pullbacks and a terminal object.

Finally, an illustration which says, more or less, that two back-to-back pullbacks define a pullback. A dual result holds for pushouts, of course.

Exercise 4.3.16. In the diagram below, suppose the two squares represent pullbacks. Then the outer rectangle is also a pullback.



5. Categories: Adjoint Functors

This chapter examines the special relationship of some pairs of functors, known as adjointness. Adjoint functors have noteworthy properties: one member of the adjoint pair will preserve limits, the other colimits. We begin, therefore, with a study of limit preservation; an introductory account of adjoint functors follows. The point of bothering with the question of when functors preserve limits is that much of the homological calculus involving modules reflects the adjointness of functors. We devote the final section of this chapter to the Hom and tensor functors, two examples which will concern us, one way or another, for a major portion of the term.

5.1 Properties Preserved by Functors. We present a brief account of limit- and colimit-preserving functors, in which the proofs of Lemmas 4.3.10 and 4.3.12 play a prominent role. **Please note that the discussion is restricted to covariant functors.** The reader is reminded that any contravariant functor $F : \mathbf{C} \rightarrow \mathbf{D}$ has a covariant counterpart, defined on the opposite category, which is formally denoted by $F^{\text{op}} : \mathbf{C}^{\text{op}} \rightarrow \mathbf{D}$.

Definition 5.1.1. Suppose that I is a small category. Let F be a functor from \mathbf{C} to \mathbf{C}' . We say that F is I -limit-preserving if for each diagram $D : I \rightarrow \mathbf{C}$, for which $(f_i : L \rightarrow D(i))_{i \in I}$ is a universal source, then $(F(f_i) : F(L) \rightarrow F(D(i)))_{i \in I}$ is a universal source for the diagram $F \cdot D : I \rightarrow \mathbf{C}'$. F is *limit-preserving* (or *preserves limits*) if it is I -limit-preserving, for every small category I .

If $F : \mathbf{C} \rightarrow \mathbf{C}'$ is I -limit-preserving, for every category I , the only morphisms of which are the identities, then F is said to be *product-preserving* (or is said to *preserve products*).

The notion of being I -colimit-preserving, *colimit-preserving*, and *coproduct-preserving* are defined dually.

Examples 5.1.2. (a) To say that $F : \mathbf{C} \rightarrow \mathbf{C}'$ *preserves equalizers* is to say that F is I -limit preserving, whenever $I = \{i, j\}$ with only identities as morphisms, plus a pair of morphisms $i \rightarrow j$.

(b) F *preserves pullbacks* if it preserves I -limits, for each $I = \{a, b, c\}$, with identity morphisms, plus two more morphisms, one $a \rightarrow c$ and another $b \rightarrow c$.

The reader should work out the duals of the examples in (a) and (b).

From the arguments in Lemmas 4.3.10 and 4.3.12 one is able to conclude the following theorem.

Theorem 5.1.3. *For a functor $F : \mathbf{C} \rightarrow \mathbf{C}'$, the following statements are equivalent:*

- (a) F *preserves limits.*
- (b) F *preserves products and equalizers.*
- (c) F *preserves products and pullbacks.*

Dually, the corresponding three statements for colimits are also equivalent.

Proof. That (a) implies (b) is trivial. Lemma 4.3.10 shows that a pullback is an equalizer of a (finite) product, which says that (b) implies (c). Lemma 4.3.12 demonstrates that any limit can be viewed as a pullback of suitably chosen products, whence (a) follows from (c). ■

One has the finitistic analogue of Theorem 5.1.3 (as well as its dual for colimits). The functor F *preserves finite limits* if it is I -limit-preserving, for every finite category I . To *preserve finite products* is to preserve all products over finite sets.

Theorem 5.1.4. *The following are equivalent for a functor $F : \mathbf{C} \rightarrow \mathbf{C}'$:*

- (a) F preserves finite limits.
- (b) F preserves finite products and equalizers.
- (c) F preserves finite products and pullbacks.
- (d) F preserves pullbacks and terminal objects.

It is a fair question to ask how common it is for a functor to preserve a certain kind of limit. Our presentation is intended, in part, to answer this question. Many of the functors we have already considered preserve either limits or colimits. Functors which preserve both, however, seem to be rare.

5.2 Adjoint Functors. We have already encountered a number of functor pairs in “adjoint” situations. The goal of this section is to codify the phenomenon, and revisit those examples. In terms of natural transformations the definition can be given very straightforwardly. Once again we stick to covariant functors.

Definition 5.2.1. Suppose that \mathbf{C} and \mathbf{C}' are categories. Consider functors $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$. The pair (F, G) is called *an adjunction* – or *an adjoint situation*, or *an adjoint pair* – if there exist natural transformations $\sigma : 1_{\mathbf{C}} \rightarrow G \cdot F$ and $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$, satisfying the following conditions:

(adj1) For each $A \in \text{obj}(\mathbf{C}')$, the composite

$$G(A) \xrightarrow{\sigma_{G(A)}} G(F(G(A))) \xrightarrow{G(\tau_A)} G(A),$$

is the identity morphism on $G(A)$; that is,

$$G(\tau_A) \cdot \sigma_{G(A)} = 1_{G(A)}.$$

(adj2) The composite

$$F(X) \xrightarrow{F(\sigma_X)} F(G(F(X))) \xrightarrow{\tau_{F(X)}} F(X),$$

is the identity morphism on $F(X)$, for each \mathbf{C} -object X ; that is to say,

$$\tau_{F(X)} \cdot F(\sigma_X) = 1_{F(X)}.$$

The above adjoint situation is encapsulated as a quadruple (F, G, σ, τ) , and it is also common to write: $F \dashv G$. In the adjoint situation (F, G, σ, τ) , the functor F is frequently called the *front* (or *left adjoint*), while G is referred to as the *back* (or *right adjoint*). Likewise, the natural transformations σ and τ are called the *front* and *back adjunctions*, respectively.

Let us now decipher this concise but cryptic business, and thereby cast it in terms that should be more familiar.

Definition & Remarks 5.2.2. Suppose that (F, G, σ, τ) is an adjunction, with $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$. Let $X \in \text{obj}(\mathbf{C})$. For each \mathbf{C} -morphism $g : X \rightarrow G(B)$, consider $g' = \tau_B \cdot F(g) : F(X) \rightarrow B$. Note that, by the natural property of σ and then (adj1),

$$G(g') \cdot \sigma_X = G(\tau_B) \cdot (G \cdot F(g)) \cdot \sigma_X = G(\tau_B) \cdot \sigma_{G(B)} \cdot g = g.$$

All of which shows that there exists a morphism $g' : F(X) \rightarrow B$ such that the triangle below commutes. The reader should have another look at the diagram following the definition of a free functor (Example 3.2.8).

$$\begin{array}{ccc}
 X & \xrightarrow{\sigma_X} & G(F(X)) & & F(X) \\
 & \searrow g & \downarrow G(g') & & \downarrow g' \\
 & & G(B) & & B
 \end{array}
 \quad (\dagger)$$

g' is unique with respect to this property, for if $G(h) \cdot \sigma_X = g$, then, owing first to the natural feature of τ , and then (adj2),

$$g' = \tau_B \cdot F(g) = \tau_B \cdot (F \cdot G(h)) \cdot F(\sigma_X) = h \cdot \tau_{F(X)} \cdot F(\sigma_X) = h.$$

We have proved half of the following theorem:

The reference (\dagger) in the formulation is to the diagram above.

Theorem 5.2.3. Suppose that $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$ are two functors, and that $\sigma : 1_{\mathbf{C}} \rightarrow G \cdot F$ and $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$ are natural transformations so that (F, G, σ, τ) is an adjunction. Then for each \mathbf{C} -morphism $g : X \rightarrow G(B)$ there exists a unique \mathbf{C}' -morphism $g' : F(X) \rightarrow B$ such that the triangle (\dagger) commutes.

Conversely, suppose that $G : \mathbf{C}' \rightarrow \mathbf{C}$ is a functor with the property that for each \mathbf{C} -object X there is a \mathbf{C}' -object $F(X)$ and a \mathbf{C} -morphism $\sigma_X : X \rightarrow G(F(X))$, so that for each \mathbf{C} -morphism $g : X \rightarrow G(B)$, there is a unique \mathbf{C}' -morphism $g' : F(X) \rightarrow B$ such that (\dagger) commutes. Then F defines a functor $\mathbf{C} \rightarrow \mathbf{C}'$ so that σ is a natural transformation, and there is also a natural transformation $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$ so that (F, G, σ, τ) is an adjoint situation.

Let's postpone the proof of the second statement in Theorem 5.2.3, in favor of a retrospective view at some examples.

Examples 5.2.4. (a) *Free Functors.* Let (\mathbf{C}, G) be a concrete category, with underlying-set functor $G : \mathbf{C} \rightarrow \mathbf{Set}$. (For example, \mathbf{C} might be a category of algebras of a certain type T , for which $\text{obj}(\mathbf{C})$ is S - and R -closed, with at least one nontrivial algebra; then by Theorem 2.4.6, each set X has a free algebra $F(X)$ over it, in \mathbf{C} .)

In any event, suppose that the conditions of Example 3.2.8 are satisfied; to be precise, suppose that

- (i) for each set X there exists a \mathbf{C} -object $F(X)$ and a one-to-one function $u_X : X \rightarrow G(F(X))$, such that
- (ii) for each function $g : X \rightarrow G(B)$, with $B \in \text{obj}(\mathbf{C})$, there exists a unique \mathbf{C} -morphism $g^* : F(X) \rightarrow B$ so that $G(g^*) \cdot u_X = g$.

We indicated in the discussion of 3.2.8 that F , so defined is a functor. This is precisely what we will prove, shortly, to complete the proof of Theorem 5.2.3. (Also to be settled is where the “other” natural transformation $F \cdot G \rightarrow 1_{\mathbf{C}}$ comes from.) Still, that is easy to describe here: for each \mathbf{C} -object A , apply (ii) to the function $1_{G(A)}$. There is a unique \mathbf{C} -morphism $v_A : F(G(A)) \rightarrow A$, such that

$$1_{G(A)} = G(v_A) \cdot u_{G(A)}.$$

Note that this is (adj1) in the definition of adjunctions.

To sum up, for any concrete category (\mathbf{C}, G) which has a free functor, we have the adjoint situation (F, G, u, v) .

(b) *Reflections.* Suppose that \mathbf{B} is a full subcategory of \mathbf{C} and that $I : \mathbf{B} \rightarrow \mathbf{C}$ is the inclusion functor. A *reflection* is an assignment R , which pairs to each \mathbf{C} -object A , a \mathbf{B} -object $R(A)$, together with a morphism $r_A : A \rightarrow I(R(A))$, and does it universally: that is, for each morphism – no need to say in which category, since the subcategory \mathbf{B} is full – $g : A \rightarrow I(B)$ ($B \in \text{obj}(\mathbf{B})$) there is a unique morphism $g^* : R(A) \rightarrow B$ so that $I(g^*) \cdot r_A = g$.

Now the reader should convince himself that – *mutatis mutandis*, in this case meaning: with a suitable alteration of the particulars – the condition defining reflections satisfies the conditions in Theorem 5.2.3 and also diagram (†), so that (R, I, r, s) is an adjunction; s is defined in the way that v was defined above in conjunction with the free adjunction.

We have seen many examples of reflections already. Let’s recall a few specific ones:

1. The full subcategory **Abel** of **Gr**, with the reflection defined by factoring out the commutator subgroup: let $\text{Ab}(G) = G/[G, G]$ and $a_G : G \rightarrow I(\text{Ab}(G)) = I(G/[G, G])$. Then (Ab, I, a, j) is an adjunction, where j is the natural transformation which assigns to each abelian group H , the map $j_H : \text{Ab}(I(H)) \rightarrow H$, namely, the inverse of the isomorphism given by the canonical map $H \rightarrow \text{Ab}(I(H)) = H/[H, H] \cong H$, since $[H, H] = \{e\}$, when H is abelian.
2. The full subcategory **TfAb** of **Abel**, with the reflection defined by $\tau_A : A \rightarrow I(\text{Tf}(A)) = A/T(A)$, where $T(A)$ denotes the torsion subgroup of A . $(\text{Tf}, I, \tau, \alpha)$ is an adjunction, where α_G is, once more, an isomorphism: $\alpha_G : \text{Tf}(I(G)) \rightarrow G$; namely, the inverse of the isomorphism (with G torsion free) $G \rightarrow \text{Tf}(I(G)) = G/T(G) \cong G$, since $T(G) = \{0\}$.
3. The full subcategory **SP** of **CRn1**. Recall that **SP** is the subcategory of all commutative rings with identity, having no nonzero nilpotent elements. Recall Exercise 3.2.12. The reflection is defined by

$$\mu_A : A \rightarrow I(\hat{A}) = I(A/n(A))$$

where $n(A)$ is the set of all nilpotent elements and $\hat{A} = A/n(A)$. Then $(\hat{\ }, I, \mu, \delta)$ is an adjunction, where δ is defined, much as in the previous two examples: for each **SP**-object S , $\delta_S : I(\hat{S}) \rightarrow S$ is the inverse of the isomorphism $S \rightarrow I(\hat{S}) = S/n(S) \cong S$, since $n(S) = \{0\}$.

Do reflect upon what the above three examples have in common: in all three the subcategory has an object class which is S - and R -closed, and in each case the reflection is being defined by factoring out a verbal congruence!

We shall return with additional examples; under the banner of “all you wanted to know but were afraid to ask”, have a look at [HS79], pp. 197–99.

Proof. Of the second part of Theorem 5.2.3. Suppose that $G : \mathbf{C}' \rightarrow \mathbf{C}$ has all the properties given in the statement. We use the notation of the theorem.

First, we prove the functoriality of the assignment F . Suppose that $g : X \rightarrow Y$ is a \mathbf{C} -morphism, and let $h = \sigma_Y \cdot g$. Then there is a unique \mathbf{C}' -morphism $g^* : F(X) \rightarrow F(Y)$ such that

$$(\#) \quad G(g^*) \cdot \sigma_X = h = \sigma_Y \cdot g.$$

Set $F(g) = g^*$; this is well defined. Now, let's establish that it's a functorial definition: if $g = 1_X$ then $g^* = 1_{F(X)}$ satisfies $(\#)$. So, by the uniqueness, $F(1_X) = 1_{F(X)}$. Next, if $g : X \rightarrow Y$ and $h : Y \rightarrow Z$ are \mathbf{C} -morphisms then

$$\begin{aligned} G(F(h) \cdot F(g)) \cdot \sigma_X &= G(F(h)) \cdot G(F(g)) \cdot \sigma_X \\ &= G(F(h)) \cdot \sigma_Y \cdot g \\ &= \sigma_Z \cdot h \cdot g \\ &= G(F(h \cdot g)) \cdot \sigma_X, \end{aligned}$$

and, once more by virtue of the uniqueness: $F(h) \cdot F(g) = F(h \cdot g)$.

Now it should also be clear from $(\#)$ that $\sigma : 1_{\mathbf{C}} \rightarrow G \cdot F$ is a natural transformation.

The next step is to produce the back adjunction $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$. To $1_{G(A)}$ we apply the assumptions: there is a unique \mathbf{C}' -morphism $\tau_A : F \cdot G(A) \rightarrow A$ so that

$$G(\tau_A) \cdot \sigma_{G(A)} = 1_{G(A)},$$

which, incidentally, is condition (adj1). Verify that τ is a natural transformation: suppose that $f : A \rightarrow B$ is a \mathbf{C}' -morphism. On the one hand:

$$G(f \cdot \tau_A) \cdot \sigma_{G(A)} = G(f) \cdot G(\tau_A) \cdot \sigma_{G(A)} = G(f),$$

while

$$\begin{aligned} G(\tau_B \cdot F(G(f))) \cdot \sigma_{G(A)} &= G(\tau_B) \cdot G(F(G(f))) \cdot \sigma_{G(A)} \\ &= G(\tau_B) \cdot \sigma_{G(B)} \cdot G(f) \\ &= 1_{G(B)} \cdot G(f) \\ &= G(f). \end{aligned}$$

By uniqueness, $f \cdot \tau_A = \tau_B \cdot F(G(f))$, which proves that $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$ is a natural transformation.

Since (adj1) has already been shown, all that remains is (adj2). But

$$G(\tau_{F(X)} \cdot F(\sigma_X)) \cdot \sigma_X = G(\tau_{F(X)}) \cdot (G(F(\sigma_X)) \cdot \sigma_X) = G(\tau_{F(X)}) \cdot \sigma_{G(F(X))} \cdot \sigma_X,$$

since σ is a natural transformation. Continuing now, and using (adj1),

$$G(\tau_{F(X)}) \cdot \sigma_{G(F(X))} \cdot \sigma_X = 1_{G(F(X))} \cdot \sigma_X = G(1_{F(X)}) \cdot \sigma_X.$$

Yet one more time the use of uniqueness: $\tau_{F(X)} \cdot F(\sigma_X) = 1_{F(X)}$, which proves (adj2).

This completes the proof of Theorem 5.2.3. ■

By dualizing, it should become clear that any adjunction (F, G, σ, τ) is symmetric. Thus, there is a dual to Theorem 5.2.3, which we now state, without further comment.

Theorem 5.2.5. *Suppose that $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$, so that (F, G, σ, τ) is an adjunction. Then, for each \mathbf{C}' -object A , and each \mathbf{C}' -morphism $h : F(Y) \rightarrow A$, there exists a unique \mathbf{C} -morphism $h^* : Y \rightarrow G(A)$, so that the following diagram commutes:*

$$\begin{array}{ccc}
 A & \xleftarrow{\tau_A} & F(G(A)) & & G(A) \\
 & \swarrow h & \uparrow F(h^*) & & \uparrow h^* \\
 & & F(Y) & & Y
 \end{array}$$

(††)

that is, $\tau_A \cdot F(h^*) = h$.

Conversely, suppose $F : \mathbf{C} \rightarrow \mathbf{C}'$ is a functor, so that for each \mathbf{C}' -object A there exists a \mathbf{C} -object $G(A)$, together with a \mathbf{C}' -morphism $\tau_A : F(G(A)) \rightarrow A$, so that for each \mathbf{C}' -morphism $h : F(Y) \rightarrow A$, there is a unique \mathbf{C} -morphism $h^* : Y \rightarrow G(A)$, such that (††) commutes, then $G : \mathbf{C}' \rightarrow \mathbf{C}$ defines a functor, $\tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$ a natural transformation, and there is a natural transformation $\sigma : 1_{\mathbf{C}} \rightarrow G \cdot F$ so that (F, G, σ, τ) is an adjoint situation.

Hot Air 5.2.6. *Dotting i's, etc.* (a) Suppose that \mathbf{B} is a full subcategory of \mathbf{C} , and $R : \mathbf{C} \rightarrow \mathbf{B}$ is a reflection; let $I : \mathbf{B} \rightarrow \mathbf{C}$ stand for the inclusion functor. Then we have an adjunction (R, I, r, s) (by Theorem 5.2.3). The back adjunction $s_B : R(I(B)) \rightarrow B$ turns out to be an isomorphism. For, according to (adj1), for each \mathbf{B} -object B , $s_B : R(I(B)) \rightarrow B$ is the unique morphism such that $I(s_B) \cdot r_{I(B)} = 1_{I(B)}$. (Uniqueness follows from (†). Note as well that I is the identity on both objects and morphisms from \mathbf{B} . We therefore omit its mention.) Hence, $r_B = r_B \cdot s_B \cdot r_B$, and by the uniqueness in (†), $r_B \cdot s_B = 1_{R(I(B))}$. Thus, s_B is an isomorphism.

(b) Here's the dual to the notion of a reflection: again \mathbf{B} is a full subcategory of \mathbf{C} , with I as the inclusion functor; we say that $\gamma : \mathbf{C} \rightarrow \mathbf{B}$ is a *coreflection* if (I, γ, j, δ) is an adjunction so that the front adjunction $j : 1 \rightarrow \gamma \cdot I$ is a natural equivalence.

Let's feature the natural transformation δ and the formulation of Theorem 5.2.5. For each $B \in \text{obj}(\mathbf{B})$, and each morphism $f : A \rightarrow B$ – again we suppress explicit mention of I – there is a unique morphism $f^* : A \rightarrow \gamma(B)$ so that $\delta_B \cdot f^* = f$.

Here's a particular example:

Exercise 5.2.7. Take **Tor**, the full subcategory of **Abel** of all torsion groups. Let $T(G)$ denote the torsion subgroup of the abelian group G . Show that (I, T, j, t) is a coreflection, for suitable choices of j and t . Describe the two.

Exercise 5.2.8. Let us suppose that $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$ are functors, and σ and τ are natural transformations, such that (F, G, σ, τ) is an adjunction. Show that $\text{Hom}_{\mathbf{C}'}(F(X), B)$ and $\text{Hom}_{\mathbf{C}}(X, G(B))$ are naturally isomorphic; that is, for each $X \in \text{obj}(\mathbf{C})$ and $B \in \text{obj}(\mathbf{C}')$ there is an isomorphism (that is to say, a bijection, since the map is taken as a morphism of **Set**)

$$\alpha_{X,B} : \text{Hom}_{\mathbf{C}'}(F(X), B) \rightarrow \text{Hom}_{\mathbf{C}}(X, G(B)),$$

so that if $g : X' \rightarrow X$ is a \mathbf{C} -morphism and $f : B \rightarrow B'$ a \mathbf{C}' -morphism, then the following square commutes:

$$\begin{array}{ccc}
\mathrm{Hom}_{\mathbf{C}'}(F(X), B) & \xrightarrow{\alpha_{X,B}} & \mathrm{Hom}_{\mathbf{C}}(X, G(B)) \\
\downarrow \mathrm{Hom}_{\mathbf{C}'}(F(g), f) & & \downarrow \mathrm{Hom}_{\mathbf{C}}(g, G(f)) \\
\mathrm{Hom}_{\mathbf{C}'}(F(X'), B') & \xrightarrow{\alpha_{X',B'}} & \mathrm{Hom}_{\mathbf{C}}(X', G(B'))
\end{array}$$

where $\mathrm{Hom}_{\mathbf{C}'}(F(g), f)(h) \equiv f \cdot h \cdot F(g)$, and $\mathrm{Hom}_{\mathbf{C}}(g, G(f))(k) \equiv G(f) \cdot k \cdot g$.

(Hint: the bijective part should flow directly from Theorems 5.2.3 and 5.2.5. The rest is dreary, but straightforward.)

For the next exercise the reader ought to refer to Exercise 3.3.8. It concerns the process of adjoining an identity to a ring.

Exercise 5.2.9. Recall the inclusion of **Rn1** in **Rn**. The first is not full in the second, as has already been noted. In the notation of Exercise 3.3.8, verify that $((\)^*, I, \delta, \varepsilon)$ is an adjoint situation. Note that the back adjunction ε is new; it does not appear in the earlier exercise.

We have not said anything about uniqueness in the adjoint situation itself. Let us close the section with that. The result which follows is a direct consequence of the uniqueness provisions in Theorems 5.2.3 and 5.2.5. We shall leave the proofs as exercises for the reader.

Proposition 5.2.10. (a) Suppose that (F, G, σ, τ) and (F, G', σ', τ') are adjoint situations. Then the functors G and G' are naturally equivalent. (Dually, by keeping G constant.)

(b) Conversely, suppose that (F, G, σ, τ) is an adjunction, and $G \cong G'$; then there exist natural transformations $\sigma' : 1 \rightarrow G' \cdot F$ and $\tau' : F \cdot G' \rightarrow 1$ such that (F, G', σ', τ') is an adjunction.

5.3 Adjointness and Preservations of Limits. This section finally lays out the material we set out to study. The goal of this chapter is to prove the following theorem:

Theorem 5.3.1. Suppose that $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$ are two functors so that (F, G, σ, τ) is an adjunction. Then G preserves limits and F preserves colimits.

Proof. By passing to the dual categories, it should be clear that we need only prove one of the claims. We show that F preserves colimits.

Suppose that $D : I \rightarrow \mathbf{C}$ is a diagram, and that a universal sink is given by $(d_i : D(i) \rightarrow K)_{i \in I}$. That $(F(d_i) : F(D(i)) \rightarrow F(K))_{i \in I}$ is a sink for $F \cdot D$ should be obvious.

Now, let $(f_i : F(D(i)) \rightarrow B)_{i \in I}$ be a sink for $F \cdot D$. Next, consider the sink $(G(f_i) \cdot \sigma_{D(i)} : D(i) \rightarrow G(B))_{i \in I}$. There is a unique \mathbf{C} -morphism $h : K \rightarrow G(B)$ so that $h \cdot d_i = G(f_i) \cdot \sigma_{D(i)}$, for each $i \in I$. By Theorem 5.2.3, there is a unique \mathbf{C}' -morphism $f : F(K) \rightarrow B$ for which $G(f) \cdot \sigma_K = h$. Composing with d_i , and applying the naturality of σ ,

$$G(f \cdot F(d_i)) \cdot \sigma_{D(i)} = G(f_i) \cdot \sigma_{D(i)},$$

whence $f \cdot F(d_i) = f_i$, and this holds for each $i \in I$, proving the universality of the $F(d_i)$ for the diagram $F \cdot D$. ■

Now comes the fun; extracting out of this theorem, and all that went before, special features of the examples we shall have occasion to use many times over. The full-flesh treatment of Hom functors is postponed until the introduction of tensor products in the next section.

Most of the corollaries that follow will be stated with little or no comment.

Corollary 5.3.2. *Suppose that (\mathbf{C}, G) is a concrete category, having a free functor F . Then:*

- (a) *the underlying-set functor preserves limits;*
- (b) *if X is the disjoint union of subsets X_i ($i \in I$), then $F(X)$ is the coproduct in \mathbf{C} of the $F(X_i)$.*

Proof. (a) follows directly from Theorem 5.3.1; (b) as well, after observing that the disjoint union is the coproduct in **Set**. ■

Corollary 5.3.3. *Suppose that \mathbf{B} is a full subcategory of \mathbf{C} . Assume that there is a reflection $R : \mathbf{C} \rightarrow \mathbf{B}$. Then R preserves all colimits and the inclusion functor preserves all limits.*

(Note: But the colimit of objects in \mathbf{B} , treated as objects in \mathbf{C} , may not agree with their colimit in \mathbf{B} . See Corollary 5.3.5(1).)

The inclusion of a full subcategory in another preserves limits, if there is a reflection going in reverse. But much more can be said: the subcategory is closed under formation of limits.

Proposition 5.3.4. *Suppose that \mathbf{B} is a full subcategory of \mathbf{C} , and that $R : \mathbf{C} \rightarrow \mathbf{B}$ is a reflection. Then if $D : I \rightarrow \mathbf{B}$ is a diagram, and $\{p_i : L \rightarrow D(i)\}$ is a universal source for D in the category \mathbf{C} , then $L = R(L)$, and $L \in \text{obj}(\mathbf{B})$. (Briefly put, any limit in \mathbf{C} of objects in \mathbf{B} in fact lies in \mathbf{B} .)*

Proof. Since $D(i) \in \text{obj}(\mathbf{B})$, for each $i \in I$, applying the functor R to the source identity $p_j \cdot D(m) = p_i$, for each $i \in I$ and morphism $m : i \rightarrow j$, yields that $R(p_j) \cdot D(m) = R(p_i)$. (Recall that the composition of R upon the inclusion functor is naturally equivalent to $1_{\mathbf{B}}$. Indeed, for any \mathbf{B} -object B , r_B is an isomorphism (5.2.6(a)).)

The reader is now asked to verify that

$$(r_{D(i)}^{-1} \cdot R(p_i) : R(L) \rightarrow D(i))_i$$

is a source for D . As the given source is universal, there is a morphism $h : R(L) \rightarrow L$ such that

$$p_i \cdot h = r_{D(i)}^{-1} \cdot R(p_i),$$

for each i . Another straightforward calculation, employing uniqueness in the universality of limits, shows that $h \cdot r_L = 1_L$. Multiplying by r_L on the left, and applying uniqueness relative to the reflection, gives $r_L \cdot h = 1_{R(L)}$. ■

Some specific instances of Theorem 5.3.1.

Corollary 5.3.5. (Refer to 5.2.4(b))

- (1) *Relative to the adjunction (Ab, I, a, j) , where $A : \mathbf{Gr} \rightarrow \mathbf{Abel}$ and I is the inclusion, with $a_G : G \rightarrow G/[G, G]$,*

$$\text{Ab}(\coprod_{i \in I} G_i) = \bigoplus_{i \in I} \text{Ab}(G_i).$$

(\coprod denotes the coproduct in \mathbf{Gr} , \bigoplus the direct sum in \mathbf{Abel} .)

(2) *Relative to the adjunction $(\text{Tf}, I, \tau, \alpha)$, where $\text{Tf} : \mathbf{Abel} \rightarrow \mathbf{TFAb}$, $\tau_G : G \rightarrow G/T(G)$ is the canonical homomorphism, and $T(G)$ denotes the torsion subgroup, Tf preserves direct sums.*

Proof. One catch to (2): it has to be shown, separately, that the direct sum is the coproduct in \mathbf{TFAb} . Then one applies Corollary 5.3.2(b). ■

Remark 5.3.6. Tf does not preserve products! Let G be the product of all \mathbb{Z}_p (over all primes p); then $\text{Tf}(\mathbb{Z}_p) = \{0\}$, for each prime, while $\text{Tf}(G)$ is not zero.

Exercise 5.3.7. Let T be the functor $\mathbf{Abel} \rightarrow \mathbf{Tor}$ which computes the torsion subgroup. Show that it preserves finite products.

Exercise 5.3.8. Let \mathbf{FGAb} denote the full subcategory of \mathbf{Abel} , of all the finitely generated abelian groups. Prove that there can be no reflection of \mathbf{Abel} in \mathbf{FGAb} .

Hot Air 5.3.9. Contravariance and Adjointness. We have scrupulously avoided contravariant functors (at least, explicitly). Nonetheless, some of the very important examples of adjunctions involve contravariant functors; (the pair involved in the Stone duality, \mathfrak{B} and Max , for instance.)

So what should be made clear here is what is meant by an adjoint situation when $F : \mathbf{C} \rightarrow \mathbf{C}'$ and $G : \mathbf{C}' \rightarrow \mathbf{C}$ are contravariant. Pass to $F^{\text{op}} : \mathbf{C} \rightarrow \mathbf{C}'^{\text{op}}$ and ${}^{\text{op}}G : \mathbf{C}'^{\text{op}} \rightarrow \mathbf{C}$, which are covariant. Say that F and G form an adjoint pair if $(F^{\text{op}}, {}^{\text{op}}G, \sigma, \tau)$ is an adjunction.

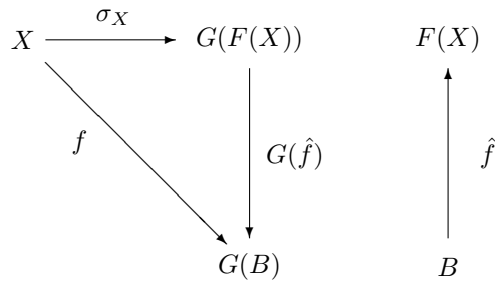
To be precise (as well as explicit), let's recite Theorem 5.2.3 for this context: for each \mathbf{C} -object X , there is a \mathbf{C} -morphism

$$\sigma : X \rightarrow {}^{\text{op}}G(F^{\text{op}}(X)) = G(F(X)),$$

so that if $f : X \rightarrow {}^{\text{op}}G(B) = G(B)$ is any \mathbf{C} -morphism, then there is a unique \mathbf{C}'^{op} -morphism $\hat{f} : F^{\text{op}}(X) \rightarrow B$, (which is to say, a \mathbf{C}' -morphism $\hat{f} : B \rightarrow F(X)$) such that

$$G(\hat{f}) \cdot \sigma = {}^{\text{op}}G(\hat{f}) \cdot \sigma_X = f.$$

(Refer to the diagrams below; the contravariance of G is suggested, to the side, by the inversion of arrows upon application of G . All ${}^{\text{op}}$'s have been omitted.)



σ and τ remain natural transformations,

$$\sigma : 1_{\mathbf{C}} \rightarrow G \cdot F, \quad \text{and} \quad \tau : F \cdot G \rightarrow 1_{\mathbf{C}'}$$

It is, therefore, consistent to denote the adjunction of F and G as (F, G, σ, τ) .

With regard to limits and colimits, the facts are these:

Theorem 5.3.10. *Suppose that F and G are contravariant functors, and that $\sigma : 1 \rightarrow G \cdot F$ and $\tau : F \cdot G \rightarrow 1$ are natural transformations, so that (F, G, σ, τ) is an adjunction. Then, if $\{\alpha_i : D(i) \rightarrow K\}$ is a universal sink for the diagram $D : I \rightarrow \mathbf{C}$, then the source $(F(\alpha_i) : F(K) \rightarrow F(D(i)))_{i \in I}$ is universal for $F \cdot D : I^{\text{op}} \rightarrow \mathbf{C}'$.*

G has the same property.

Proof. Obvious, since $(F^{\text{op}}, {}^{\text{op}}G, \sigma, \tau)$ is an adjunction of covariant functors. To say that F^{op} preserves colimits (which it does by Theorem 5.3.1) is to do precisely what is claimed in the theorem. As to G , note that ${}^{\text{op}}G$ preserves limits, which also says that G does what is claimed. ■

Loosely speaking then, F and G take colimits to limits. For example, F (or G), applied to a coproduct, will yield a product; applied to a coequalizer, will produce an equalizer, etc.

The reader is encouraged to wax philosophical on the uniformity of the situation involving adjointness of contravariant functors. We shall have more to say about contravariant functors later on.

5.4 Hom Functors and Tensor Products. This section concerns particulars about functors of the form $\text{Hom}_{\mathbf{C}}(A, _)$, and its contravariant partner. This discussion will serve as a springboard for the remainder of the term. In most of the examples \mathbf{C} will be the category of (unital) left or right modules over a ring with identity. Presently, the tensor product will be defined, and its adjoint relationship to $\text{Hom}_{\mathbf{C}}(A, _)$ will be addressed.

Definition 5.4.1. *For the Record.* We start with, and fix, a (not-necessarily commutative) ring R with identity. We shall work in the categories \mathbf{RMod} and \mathbf{ModR} of left *unital* and, respectively, right unital R -modules and all R -homomorphisms. (The term “unital” simply recalls that $1 \cdot m = m$, for each m in the left R -module M , and $m \cdot 1 = m$, when the scalars operate on the right. 1 denotes the identity of R .)

Henceforth we will drop the adjective “unital”.

Remarks 5.4.2. *Old Hat.* The object classes of both \mathbf{RMod} and \mathbf{ModR} are equational classes (in their respective types). Both categories are complete and cocomplete. The product (in either one) is the direct product, while the coproduct is the direct sum. In \mathbf{RMod} the free R -module on the set X , $F(X)$, is the direct sum of X copies of R , with left scalar multiplication. In \mathbf{ModR} it is the same thing, but with right scalar multiplication.

Let us also resuscitate a matter brought up back in 1.5.5: the relationship between congruences and subobjects, in \mathbf{RMod} . Similar comments apply in \mathbf{ModR} . If A is a left R -module and σ is any congruence on A , then $K_\sigma = \sigma[0]$ (review the notation from Chapter 1) is a submodule of A . Conversely, if K is any submodule of A , then let

$$\sigma_K = \{ (a, b) \in A^2 : a - b \in K \}.$$

Then σ_K is a congruence on A . The correspondences $K \mapsto \sigma_K$ and $\sigma \mapsto K_\sigma$ are mutually inverse lattice isomorphisms. Moreover, for any congruence σ , A/σ (as defined in Chapter 1) and A/K_σ (traditionally interpreted, as a module of cosets), are identical.

Remarks 5.4.3. *Modus Operandi.* The game, for now, will be to establish generally applicable facts about \mathbf{RMod} and \mathbf{ModR} . So R is indeed, for now, a fixed, though otherwise arbitrary ring with 1.

We will say some intelligent things about commutative rings, now and then; it should be clear that, if R is commutative, then \mathbf{RMod} and \mathbf{ModR} are naturally equivalent categories. (It is a

triviality, but satisfy yourself on this point!) When R is commutative, we shall write $\text{Hom}_R(A, B)$ for the set of all R -homomorphisms from A to B .

We've come to one of the most important definitions contained in these notes. The tensor product is a difficult construction to interpret in practice; we shall try to do this successfully in this section and in the succeeding chapters.

Definition 5.4.4. Suppose that A is a right R -module and B is a left R -module. A map $\theta : A \times B \rightarrow G$ ($G \in \text{obj}(\mathbf{Abel})$) is said to be R -bilinear if both $\theta(a, \cdot) : B \rightarrow G$ and $\theta(\cdot, b) : A \rightarrow G$ are group homomorphisms, for each $a \in A$, and $b \in B$, such that, in addition, for each $r \in R$,

$$\theta(ar, b) = \theta(a, rb), \forall a \in A, b \in B.$$

A *tensor product* of A and B is a pair (T, τ) , where T is an abelian group, and $\tau : A \times B \rightarrow T$ an R -bilinear map such that, whenever $\theta : A \times B \rightarrow G$ is an R -bilinear map, then there is a unique group homomorphism $\theta^* : T \rightarrow G$ so that $\theta^* \cdot \tau = \theta$.

The first result on tensor products establishes the existence and uniqueness.

Theorem 5.4.5. *For each right R -module A and left R -module B , the tensor product exists, and it is unique up to a group isomorphism.*

Proof. We give the outlines. Both parts are similar to several arguments given in earlier situations.

(Existence) Consider the set $X = A \times B$, and let F be the free abelian group over X as a set. Denote the universal embedding of X in F by u . Let K be the subgroup of F , generated by all the following elements:

- (a) $u(ar, b) - u(a, rb), \forall r \in R, a \in A$ and $b \in B$;
- (b) $u(a_1 + a_2, b) - u(a_1, b) - u(a_2, b), \forall a_1, a_2 \in A, b \in B$;
- (c) $u(a, b_1 + b_2) - u(a, b_1) - u(a, b_2), \forall a \in A, b_1, b_2 \in B$.

Let $T = F/K$, and $\tau : A \times B \rightarrow T$ be the restriction of the canonical map; thus,

$$\tau(a, b) = u(a, b) + K, \forall a \in A, b \in B.$$

We leave it to the reader to verify that τ is R -bilinear.

Now if $\theta : A \times B \rightarrow G$ is an R -bilinear map into the abelian group G , denote by $g : F \rightarrow G$ the unique homomorphism for which $g \cdot u = \theta$. (Note: all underlying-set functors are henceforth suppressed.) Verify, next, that this means that $K \leq \ker(g)$. Then apply the Induced Homomorphism Theorem to conclude that there is a unique homomorphism $\theta^* : T \rightarrow G$, so that $\theta^*(y + K) = g(y)$. This means that

$$\theta^* \cdot \tau(a, b) = \theta^*(u(a, b) + K) = g(u(a, b)) = \theta(a, b),$$

and, after checking that θ determines θ^* uniquely, the existence is proved.

(Uniqueness) Suppose that (T', τ') is a pair satisfying the defining conditions of a tensor product. On the one hand, there is a group homomorphism $\alpha : T \rightarrow T'$ so that $\alpha \cdot \tau = \tau'$; likewise, we have a group homomorphism $\beta : T' \rightarrow T$ so that $\beta \cdot \tau' = \tau$. Composing, we get:

$$(\beta \cdot \alpha) \cdot \tau = \tau \quad \text{and} \quad (\alpha \cdot \beta) \cdot \tau' = \tau',$$

and by uniqueness it follows that α and β are mutually inverting group isomorphisms. ■

Definition & Remarks 5.4.6. We shall denote the tensor product of A and B by $T = A \otimes B$, and the image of the universal map τ on the pair (a, b) by $a \otimes b$.

The assignment $() \otimes B$ is a covariant functor from \mathbf{Mod}_R to \mathbf{Abel} . For if $f : A \rightarrow A'$ is a right R -homomorphism, then $\theta_f(a, b) = f(a) \otimes b$, for all $a \in A$ and $b \in B$, defines an R -bilinear map θ_f on $A \times B$, into $A' \otimes B$. This gives rise to a group homomorphism

$$f \otimes B \equiv \theta_f^* : A \otimes B \rightarrow A' \otimes B,$$

such that $(f \otimes B)(a \otimes b) = f(a) \otimes b$.

We leave it to the reader to verify the functorial properties. One should add that the assignment $A \otimes ()$ defines a covariant functor from \mathbf{RMod} to \mathbf{Abel} , in the same way.

Oh, yes, the goal:

Hot Air 5.4.7. The goal is to show that the functors

$$\mathrm{Hom}_{\mathbb{Z}}(B, \) \text{ and } () \otimes B$$

are adjoint, for each left R -module B . We still need to clarify in what sense $\mathrm{Hom}_{\mathbb{Z}}(B, \)$ is to be regarded as going from \mathbf{Abel} to \mathbf{Mod}_R . (Recall: \mathbb{Z} stands for the ring of integers. Let us recall the obvious here: namely, that \mathbb{Z} -modules are abelian groups, and vice versa.)

Let B be a left R -module and G be an abelian group. Define a group structure on $\mathrm{Hom}_{\mathbb{Z}}(B, G)$ as follows: for $f, g \in \mathrm{Hom}_{\mathbb{Z}}(B, G)$,

$$(f + g)(b) = f(b) + g(b), \quad \forall b \in B;$$

it is routine to check that $f + g$ is a group homomorphism from B into G . The zero homomorphism $0 : B \rightarrow G$ sends every element of B to 0. The additive inverse of $f \in \mathrm{Hom}_{\mathbb{Z}}(B, G)$ is $(-f)(b) = -f(b)$, for each $b \in B$. We let the reader verify that this does make of $\mathrm{Hom}_{\mathbb{Z}}(B, G)$ an abelian group.

The novelty is this: $\mathrm{Hom}_{\mathbb{Z}}(B, G)$ can be made into a right R -module, as follows: for each $r \in R$, and $f \in \mathrm{Hom}_{\mathbb{Z}}(B, G)$ define

$$(f \cdot r)(b) \equiv f(rb).$$

Let's verify a few of the axioms for scalar multiplication. First, the unital feature: $(f \cdot 1)(b) = f(b)$, for each $b \in B$, so that $f \cdot 1 = f$. Also, if $s \in R$, then

$$\begin{aligned} (f \cdot (s + r))(b) = f((s + r)b) &= f(sb + rb) = f(sb) + f(rb) \\ &= (f \cdot s)(b) + (f \cdot r)(b) \\ &= (f \cdot s + f \cdot r)(b), \end{aligned}$$

for each $b \in B$, whence $f \cdot (s + r) = f \cdot s + f \cdot r$.

This proves most of the next proposition; the rest of it is left to the reader.

Proposition 5.4.8. *Suppose that B is a left R -module. Then the assignment $\mathrm{Hom}_{\mathbb{Z}}(B, \)$ is a covariant functor from \mathbf{Abel} to \mathbf{Mod}_R .*

Proof. One has to verify that if $g : G \rightarrow G'$ is a group homomorphism, then $\mathrm{Hom}_{\mathbb{Z}}(B, g)$ is a right R -homomorphism. ■

For each right R -module A , define $s_A : A \rightarrow \mathrm{Hom}_{\mathbb{Z}}(B, A \otimes B)$ by setting $s_A(a)(b) = a \otimes b$, for all $a \in A$ and $b \in B$. It is obvious from the definition of R -bilinear maps that (for each $a \in A$) $s_A(a)$ is a group homomorphism; also, that $s_A : A \rightarrow \mathrm{Hom}_{\mathbb{Z}}(B, A \otimes B)$ is a right R -homomorphism.

On the other hand, the map $\mu : \text{Hom}_{\mathbb{Z}}(B, G) \times B \rightarrow G$, defined by $\mu(f, b) = f(b)$, is (by definition of the right scalar multiplication) R -bilinear. Consequently, there is a unique group homomorphism

$$t_G : \text{Hom}_{\mathbb{Z}}(B, G) \otimes B \rightarrow G,$$

so that $t_G(f \otimes b) = f(b)$, for each $f \in \text{Hom}_{\mathbb{Z}}(B, G)$ and $b \in B$.

And now, with feeling \dots

Theorem 5.4.9. $((\) \otimes B, \text{Hom}_{\mathbb{Z}}(B, \), s, t)$ is an adjunction between the categories \mathbf{Mod}_R and \mathbf{Abel} , for each left R -module B .

Proof. We show that the conditions of Theorem 5.2.5 are satisfied. This will show that s and t are natural transformations.

Suppose that $g : A \otimes B \rightarrow G$ is a group homomorphism. We are supposed to come up with a right R -homomorphism $\tilde{g} : A \rightarrow \text{Hom}_{\mathbb{Z}}(B, G)$, so that $t_G \cdot (\tilde{g} \otimes B) = g$. Define $\tilde{g}(a)(b) = g(a \otimes b)$. The reader can easily verify, using the bilinearity of tensors, that

- (a) $\tilde{g}(a)$ is a group homomorphism, for each $a \in A$;
- (b) \tilde{g} is a right R -homomorphism, from A to $\text{Hom}_{\mathbb{Z}}(B, G)$. (Let us check it for scalar multiplication:

$$\tilde{g}(ar)(b) = g(ar \otimes b) = g(a \otimes rb) = \tilde{g}(a)(rb) = (\tilde{g}(a) \cdot r)(b),$$

whence $\tilde{g}(ar) = \tilde{g}(a) \cdot r$.)

- (c) \tilde{g} is the correct morphism, and it is unique relative to the condition $t_G \cdot (h \otimes B) = g$.

■

Hot Air 5.4.10. *Tensor Product Duality as an Exponential Law.* In 5.4.7, $\text{Hom}_{\mathbb{Z}}(B, G)$ was given a group structure, whenever B is a left R -module and G is an abelian group. Now, suppose that G too is a left R -module; we examine $\text{Hom}_{R\mathbf{Mod}}(B, G)$ as a subset of $\text{Hom}_{\mathbb{Z}}(B, G)$. It is easy to see that the sum of two left R -homomorphisms preserves the left scalars, and from that also easy to realize that $\text{Hom}_{R\mathbf{Mod}}(B, G)$ is a subgroup of $\text{Hom}_{\mathbb{Z}}(B, G)$. (Note: the right R -module structure of $\text{Hom}_{\mathbb{Z}}(B, G)$ rarely carries over.)

If B and G are right R -modules then $\text{Hom}_{\mathbf{Mod}_R}(B, G)$ is an abelian group as well, and in a similar way.

Given a right R -module A , a left R -module B , and an abelian group G , we now compare

$$\text{Hom}_{\mathbf{Mod}_R}(A, \text{Hom}_{\mathbb{Z}}(B, G)) \quad \text{and} \quad \text{Hom}_{\mathbb{Z}}(A \otimes B, G).$$

On account of Theorem 5.4.9 and Exercise 5.2.8, there is a natural bijection between these. But more is true; we state this formally in the next exercise.

First, we point out that the isomorphism being described in Exercise 5.4.11 is often rendered as an exponential law. The comment which follows is only intended to help with intuition and memory.

Imagine the notation $\text{Hom}_{\mathbf{C}}(A, B)$ (for a certain category \mathbf{C}) written exponentially ${}^A B$, and that the category in question is understood. Then the isomorphism in Exercise 5.4.11 can be read as

$$({}^{A \otimes B})G \cong A({}^B G),$$

which is an exponential law. Take it with a grain of salt!

Exercise 5.4.11. For each left R -module B , there is a natural group isomorphism

$$\alpha_{A,G} : \text{Hom}_{\mathbb{Z}}(A \otimes B, G) \longrightarrow \text{Hom}_{\mathbf{Mod}_R}(A, \text{Hom}_{\mathbb{Z}}(B, G)).$$

(Hint: have another look at Exercise 5.2.8, and interpret the naturality. The point is to show that the map in 5.2.8 is a group homomorphism.)

Frequently, modules have scalar structures over two or more rings, and on both sides. If A is both a left R -module and a right S -module, then one assumes that the two scalar multiplications are linked by the identity $(ra)s = r(as)$, for all $r \in R$, $s \in S$ and $a \in A$. (In the literature it is said that A is a *bimodule* in this case. We shall not have much occasion to use this terminology.)

The following exercises give the particulars in such situations for Hom and the tensor product. (See [Hu74], Theorem 4.8, p. 203. and Theorem 5.5, p. 210.)

Exercise 5.4.12. Suppose that R and S are rings with identity, and indicate that A is a left (resp. right) R -module, by writing ${}_R A$ (resp. A_R). If ${}_R A$, ${}_R B_S$, ${}_R C_S$ and ${}_R D$ are modules as indicated, then

- (a) $\text{Hom}_{\mathbf{Mod}_R}(A, B)$ is a right S -module, with the scalar multiplication given by $(f \cdot s)(a) = f(a)s$, for all $f \in \text{Hom}_{\mathbf{Mod}_R}(A, B)$, $a \in A$ and $s \in S$. If E_S is a module, then $\text{Hom}_{\mathbf{Mod}_S}(C, E)$ is a right R -module, with the definition of scalar product $(f \cdot r)(c) = f(rc)$, for all $r \in R$, $c \in C$ and $f \in \text{Hom}_{\mathbf{Mod}_S}(C, E)$.
- (b) $\text{Hom}_{\mathbf{Mod}_R}(C, D)$ is a left S -module, with scalar multiplication given by $(s \cdot f)(c) = f(cs)$, for all $f \in \text{Hom}_{\mathbf{Mod}_R}(C, D)$, $c \in C$ and $s \in S$.
- (c) If $g : A \longrightarrow A'$ is a left R -homomorphism, then the induced $\text{Hom}_{\mathbf{Mod}_R}(g, B)$ is a right S -homomorphism. Likewise, if $h : D \longrightarrow D'$ is a left R -homomorphism, then $\text{Hom}_{\mathbf{Mod}_R}(C, h)$ is a left S -homomorphism.
- (d) $\text{Hom}_{\mathbf{Mod}_R}(C, _)$ is a covariant functor from \mathbf{Mod}_R to \mathbf{Mod}_S .
 $\text{Hom}_{\mathbf{Mod}_R}(_, B)$ is a contravariant functor from \mathbf{Mod}_R to \mathbf{Mod}_S .

Exercise 5.4.13. As in 5.4.12, R , S and T are rings with identity. Let ${}_S A_R$ and ${}_R B_T$ be modules as indicated. Then

- (a) $A \otimes_R B$ is a left S - and right T -module, by defining

$$s(a \otimes b)t = (sa) \otimes (bt),$$

for all $a \in A$, $b \in B$, $s \in S$ and $t \in T$. (Note: \otimes_R simply emphasizes that the tensor product is R -bilinear.)

- (b) If $f : A \longrightarrow A'$ is a (left S -) and (right R -)homomorphism, then $f \otimes_R B \in \text{mor}(\mathbf{Mod}_S)$. Likewise, if $g : B \longrightarrow B'$ is a (left R -) and (right T -)homomorphism, then $A \otimes_R g \in \text{mor}(\mathbf{Mod}_T)$.
- (c) (With $S = \mathbb{Z}$) $(_) \otimes_R B$ is a covariant functor from \mathbf{Mod}_R to \mathbf{Mod}_T , and (with $T = \mathbb{Z}$) $A \otimes_R (_)$ is a functor from \mathbf{Mod}_R to \mathbf{Mod}_S .

Exercise 5.4.14. Suppose that R and S are rings with identity. Then for each module ${}_R B_S$, the quadruple

$$((_) \otimes_R B, \text{Hom}_{\mathbf{Mod}_S}(B, _), s, t)$$

forms an adjunction, with functors $(_) \otimes_R B : \mathbf{Mod}_R \longrightarrow \mathbf{Mod}_S$ and $\text{Hom}_{\mathbf{Mod}_S}(B, _)$ in the reverse direction, and natural transformations s and t as in Theorem 5.4.9.

Remark 5.4.15. Finally, in this section, let us see what happens if the ring R is commutative. First, we identify the categories \mathbf{RMod} and \mathbf{Mod}_R . Next, for any pair of R -modules A and B , $A \otimes B$ has a natural R -module structure (defined below) making $(\) \otimes B$ (and $A \otimes (\)$ too) a functor from \mathbf{RMod} to itself.

On the other hand, if B and C are R -modules, then $\text{Hom}_R(B, C)$ is an R -module (under the scalar multiplication in 5.4.7). Theorem 5.4.9 (with suitable preliminaries) then reads as follows:

- (a) For each $r \in R$, the map $\phi_r(a, b) = ra \otimes b$ (for all $a \in A$ and $b \in B$) is R -bilinear.
- (b) There is a unique scalar multiplication on $A \otimes B$, for which

$$r(a \otimes b) = (ra) \otimes b = a \otimes (rb),$$

for all $r \in R$, $a \in A$ and $b \in B$, making $A \otimes B$ an R -module.

- (c) For each R -module B , the quadruple

$$((\) \otimes B, \text{Hom}_R(B, \), s, t)$$

is an adjunction on \mathbf{RMod} (from the category to itself), where s and t are the natural transformations defined for 5.4.9 (which now are R -homomorphisms).

- (d) There is a natural R -isomorphism $\alpha_{A,C}$ from $\text{Hom}_R(A \otimes B, C)$ onto $\text{Hom}_R(A, \text{Hom}_R(B, C))$.

6. Rings and Homology: Exact Sequences

This chapter introduces the elements of homological algebra: exact sequences; projective and injective modules. The first section still involves basic properties of the Hom and tensor functors; almost all the results follow immediately from the theory in the previous chapter.

6.1 Properties of Hom and Tensor Products. The isomorphisms which follow are all natural; we leave the precise formulation of naturality to the reader. In Chapter 8 (Proposition 8.2.2) we shall explore the naturality, in the case of Theorem 6.1.1(d).

Theorem 6.1.1. *Suppose that R is a ring with identity. Then*

(a)

$$\mathrm{Hom}_{\mathbb{Z}}(B, \prod_{i \in I} G_i) \cong \prod_{i \in I} \mathrm{Hom}_{\mathbb{Z}}(B, G_i),$$

for each left R -module B and abelian groups G_i ($i \in I$). This is a natural isomorphism of right R -modules.

A similar property holds for right R -modules.

The next four isomorphisms are of abelian groups and are natural.

(b)

$$\mathrm{Hom}_{\mathbf{RMod}}(B, \prod_{i \in I} C_i) \cong \prod_{i \in I} \mathrm{Hom}_{\mathbf{RMod}}(B, C_i),$$

where B and the C_i are left R -modules. The dual for right modules also holds.

(c)

$$\mathrm{Hom}_{\mathbf{RMod}}(\bigoplus_{i \in I} B_i, C) \cong \prod_{i \in I} \mathrm{Hom}_{\mathbf{RMod}}(B_i, C),$$

where the B_i and C are left R -modules. The dual for right modules also holds.

(d)

$$\left(\bigoplus_{i \in I} A_i\right) \otimes B \cong \bigoplus_{i \in I} (A_i \otimes B),$$

for all right R -modules A_i and each left R -module B . The dual, with direct sums in the second variable, also holds.

(e)

$$\left(\varinjlim_{i \in I} A_i\right) \otimes B \cong \varinjlim_{i \in I} (A_i \otimes B),$$

where each A_i is a right R -module, and B is any left R -module.

If R is commutative, then the isomorphisms in (b), (c), (d) and (e) are natural R -module isomorphisms.

Proof. (a), (d) and (e) are direct consequences of the adjunctions of Hom and tensors, Theorems 5.4.9 and 5.3.1. As to (b), it's a question of showing that the isomorphism of (a) restricts to that isomorphism. We leave the details as an exercise, as we do (c) (See the exercise which follows). ■

Exercise 6.1.2. Give a proof of 6.1.1(c). The quirk is in the contravariance. Although one can formulate an adjoint situation, involving a kind of contravariant tensor product, the exercise is more trouble than it is worth. So it is simpler, in this instance to prove directly that this Hom functor converts coproducts to products.

Refer to Exercises 5.4.13 and 5.4.14 for the following proposition. Keep in mind that the ring R is a module over itself, on both sides. In interpreting (a) of Proposition 6.1.3, which is a left R -module isomorphism, $\text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, B)$ is a left module, according to the scalar product of Exercise 5.4.13(b), taking $R = S$, $C = R$, and $D = B$ there.

Proposition 6.1.3. *Suppose that R is a ring with identity. Then for each left module B , one has a natural isomorphism*

$$(a) \quad \text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, B) \cong B.$$

$$(b) \quad (R_R) \otimes_R B \cong B.$$

Proof. To establish (a), note that any left R -homomorphism of R into B is completely determined by the image of the identity. So define, for each $f \in \text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, B)$, $\theta_B(f) = f(1)$. It is a routine matter to show that θ is a left R -isomorphism.

As to the naturality, it is required to check that if $h : B \rightarrow B'$ is a left R -homomorphism, then

$$h \cdot \theta_B = \theta_{B'} \cdot \text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, h).$$

As to that, if $f \in \text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, B)$, then the left side of the above equation is $h(f(1))$, whereas the right is

$$\theta_{B'} \cdot \text{Hom}_{\mathbf{R}\text{Mod}}({}_R R, h)(f) = \theta_{B'}(h \cdot f),$$

which is also $h(f(1))$.

As for (b), the map $\rho_B : R \otimes B \rightarrow B$ given by $\rho_B(r, b) = rb$, is easily shown to be R -bilinear, so that there is a homomorphism $\hat{\rho}_B : R \otimes_R B \rightarrow B$ for which $\hat{\rho}_B(r \otimes b) = rb$. One needs to verify that it preserves left scalars, and for this it suffices to show that

$$r_1 \hat{\rho}_B(r_2 \otimes b) = \hat{\rho}_B(r_1 r_2 \otimes b),$$

for all $r_1, r_2 \in R$ and $b \in B$. (Why?) But the latter identity should be obvious.

On the other hand, the map $\alpha : B \rightarrow R \otimes_R B$, defined by $\alpha(b) = 1 \otimes b$ is an R -homomorphism. Finally note that $\hat{\rho}_B(\alpha(b)) = b$, for all $b \in B$, and that

$$\alpha(\hat{\rho}_B(r \otimes b)) = \alpha(rb) = r\alpha(b) = r(1 \otimes b) = r \otimes b,$$

whence $\alpha \cdot \hat{\rho}_B = 1$ (on $R \otimes_R B$).

One also needs to verify that $\hat{\rho}$ is natural, but this we leave to the reader. ■

Next, we list a few exercises, adapted from [Hu74].

Exercise 6.1.4. ([Hu74], p. 206, 1.) In **Abel**, show the following:

- (a) For any group A and every positive integer m one has the natural isomorphism

$$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_m, A) \cong A[m] \equiv \{a \in A : ma = 0\}.$$

- (b) $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}_m, \mathbb{Z}_n) \cong \mathbb{Z}_{(m,n)}$.

Exercise 6.1.5. ([Hu74], p. 216, 2.) In **Abel**, prove the following:

- (a) For any group A and positive integer m there is a natural isomorphism

$$A \otimes_{\mathbb{Z}} \mathbb{Z}_m \cong A/mA.$$

So describe $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n$.

- (b) Given A and B , two finitely generated groups, describe $A \otimes_{\mathbb{Z}} B$, using the Fundamental Theorem on Finitely Generated Abelian Groups.

Exercise 6.1.6. ([Hu74], p. 216, 3.) Again in **Abel**, prove the following:

- (a) If A is a torsion abelian group, then $A \otimes_{\mathbb{Z}} \mathbb{Q} = 0$,
- (b) Recall that B is divisible if for each $b \in B$, and each integer $n \neq 0$, $nx = b$ is solvable. Prove that if A is a torsion group and B is divisible $A \otimes_{\mathbb{Z}} B = 0$.
- (c) If A is torsion free and divisible, then

$$\mathbb{Q} \otimes_{\mathbb{Z}} A \cong A,$$

naturally.

We also note the following items, leaving the proofs as exercises. The first is an associative law for tensor products.

Exercise 6.1.7. Suppose that R and S are rings with identity, and that A_R , ${}_R B_S$ and ${}_S C$ are modules, as indicated. Then

$$(A \otimes_R B) \otimes_S C \cong A \otimes_R (B \otimes_S C),$$

as abelian groups. If $R = S$ is commutative, then the above is an R -isomorphism.

Exercise 6.1.8. If R is commutative, then $A \otimes_R B \cong B \otimes_R A$, as R -modules.

6.2 Exact Sequences. The notion of an exact sequence, and more particularly, a short exact sequence, is fundamental in homological algebra. In this section “module” means “left module”. R is a ring with identity. “Hom” (without subscript) here means “ $\mathrm{Hom}_{\mathbf{R}\text{Mod}}$ ”, and the tensor symbol will also occur without subscript, in this section.

Definition & Remarks 6.2.1. A sequence

$$\dots \xrightarrow{f_{i-1}} A_i \xrightarrow{f_i} A_{i+1} \xrightarrow{f_{i+1}} \dots$$

of R -homomorphisms and R -modules is said to be *exact at* A_{i+1} if $f_i(A_i) = \ker(f_{i+1})$. Observe that $f_{i+1} \cdot f_i = 0$ if and only if $f_i(A_i) \leq \ker(f_{i+1})$. If the sequence is exact at each A_i then we say that it is an *exact sequence*. (Note: in the sequel 0 stands for the homomorphism which sends every element to 0.)

If the sequence

$$(E) \quad 0 \xrightarrow{0} A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{0} 0$$

is exact, we say that it is a *short exact sequence*. (In the future we omit labelling the zero maps into or out of the 0 module, as there can be no other such homomorphism.) Thus, (E) is short exact if and only if (i) $\ker(f) = 0$ (i.e., f is one-to-one), (ii) $f(A) = \ker(g)$, and $g(B) = \ker(0) = C$; i.e. g is onto.

The prototype of a short exact sequence arises as follows: let B be an R -module and A a submodule; then the sequence

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\mu} C \longrightarrow 0$$

where α is the inclusion and μ is the canonical homomorphism, is exact. On the other hand, if (E) above is exact, then $C \cong B/f(A)$.

Definition 6.2.2. The short exact sequence (E) above is said to be *split exact* (or it is said that (E) *splits*) if there exists an R -homomorphism $h : C \longrightarrow B$ such that $g \cdot h = 1_C$. This definition, although apparently asymmetric, is not so, as the next proposition demonstrates.

Proposition 6.2.3. *Suppose that*

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is short exact. Then the following are equivalent:

- (a) (E) is split exact.
- (b) There is an R -homomorphism $k : B \longrightarrow A$ so that $k \cdot f = 1_A$.
- (c) There is a submodule B' of B , isomorphic to C , such that $B = B' \oplus \ker(g)$.

Proof. We show that (a) and (c) are equivalent; the equivalence of (b) and (c) is shown in a similar way.

(a) \Rightarrow (c): We suppose that $h : C \longrightarrow B$ is an R -homomorphism, so that $g \cdot h = 1_C$. Let $B' = h(C)$. We must show that $B' \cap \ker(g) = 0$, and $B = B' + \ker(g)$. First, if $b \in B' \cap \ker(g)$, then $b = h(c)$, for some $c \in C$, and $g(b) = 0$. But then $c = g(h(c)) = 0$, which implies that $b = 0$. Next, if $x \in B$, consider $y = h(g(x))$; evidently, $y \in B'$, and

$$g(x - y) = g(x) - g(h(g(x))) = g(x) - g(x) = 0,$$

so that $x - y \in \ker(g)$, whence $x \in B' + \ker(g)$. Since h is one-to-one, it is clear that B' and C are isomorphic.

(c) \Rightarrow (a): If $B = B' + \ker(g)$ as stated in (c), then, for each $c \in C$, there is a $z \in B$, so that $g(z) = c$. Write $z = z_1 + z_2$, with $z_1 \in B'$, and $z_2 \in \ker(g)$, and define $h(c) \equiv z_1$. This choice is independent of z : if $g(w) = c$, and $w = w_1 + w_2$, with $w_1 \in B'$ and $w_2 \in \ker(g)$, then

$$0 = g(z - w) = g(z_1 - w_1) + g(z_2 - w_2) = g(z_1 - w_1),$$

and so $z_1 - w_1 \in \ker(g) \cap B' = 0$; that is, $z_1 = w_1$.

One easily shows that h , so defined, is an R -homomorphism. This we leave to the reader. Finally,

$$g(h(c)) = g(z_1) = g(z_1 + z_2) = g(z) = c,$$

that is, $g \cdot h = 1_C$. ■

Note: the phrase “isomorphic to C ” is added for emphasis. By the First Isomorphism Theorem $B/f(A) \cong C$ and also to B' , so that the isomorphy of these two is forced. Notice that the isomorphy is not explicitly referred to in the proof that (c) implies (a).

Hot Air 6.2.4. The study of the effect of the functors Hom and \otimes on exact sequences is a major preoccupation in homological algebra, which can be characterized as the study of the structure of rings with identity through their modules and the exact sequences between them.

We shall restrict our attention to the effect of these functors on short exact sequences. In this section the basic results are presented, with various developments to follow in later sections and chapters.

Proposition 6.2.5. (Hom on Short Exact Sequences.) *Suppose that the sequence*

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is short exact. Then, for each R -module M ,

(a)

$$0 \longrightarrow \text{Hom}(M, A) \xrightarrow{\text{Hom}(M, f)} \text{Hom}(M, B) \xrightarrow{\text{Hom}(M, g)} \text{Hom}(M, C)$$

is exact (at $\text{Hom}(M, A)$ and $\text{Hom}(M, B)$).

(b)

$$0 \longrightarrow \text{Hom}(C, M) \xrightarrow{\text{Hom}(g, M)} \text{Hom}(B, M) \xrightarrow{\text{Hom}(f, M)} \text{Hom}(A, M)$$

is exact.

Exactness at $\text{Hom}(M, C)$ and $\text{Hom}(A, M)$ fails, in general.

Proof. We prove (a), leaving (b) as an exercise.

BeginProof that $\text{Hom}(M, f)$ is one-to-one: if $\text{Hom}(M, f)(h) = 0$, then (by definition) $f \cdot h = 0$, which is to say that $h(M) \leq \ker(f) = 0$, and so $h = 0$.

BeginProof of exactness at $\text{Hom}(M, B)$: since $g \cdot f = 0$, and $\text{Hom}(M, \)$ is functorial, it follows that

$$\text{Hom}(M, g) \cdot \text{Hom}(M, f) = 0.$$

It therefore suffices to show that $\ker(\text{Hom}(M, g)) \leq \text{Hom}(M, f)(\text{Hom}(M, A))$. So suppose that $k \in \text{Hom}(M, B)$ and $g \cdot k = 0$. This means that $k(M) \leq \ker(g) = f(A)$. Then define $l \in \text{Hom}(M, A)$ as follows: $l(x) = a$, where $f(a) = h(x)$. This is unambiguous, because f is one-to-one. It is routine to show that l is an R -homomorphism, and clearly $\text{Hom}(M, f)(l) = k$.

Examples of the inexactness at the ends of the sequences in (a) and (b) are highlighted next. ■

Exercise 6.2.6. In this exercise $R = \mathbb{Z}$, the ring of integers.

- (a) In (a) of Proposition 6.2.5, let $M = C = \mathbb{Z}_2$, $B = \mathbb{Z}_4$ and $A = 2\mathbb{Z}_4$. f is the inclusion map, and g the canonical homomorphism $\mathbb{Z}_4 \rightarrow \mathbb{Z}_2$. Show that the sequence is not exact at $\text{Hom}(M, C)$.
- (b) Find an example to show that in (b) of Proposition 6.2.5 the sequence need not exact at $\text{Hom}(A, M)$. (Hint: let $M = \mathbb{Z}$ itself.)

Now the effect of tensors on a short exact sequence.

Proposition 6.2.7. (Tensor Products on Short Exact Sequences.) *Suppose that*

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is a short exact sequence of left R -modules. Then for each right R -module M , the following sequence is exact:

$$M \otimes A \xrightarrow{M \otimes f} M \otimes B \xrightarrow{M \otimes g} M \otimes C \longrightarrow 0.$$

Exactness at $M \otimes A$ fails, in general. Also a dual result holds for short exact sequences of right R -modules.

Proof. The proof, for the most part, goes to the heart of the definition of tensor products. The only trivial part is the identity $(M \otimes g) \cdot (M \otimes f) = 0$, because $M \otimes ()$ is a functor, and, clearly, $M \otimes 0 = 0$.

Exactness at $M \otimes C$: each element of $M \otimes C$ is a sum of tensors of the form $m \otimes c$, each of which can be written as $m \otimes c = m \otimes g(b)$, for a suitable $b \in B$, as g is surjective.

Exactness at $M \otimes B$: we already know that $(M \otimes f)(M \otimes A) \leq \ker(M \otimes g)$. Denote

$$M \otimes B / (M \otimes f)(M \otimes A) \cong G$$

and let $\mu : M \otimes B \rightarrow G$ be the canonical homomorphism. By the Induced Homomorphism Theorem, there is a homomorphism $g^* : G \rightarrow M \otimes C$, such that $g^* \cdot \mu = M \otimes g$. On the other hand, define a map $\theta : M \times C \rightarrow G$ by $\theta(m, c) = \mu(m \otimes b)$, where b is chosen so that $g(b) = c$. We let the reader check that this is well defined: $\mu(m \otimes b)$ does not depend on the choice of b . It is also easy to verify that θ is R -bilinear. This means that there is a homomorphism $\theta^* : M \otimes C \rightarrow G$ with the property that

$$\theta^*(m \otimes c) = \mu(m \otimes b),$$

whenever $g(b) = c$. From this we get that

$$\theta^*(g^*(\mu(m \otimes b))) = \theta^*(m \otimes g(b)) = m \otimes b,$$

from which we conclude (using the uniqueness provision of the Induced Homomorphism Theorem) that $\theta^* \cdot g^* = 1_G$. This makes g^* one-to-one. On the other hand, since $M \otimes g$ is onto, so is g^* , and therefore g^* is an isomorphism. This outcome means that $\ker(M \otimes g) = (M \otimes f)(M \otimes A)$. ■

Example 6.2.8. For $R = \mathbb{Z}$, consider the short exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\alpha} \mathbb{Q} \xrightarrow{\mu} \mathbb{Q}/\mathbb{Z} \longrightarrow 0,$$

where α is the inclusion map, and μ is the canonical homomorphism. Tensor with \mathbb{Z}_2 : by Exercise 6.1.6, $\mathbb{Z}_2 \otimes \mathbb{Q} = 0$, whereas $\mathbb{Z}_2 \otimes \mathbb{Z} = \mathbb{Z}_2$, according to Theorem 6.1.1(b). Thus, $\mathbb{Z}_2 \otimes \alpha$ is the zero map, and not one-to-one. Exactness at $M \otimes A$ then, in Proposition 6.2.7, can indeed fail.

The next exercise is Exercise 9, p. 217, [Hu74].

Exercise 6.2.9. (a) Suppose that B is a module, and I is a right ideal of R . Then

$$(R/I) \otimes B \cong B/IB \quad (\text{as groups}),$$

where IB is the subgroup of B generated by all elements of the form xb , with $x \in I$ and $b \in B$. Moreover, the functors $(R/I) \otimes ()$ and \mathcal{M}_I defined by

(i) $\mathcal{M}_I(B) \equiv B/IB$, on objects, and

(ii) $\mathcal{M}_I(g)(IB + x) = IC + g(x)$, for each $g \in \text{Hom}(B, C)$ and each $x \in B$,

are naturally equivalent.

(b) If R is commutative, and I and J are ideals of R , then

$$(R/I) \otimes (R/J) \cong R/(I + J) \quad (\text{as } R\text{-modules}).$$

We conclude the section with the following observation, which ought to be clear from Theorem 6.1.1, (b), (c) and (d), together with Proposition 6.2.7. In any event, we leave the proof to the reader.

Proposition 6.2.10. *If the sequence*

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is split exact, then so are

(a)

$$0 \longrightarrow \text{Hom}(M, A) \xrightarrow{\text{Hom}(M, f)} \text{Hom}(M, B) \xrightarrow{\text{Hom}(M, g)} \text{Hom}(M, C) \longrightarrow 0,$$

(b)

$$0 \longrightarrow \text{Hom}(C, M) \xrightarrow{\text{Hom}(g, M)} \text{Hom}(B, M) \xrightarrow{\text{Hom}(f, M)} \text{Hom}(A, M) \longrightarrow 0,$$

and

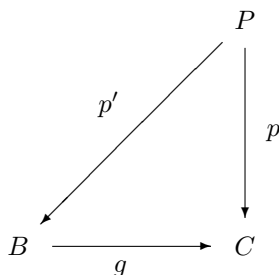
(c)

$$0 \longrightarrow K \otimes A \xrightarrow{K \otimes f} K \otimes B \xrightarrow{K \otimes g} K \otimes C \longrightarrow 0,$$

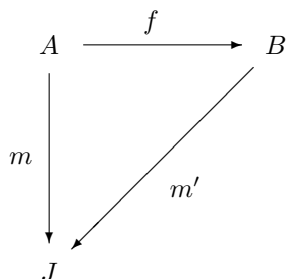
for all left R -modules M and right R -modules K .

6.3 Projective and Injective Modules. Here we introduce the modules which correct the defects in exactness of Proposition 6.2.5. (In Chapter 8 we take up the modules which do the same, with respect to tensor products.) Once again, we restrict ourselves to left R -modules (unless there is an indication to the contrary). In particular, the definitions of projective and injective modules should be seen as categorical duals, in the category \mathbf{RMod} . R still stands for a fixed ring with identity.

Definition 6.3.1. An R -module P is said to be projective if for each surjective R -homomorphism $g : B \longrightarrow C$, and each R -homomorphism $p : P \longrightarrow C$, there is an R -homomorphism $p' : P \longrightarrow B$ (not necessarily unique) such that the diagram below commutes:



The concept of an *injective* module is dual, that is, with all arrows turned around: J is injective if, for each one-to-one R -homomorphism $f : A \rightarrow B$, and each R -homomorphism $m : A \rightarrow J$, there is an R -homomorphism $m' : B \rightarrow J$, so that the dual diagram commutes:



The following is straightforward, and it is left to the reader:

Proposition 6.3.2. (a) P is a projective module if and only if for every short exact sequence

$$(E) \quad 0 \rightarrow A \xrightarrow{f} B \xrightarrow{g} C \rightarrow 0,$$

the induced sequence

$$0 \rightarrow \text{Hom}(P, A) \xrightarrow{\text{Hom}(P, f)} \text{Hom}(P, B) \xrightarrow{\text{Hom}(P, g)} \text{Hom}(P, C) \rightarrow 0$$

is short exact.

(b) J is an injective module if and only if for each short exact

$$(E) \quad 0 \rightarrow A \xrightarrow{f} B \xrightarrow{g} C \rightarrow 0,$$

the induced sequence

$$0 \rightarrow \text{Hom}(C, J) \xrightarrow{\text{Hom}(g, J)} \text{Hom}(B, J) \xrightarrow{\text{Hom}(f, J)} \text{Hom}(A, J) \rightarrow 0$$

is short exact.

Next, we turn to properties of projective modules. A full treatment of injectives will have to wait a while, until we are ready to prove that every R -module can be embedded in an injective one.

Theorem 6.3.3. (a) Every free R -module is projective (but the converse is, in general, false).

(b) Suppose that $P = \bigoplus_{i \in I} P_i$. Then P is projective if and only if each P_i is projective.

(c) An R -module P is projective if and only if there is a free module F and a submodule M of F , so that $F = P \oplus M$.

Proof. (a) Suppose that F is a free module, on the set of generators X . Suppose that $g : B \rightarrow C$ is a surjective homomorphism, and $p : F \rightarrow C$ is a homomorphism. For each $x \in X$, there is an element $b_x \in B$, such that $g(b_x) = p(x)$. Define $\tilde{p} : X \rightarrow B$ by $\tilde{p}(x) = b_x$. Notice that $g \cdot \tilde{p} = p$, by definition. Since F is free on X , there is a unique homomorphism $p' : F \rightarrow B$ so that $p'(x) = \tilde{p}(x)$, for each $x \in X$. Since $g \cdot p' = p$ restricted to X , the two sides agree on F , by uniqueness of extensions, or else because X generates F .

(b) The proof that P is projective if each P_i is projective, is an illustration of the fact that the direct sum is the coproduct in \mathbf{RMod} . Details left to the reader.

As to the converse, suppose that P is projective. Let π_i and δ_i denote (respectively) the i -th projection on P_i , and the i -th coprojection of $P_i \rightarrow P$. Suppose that $g : B \rightarrow C$ is surjective, and $p : P_i \rightarrow C$ is an R -homomorphism. Applying projectivity to $p \cdot \pi_i$, we get a homomorphism $q : P \rightarrow B$ for which $g \cdot q = p \cdot \pi_i$. Composing on the right:

$$g \cdot q \cdot \delta_i = p \cdot \pi_i \cdot \delta_i = p \cdot 1_{P_i} = p.$$

Thus, $q \cdot \delta_i : P_i \rightarrow B$ fulfills the requirement.

(c) Sufficiency is a direct consequence of (a) and (b). So suppose now that P is a projective module. Let F be the free module over the underlying set of P . This means that there is a homomorphism $h : F \rightarrow P$, for which $h(x) = x$, for each $x \in P$. In particular, h is onto P . Apply projectivity now to the identity 1_P to obtain a homomorphism $m : P \rightarrow F$ such that $h \cdot m = 1_P$. To conclude, notice that this says that the exact sequence

$$0 \longrightarrow \ker(h) \xrightarrow{i} F \xrightarrow{h} P \longrightarrow 0,$$

in which i is the inclusion, splits. (Proposition 6.2.3) Thus, F is a direct sum of $\ker(h)$ and another submodule, isomorphic to P . ■

Presently we will prove that if R is a PID (= principal ideal domain), then every submodule of a free module is free (and, hence, projective). First, however, an example of a projective module which is not free. It is the same example provided by [Hu74], p. 193.

Note that if A is an R -module and B is a submodule, so that $A = B \oplus C$, for some submodule C , then B is said to be a *summand* of A .

Example 6.3.4. Let $R = \mathbb{Z}_6$. As is well known, $\mathbb{Z}_6 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_3$, as abelian groups. However, \mathbb{Z}_2 and \mathbb{Z}_3 are, in fact, \mathbb{Z}_6 -modules, and this isomorphism is one as \mathbb{Z}_6 -modules. So, by Theorem 6.3.3(c), both summands are projective, but they are certainly not free \mathbb{Z}_6 -modules.

Exercise 6.3.5. Show that \mathbb{Q} is not a projective abelian group.

Here is a useful characterization of projectivity.

Exercise 6.3.6. Over any ring with identity R , prove that P is a projective R -module if and only if every short exact sequence

$$0 \longrightarrow A \longrightarrow B \longrightarrow P \longrightarrow 0$$

splits.

Theorem 6.3.7. *If R is a PID, then every submodule of a free R -module is free (and therefore projective).*

Proof. This argument requires the principle of transfinite induction, as well as the set-theoretic axiom that every set can be well-ordered.

Suppose that F is a free R -module. Then $F = \bigoplus_{i \in I} R_i$, where each R_i is an isomorphic copy of ${}_R R$. Now, we assume that I has been well-ordered. For each $j \in I$, let

$$F_j = \bigoplus_{i < j} R_i;$$

it should be clear that each $F_i \leq F_{i+1}$, and that if j has no predecessor in the ordering, then $F_j = \bigcup_{i < j} F_i$. Finally, $F = \bigcup_{i \in I} F_i$. (Note that F_1 is trivial.)

Now suppose that G is a submodule of F , and let $G_i = F_i \cap G$. Again observe that (i) $G_i \leq G_{i+1}$ and (ii) $G_j = \bigcup_{i < j} G_i$, if j has no predecessor. Also, (iii) $G = \bigcup_{i \in I} G_i$. By one of the isomorphism theorems, for each $i \in I$,

$$G_{i+1}/G_i \cong G_{i+1}/(F_i \cap G_{i+1}) = (F_i + G_{i+1})/F_i,$$

and the latter is a submodule of F_{i+1}/F_i , which is isomorphic to R , because R is a PID. Thus, G_{i+1}/G_i is isomorphic to an ideal J_i of R , which is either trivial, or else principal, and free in either event.

Since free modules are projective, it follows that either $G_i = G_{i+1}$, or else the canonical short exact sequence

$$0 \longrightarrow G_i \longrightarrow G_{i+1} \longrightarrow G_{i+1}/G_i \longrightarrow 0$$

splits, by Exercise 6.3.6. This means that $G_{i+1} = G_i \oplus S_i$, where S_i is either trivial or an isomorphic copy of R . Suppose now that, for each $i < j$, it has been demonstrated that $G_i = \bigoplus_{i' < i} S_{i'}$. (Note that $S_1 = G_2$.) If j is the successor of j' , then $G_j = G_{j'+1} = G_{j'} \oplus S_{j'}$, and it follows that $G_j = \bigoplus_{i < j} S_i$. If j has no predecessor, then it is the union of the G_i (all $i < j$), each of which is a direct sum of S_i 's, which, in turn, implies that $G_j = \bigoplus_{i < j} S_i$. By the principle of transfinite induction, it follows that

$$G = \bigoplus_{i \in I} S_i,$$

and hence free. ■

Hot Air 6.3.8. *Fastforward.* The central theorem of Chapter 7 is the Wedderburn–Artin Theorem, which gives necessary and sufficient conditions for every left R -module to be projective. In the next proposition we shall give (the easier) part of this characterization; it sounds nice, but does not give any significant information about the ring.

Before getting to this proposition, however, note the following exercise, which tells us when we can expect every module over R to be free. It is exercise 14, p. 199, [Hu74].

Exercise 6.3.9. Prove that if every left R -module is free then R is a division ring. (Recall that a *division ring* is a ring with identity, for which the nonzero elements form a group with respect to multiplication.)

Hot Air 6.3.10. *Before Anybody Asks.* The converse of the preceding exercise is also true; see Theorem 2.4, p. 183 of [Hu74]. In a nutshell, what one has to become convinced of is that all the standard material for bases and dimension of a vector space, over a field, goes through *verbatim* for modules over a division ring. The reader will, of course, take care to become convinced of this!

Proposition 6.3.11. *For a ring with identity R , the following are equivalent.*

- (a) *Every R -module is projective.*
- (b) *Every short exact sequence splits.*
- (c) *Every R -module is injective.*

Proof. We prove that (a) is equivalent to (b), and leave the equivalence of (b) and (c) to the reader.

If every module is projective, and

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0,$$

is short exact, then (since C is projective) the sequence splits, according to Exercise 6.3.6. Conversely, if every short exact sequence splits, let P be an R -module. As in the proof of Theorem 6.3.3(c), there is a free module F and a surjective homomorphism $g : F \longrightarrow P$. The sequence below, with inclusion i ,

$$0 \longrightarrow \ker(g) \xrightarrow{i} F \xrightarrow{g} P \longrightarrow 0,$$

then splits. Applying 6.3.3(c), P is projective. ■

Remark 6.3.12. *Something to Think About.* Although there surely is the “right module” dual of Proposition 6.3.11, it is by no means obvious that Proposition 6.3.11, for left R -modules, implies the same proposition, for right modules. This lovely symmetry is part of the Wedderburn–Artin Theorem.

To conclude this section, here is a partial counterpart, for injectives, to Theorem 6.3.3. Notice that on the face of it, we still have no clue that any injectives exist. After the proposition, it is shown that, at least, over the integers, they do; they are the divisible groups.

The proof of (a) of the next proposition is left to the reader.

Proposition 6.3.13. (a) *Suppose that $J = \prod_{i \in I} J_i$. Then J is injective if and only if each J_i is injective.*

- (b) *If J is an injective R -module, then every short exact sequence*

$$0 \longrightarrow J \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

splits.

(Note: the converse is true, but will have to wait until we can demonstrate that “enough” injectives exist. We also note that (a) and (b) suffice to show that (b) and (c) in Proposition 6.3.11 are equivalent.)

Proof. Of (b): If J is injective, then apply the definition to f and the identity 1_J to produce a homomorphism $h : B \longrightarrow J$, such that $1_j = h \cdot f$. ■

Finally, we characterize the injective abelian groups. It is given as Lemma 3.9, p. 195, [Hu74].

Theorem 6.3.14. *An abelian group is injective if and only if it is divisible.*

Proof. Suppose that G is an injective group, $g \in G$, and n is a nonzero integer. Define $\alpha : n\mathbb{Z} \rightarrow G$ by $\alpha(nk) = kg$; this is a group homomorphism. Since G is injective, α extends to a homomorphism $\beta : \mathbb{Z} \rightarrow G$. Note now that $g = \alpha(n) = n\beta(1)$, whence G is divisible.

Conversely, suppose that D is a divisible group, $f : A \rightarrow D$ is a homomorphism, with A a subgroup of B . Let $\text{Ex}(f)$ denote the set of all pairs (C, h) , where C is a subgroup of B containing A , and $h : C \rightarrow D$ is a homomorphism which extends f . We prepare to show that Zorn's Lemma can be applied to $\text{Ex}(f)$. First, we must partially order this set:

$$(C, h) \leq_{\text{def}} (C', h')$$

if $C \leq C'$ and h' extends h . We let the reader check that this is a partial order. Next, suppose that $\{(C_i, h_i) : i \in I\}$ is a chain and that I is ordered too, so that if $i < j$ then $C_i \leq C_j$. Let $\tilde{C} = \cup_i C_i$, and define $\tilde{h} : \tilde{C} \rightarrow D$ as $\tilde{h}(x) = h_i(x)$, provided that $x \in C_i$. It must be shown that this definition is independent of i , and that \tilde{h} is a homomorphism. This is routine, however. It should also be obvious that $(C_i, h_i) \leq (\tilde{C}, \tilde{h})$.

We have shown that Zorn's Lemma applies to $\text{Ex}(f)$. Now suppose that (B', g') is a maximal member of $\text{Ex}(f)$. It is to be hoped that $B' = B$. Suppose not, however. Suppose that $y \in B \setminus B'$, and consider $C = B' + \langle y \rangle$. Note that $B' < C$; in C each element has the form $b' + ky$, for suitable $b' \in B'$ and integer k . However, the expression might not be unique, and this is precisely the rub in the argument that follows.

Let t be the least nonnegative integer such that $ty \in B'$; if $t = 0$, this means that $B' \cap \langle y \rangle = 0$, and so C is the direct sum of B' and $\langle y \rangle$, and we can choose any $d \in D$, and unambiguously define $h : C \rightarrow D$ as follows:

$$h(b' + ky) = g'(b') + kd.$$

However, we have the pair $(C, h) > (B', g')$, violating the maximality of (B', g') .

Thus, t , in the preceding paragraph, is positive. In D , let d be a solution to the equation $g'(ty) = td$. This time define $h : C \rightarrow D$ by

$$h(b' + ky) = g'(b') + kd.$$

Once it is demonstrated that h is well-defined, it is easy to show that it is a homomorphism, which once more makes the pair (C, h) violate the maximality of (B', g') . So we finish by proving that h is well-defined. Suppose that $b' + ky = b'' + ly$; then $(k - l)y = b'' - b' \in B'$, from which it follows that t divides $k - l$. (WHY? Think hard about this, and its consequence for generalization. See the comment after the proof.) Resuming: suppose that $k - l = mt$; then

$$g'(b'') - g'(b') = g'(b'' - b') = g'((k - l)y) = mg'(ty) = mtd = (k - l)d;$$

this implies that $g'(b') + kd = g'(b'') + ld$, and shows that h is well-defined. ■

Remark 6.3.15. *A Question, Really.* If one defines “divisible module” appropriately over a PID, doesn't Theorem 6.3.14 follow, word for word?

Now on to the Wedderburn–Artin Theorem!

7. Rings and Homology: The Wedderburn–Artin Theorem

The central theme of this chapter is the Wedderburn–Artin Theorem (Theorem 7.4.4), which characterizes, in a variety of ways according to this presentation, the rings which are left Artinian and have no nonzero nilpotent left ideals.

7.1 Simple Modules. Once again, R stands for a ring with identity, and “module” throughout this section means “left module”. $\text{Hom}(A, B)$ stands for $\text{Hom}_{R\text{-Mod}}(A, B)$.

In this section we characterize the rings in question in terms of decompositions into simple submodules.

Definition 7.1.1. An R -module S is *simple* if it has no nontrivial, proper submodules. Note that S is simple if and only if $Ra = S$, for each nonzero $a \in S$. Simple modules are also called *irreducible*.

The first observation about simple modules is straightforward, but crucial. It is a version of Schur’s Lemma; the proof is omitted.

Proposition 7.1.2. *Suppose that S is an R -module. Then the following are equivalent.*

- (a) S is simple.
- (b) Every nonzero member of $\text{Hom}(S, A)$ is one-to-one, for each R -module A .
- (c) Every nonzero member of $\text{Hom}(A, S)$ is onto, for each R -module A .

Corollary 7.1.3. *A nonzero module S is simple if and only if it is isomorphic to R/M , for a suitable maximal left ideal M of R .*

Proof. If $R/M = S$, for some maximal left ideal M , then because there is a one-to-one order preserving correspondence between the R -submodules of S and the left ideals of R which contain M , it follows that S is simple.

Conversely, suppose that S is nonzero and simple; let $a \in S$ be any nonzero element. The R -homomorphism $h(r) = ra$, $h \in \text{Hom}(R, S)$, is nonzero, and, therefore surjective, according to Proposition 7.1.2(c). By the First Isomorphism Theorem, $S = Ra = R/\ker(h)$, and, by the same argument as in the first paragraph, $\ker(h)$ must be maximal. ■

Definition & Remarks 7.1.4. If A is any R -module and $a \in A$, consider the set

$$\text{Ann}(a) = \{r \in R : ra = 0\}.$$

It is a left ideal, called the (*left*) *annihilator* of a . Note, from the proof of Corollary 7.1.3, that Ra is simple if and only if $\text{Ann}(a)$ is a maximal left ideal of R , and then $R/\text{Ann}(a) = Ra$.

If A is an R -module, put

$$\text{Ann}(A) = \cap \{ \text{Ann}(a) : a \in A \}.$$

It is easily checked that $\text{Ann}(A)$ is a twosided ideal. $\text{Ann}(A)$ is the *annihilator* of A . We emphasize, for later use, that, if S is simple, then $\text{Ann}(S)$ is an intersection of maximal left ideals M_i , for which the factor modules R/M_i are isomorphic.

The following remark should put the above in perspective.

Remark 7.1.5. Fix a prime number p . Recall that an abelian group G is called an *elementary p -group* if $pg = 0$, for each $g \in G$. Note that $\text{Ann}(G) = p\mathbb{Z}$, for every elementary p -group (by definition), which is maximal, regardless of whether G is simple or not.

In the next section (Example 7.2.2) we shall find a simple module, with trivial annihilator. The trivial ideal is not a maximal left ideal; more about this in §7.3. Thus, even though a module S is (left) R -simple its annihilator need not be a maximal left ideal.

The next exercise serves as a good lead-in for the theorem which follows, and shows that the situation with elementary p -groups is fairly typical.

Exercise 7.1.6. Suppose that A is a nonzero R -module, with the property that $\text{Ann}(A)$ is a maximal left ideal. Then prove that A is R -isomorphic to $\bigoplus_{i \in I} S_i$, where each S_i is a simple submodule, which is R -isomorphic to $R/\text{Ann}(A)$.

(Hint: first show that for any two non-zero $a, b \in A$, either Ra and Rb are the same, or else their intersection is 0. Now show that Zorn's Lemma can be applied to come up with a family $\{S_i : i \in I\}$ of simple submodules of A , with the property that each S_j intersects the submodules generated by the others trivially, and which is maximal with respect to this feature. Then $A = \bigoplus_{i \in I} S_i$.)

The first three conditions in the next result are the three in Proposition 6.3.11. The new item can be viewed as a generalization of what happened in the preceding exercise: under the conditions of the exercise, a module A is isomorphic to the direct sum of a number of copies of the same simple module (namely $R/\text{Ann}(A)$), while in (d) of Theorem 7.1.7 there could be nonisomorphic simple summands.

A module which satisfies (d) of Theorem 7.1.7 is said to be *semisimple*.

Theorem 7.1.7. For a ring R with identity, the following statements are equivalent.

- (a) Every R -module is projective.
- (b) Every short exact sequence splits.
- (c) Every R -module is injective.
- (d) Each R -module A is a direct sum of simple submodules.

Proof. (a), (b) and (c) \Rightarrow (d): The first thing that must be demonstrated is that any nonzero R -module possesses simple submodules. (Note that (d) trivially follows if $A = 0$; we therefore assume that A is nonzero.)

For each $a \in A$, $a \neq 0$, let N be a submodule of A which is maximal with respect to the property of not containing a . (Such an N exists, by Zorn's Lemma.) Now put

$$N^* = \bigcap \{ B \leq A : N < B \};$$

observe that, since each submodule B which contains N properly also contains a , $a \in N^*$, which is to say that $N \neq N^*$. Verify that N^*/N is simple. However, by (b), the canonical short exact sequence

$$0 \longrightarrow N \longrightarrow N^* \longrightarrow N^*/N \longrightarrow 0$$

splits, and this means that there is a simple submodule S of A , so that $N^* = N \oplus S$; (and $S = N^*/N$, of course!)

We note (again), in passing, that if S and T are any two simple submodules, then either $S \cap T = 0$, or else $S = T$.

Suppose now, as outlined in the hint of the previous exercise, that $\{S_i : i \in I\}$ is a family of nonzero simple submodules of A , so that each S_j intersects the submodule generated by the others trivially, and which is maximal with respect to this property. (Mr. Zorn strikes again!) Let

$$C = \sum_{i \in I} S_i.$$

If $C \neq A$, then A/C is a nonzero R -module, which, according to the argument of the second paragraph of this proof, contains a simple submodule T . There is a unique submodule $T' \geq C$, such that $T'/C = T$; again the canonical short exact sequence

$$0 \longrightarrow C \longrightarrow T' \longrightarrow T \longrightarrow 0$$

splits, which provides a submodule \tilde{T} of A , for which $\tilde{T} \cap C = 0$, and $\tilde{T} \cong T$, whence \tilde{T} is simple. Evidently then the family consisting of all the S_i with \tilde{T} adjoined, violates the maximality in the assumption. It follows that $C = A$, and (d) is proved.

(d) \Rightarrow (b): Proving (b) is equivalent to showing that each submodule B of an arbitrary R -module A is a summand. If $B = A$ then there is nothing to prove, so we might as well assume that B is a proper submodule. The module A is, by hypothesis, a direct sum of simple submodules $\{S_i : i \in I\}$. Now, for each $i \in I$, $S_i \cap B$ is either 0 or S_i , and the first event must occur for some $i \in I$, else $B = A$.

Applying Zorn's Lemma yet again, find a subfamily $\{S_i : i \in J\}$ (with $J \subseteq I$) which is maximal with respect to $B \cap (\sum_{j \in J} S_j) = 0$. We claim that $A = B \oplus (\sum_{j \in J} S_j)$. For if not, then there is an index k in I , so that

$$S_k \cap (B \oplus (\sum_{j \in J} S_j)) = 0.$$

This S_k , adjoined to the set $\{S_j : j \in J\}$ then violates the maximality of the latter. (Think about this!)

This completes the proof of the theorem. ■

Remarks 7.1.8. *Aftermath.* Notice, in the proof that (d) implies (b), that the following were demonstrated:

- (i) If A is a semisimple module, then so is any submodule.

Proof. From the proof above, B is isomorphic to the direct sum of the S_i , with $i \in I \setminus J$, as noted there. ■

- (ii) If A is a semisimple module, then so is any homomorphic image of A .

Proof. Exercise. ■

We note, in closing this section, that the decomposition in Theorem 7.1.7 is not necessarily unique: the Klein 4-group can be written as a direct sum of copies of \mathbb{Z}_2 three ways; an n -dimensional vector space (over any field) admits lots of different direct sum decompositions into one-dimensional subspaces; (infinitely many, if the field is infinite.)

7.2 The Jacobson Density Theorem. En route to the Wedderburn–Artin Theorem (Theorem 7.4.4), we pick up another important theorem in the theory of rings: the Jacobson Density Theorem. It broaches the subject of rings of endomorphisms, which is one we shall not pursue in this course, except to merely skim the surface.

It should be noted that the presentation given here can be generalized to rings without identity. The complications involved are mostly technical ones, which (in the line of our approach) are not worth the trouble. See [Hu74], Chapter IX, which does deal with this generalized setting.

As in the previous section, our universe of discourse is the category of left R -modules and left R -homomorphisms.

Definition 7.2.1. An R -module A is said to be *faithful* if $\text{Ann}(A) = 0$. The ring R is *primitive* if there is a simple, faithful module.

Here is the principal model of a primitive ring.

Example 7.2.2. (See [Hu74], p. 418.) Suppose that D is a division ring, and V is a vector space over D ; (i.e., a D -module.) Let $\text{End}_D(V, V) = \text{Hom}_D(V, V)$; this is a ring, with composition of D -linear transformations.

Now denote $R = \text{End}_D(V, V)$, and regard V as an R -module, where $f \cdot v = f(v)$, for each $f \in R$ and $v \in V$. Observe, first of all, that for each pair of vectors $v, w \in V$, there exists an $f \in R$ such that $f \cdot v = w$; (by Zorn's Lemma, extend the singleton $\{v\}$ to a basis \mathcal{B} for V , and define $f(v) = w$, while $f(v') = 0$, for each $v' \in \mathcal{B}$, $v' \neq v$. Recall that over a division ring, every module is free.) This shows that V is a simple R -module. It is also faithful, as $f \cdot v = 0$, for all $v \in V$, clearly implies that $f = 0$.

Definition 7.2.3. A ring R is said to be *simple* if it has no nonzero proper twosided ideals.

The following are easy to prove; we provide sketches. The proofs can be found on p. 418 of [Hu74].

Proposition 7.2.4. (a) *A simple ring is primitive.*

(b) *If R is commutative, then it is primitive if and only if it is a field.*

Proof. (a) Let $A = R/M$, where M is a maximal left ideal. Then it should be clear that A is simple; it is faithful because $\text{Ann}(A) \leq M$, and the annihilator is a twosided ideal, which must therefore be trivial.

(b) It is clear that a field is simple, hence primitive. Conversely, if R is commutative, and B is any cyclic R -module, then $\text{Ann}(B) = \text{Ann}(b)$, with $B = \langle b \rangle$. Thus, if B is simple and faithful it must be of the form $B \cong R/M$, with M a maximal ideal, and M is necessarily 0. This means that R is a field. ■

Definition & Remarks 7.2.5. Suppose that V is a vector space over the division ring D , and regard V as a left $\text{End}_D(V, V)$ -module, as in Example 7.2.2. Let R be a subring of $\text{End}_D(V, V)$. We say that R is a *dense ring of endomorphisms* if for each positive integer n , and each D -linearly independent set $\{v_1, \dots, v_n\}$ in V , and every subset $\{w_1, \dots, w_n\}$, there is an $f \in R$ for which $f(v_i) = w_i$, for each $i = 1, 2, \dots, n$.

If V is finite dimensional over D , then it can be difficult to find proper subrings of $\text{End}_D(V, V)$ which are dense; see Corollary 7.2.8. Here is an example in the infinite dimensional case.

Suppose that V is infinite dimensional over D , and let S be the subring of $R = \text{End}_D(V, V)$ generated by $1 \in R$ and all the D -linear transformations of finite rank. (A D -linear transformation has *finite rank* if its image is finite dimensional.) We leave it to the reader to check that S is dense.

Observe from the definition of density, by interpreting the case $n = 1$, that if R is a dense subring of $\text{End}_D(V, V)$, then $Rv = V$, for each $v \neq 0$. Thus, V is automatically R -simple. As in Example 7.2.2, $\text{Ann}(V) = 0$ as well, and so V is a faithful R -module, whence any dense ring of endomorphisms is primitive.

The Jacobson Density Theorem states that the converse is true.

Hot Air 7.2.6. *Make a Note.* Suppose that S is a simple R -module. Then $D = \text{Hom}(S, S)$ is a division ring, by Proposition 7.1.2. We may then consider S as a vector space over D , by setting $d \cdot s = d(s)$, for all $d \in D$ and $s \in S$.

Here is the main theorem; compare with Theorem 1.12, p. 420, in [Hu74].

Theorem 7.2.7. (The Jacobson Density Theorem.) *Suppose that R is a primitive ring, and that S is a faithful, simple R -module. Then, viewing R as a vector space over $D = \text{Hom}(S, S)$, R is isomorphic to a dense subring of $\text{End}_D(S, S)$.*

Proof. Minus a technical observation, which we shall leave as an exercise (7.2.9), we outline the essential parts of the argument.

For each $r \in R$, let $\phi_r : S \rightarrow S$ denote the map $\phi_r(s) = rs$, for each $s \in S$. Observe that ϕ_r is a D -endomorphism of S : for if $d \in D$, then

$$\phi_r(d \cdot s) = r(d(s)) = d(rs) = d \cdot (\phi_r(s)),$$

for each $s \in S$.

Next, $\phi_{r+r'} = \phi_r + \phi_{r'}$, and $\phi_{r'r} = \phi_{r'} \cdot \phi_r$, for each $r, r' \in R$; this means that the assignment $\Phi(r) = \phi_r$ is a ring homomorphism from $R \rightarrow \text{End}_D(S, S)$. Because S is a faithful R -module, the map Φ is one-to-one.

What remains then is to show that the image $\Phi(R)$ is dense. Suppose then that $\{v_1, \dots, v_n\}$ is a D -linearly independent set in S , and $\{w_1, \dots, w_n\}$ is any subset of S . For each $i \in I$, let V_i be the D -subspace of S spanned by all the v_j except v_i . As the v_i are D -linearly independent, it follows that $v_j \notin V_j$.

We now require the lemma which was alluded to at the start, and the reader is referred to it, in Exercise 7.2.9. Here it applies as follows: there is an element $r_i \in R$ such that $r_i v_i \neq 0$ and $r_i V_i = 0$. As each $r_i v_i \neq 0$ and S is simple, $R r_i v_i = S$. This means that there exists a $t_i \in R$, such that $t_i r_i v_i = w_i$ (for each $i = 1, 2, \dots, n$.) Now let $r = t_1 r_1 + \dots + t_n r_n$. The reader then easily checks that r does what we want: $r v_i = w_i$, for each i . ■

The next corollary, due to Wedderburn, completes the story of dense subrings, in the finite dimensional case, as foreshadowed in 7.2.5. The proof should be clear from the definition of density.

Corollary 7.2.8. *Suppose that S is a faithful, simple R -module. Let $D = \text{Hom}(S, S)$, and suppose that S is finite dimensional over D . Then $R = \text{End}_D(S, S)$; (that is, the embedding of the Density Theorem is surjective.)*

Exercise 7.2.9. Suppose that S is a simple R -module. View S as a vector space over the division ring $D = \text{Hom}(S, S)$. If V is any finite dimensional D -subspace of S , and $a \in S \setminus V$, then there is an element $r \in R$ so that $rV = 0$ but $ra \neq 0$. (This is Lemma 1.11, p. 420, in [Hu74].)

The reader is encouraged to look at Exercise 1, pp. 423-24, in [Hu74].

7.3 Semisimple Rings. In this section R is a ring with identity. We shall begin to compare left oriented features of R with their right oriented counterparts.

We begin with a theorem, which defines the so-called Jacobson radical. The reader would do well, at this point, to review the remarks in 7.1.4.

Theorem 7.3.1. *The following all describe the same subset $\mathfrak{J}(R)$ of R .*

(a)

$$\{a \in R : 1 + ra \text{ is left invertible, } \forall r \in R\}.$$

(b)

$$\bigcap \{N \leq R : N \text{ is the annihilator of a simple } R\text{-module}\}.$$

(c)

$$\bigcap \{M \leq R : M \text{ is a maximal left ideal}\}.$$

$\mathfrak{J}(R)$ is a twosided ideal of R .

Proof. Observe that, since annihilators of modules are twosided, the set described in (b) is a twosided ideal. So, if all else works out, the final claim surely follows.

Let $J^{(a)}$, $J^{(b)}$ and $J^{(c)}$ be the sets described in (a), (b) and (c), respectively. First, realize that, since every annihilator of a simple module is an intersection of maximal left ideals, it should be clear that $J^{(c)} \leq J^{(b)}$.

If a fails to be in the maximal left ideal M , then one can express $1 = ra + m$, where $r \in R$ and $m \in M$, but then $1 + (-r)a \in M$, which then cannot be left invertible. Thus, $J^{(a)} \leq J^{(c)}$. Conversely, if there is an element $b \in R$ such that $1 + ba$ is not left invertible, then $1 + ba$ generates a proper left ideal, which is then contained in some maximal left ideal N . But then $a \notin N$, for otherwise $ba \in N$, and this implies that $1 \in N$, a contradiction. Hence $J^{(a)} = J^{(c)}$.

Finally, suppose that M is a maximal left ideal; form the simple left R -module $A = R/M$. Claim: if J is the intersection of all the maximal left ideals N for which $A \cong R/N$, then $\text{Ann}(A) = J$. For if $c \in J$, then for each $x \in A$, $x \neq 0$, $Rx = A$, whence $R/N \cong A$, where $N = \ker(r \mapsto rx)$. Since $N = \text{Ann}(x)$, it follows (since $J \leq N$) that $cx = 0$, and so $J \leq \text{Ann}(A)$.

Conversely, if $s \notin J$, there is a maximal left ideal K which does not contain s , yet $R/K \cong A$. The element $a \in A$ corresponding to the coset $K + s$ is therefore not zero, and $s(K + 1) = a$, which shows that $s \notin \text{Ann}(A)$. This proves the claim in the previous paragraph.

Conclusion: the relation \sim , linking two maximal left ideals M and N provided $A/M \cong A/N$, is an equivalence relation, and the intersection of all the members of one equivalence class is the annihilator of a simple R -module. Thus, $J^{(c)} = J^{(b)}$, and we're done. ■

Definition 7.3.2. The ideal $\mathfrak{J}(R)$ defined by means of Theorem 7.3.1 is called the *Jacobson radical* of R . So far its definition appears to be left oriented. The next proposition establishes that it is not; $\mathfrak{J}(R)$ is also each of the following:

(i) the intersection of all maximal right ideals;

(ii) the intersection of all annihilators of simple right R -modules;

(iii)

$$\mathfrak{J}(R) = \{a \in R : 1 + ar \text{ is right invertible, } \forall r \in R\}.$$

By left–right duality, (i), (ii) and (iii) certainly describe the same set.

Remarks 7.3.3. *Picking up Loose Ends.* Before proceeding, we ought to join the remarks in 7.1.4 with some of the arguments in the proof of Theorem 7.3.1, by way of a summary.

If S is a simple left module then $\text{Ann}(S)$ is the intersection of maximal left ideals M_i ($i \in I$), so that any two R/M_i are isomorphic. Conversely, if M is a maximal left ideal, and $\text{Iso}(M)$ is the set of all maximal left ideals N for which $R/N \cong R/M$, then $\bigcap \text{Iso}(M)$ is the annihilator of the simple module R/M .

Proposition 7.3.4. *Suppose that I is a left ideal of R , and that for each $a \in I$, $1 + a$ is left invertible. Then $1 + a$ is right invertible, for each $a \in I$.*

Proof. Suppose that each element of I has the stated property. Observe the following: $1 + y$ is left invertible if and only if there exists an $x \in R$ such that $x + y + xy = 0$. (To see this, in one direction, suppose that $1 + y$ is left invertible, and write the left inverse as $1 + x$. It is easy to verify that $x + y + xy = 0$. The converse should be obvious.)

So suppose now that $a \in I$. Find $b \in R$ so that $a + b + ba = 0$; note that $b = -(1 + b)a \in I$. So there exists a $c \in R$ so that $c + b + cb = 0$. An easy computation then yields that $a = c$. Thus, $a + b + ab = 0$, whence $(1 + a)(1 + b) = 1$, meaning that $1 + a$ is right invertible. ■

Corollary 7.3.5. *For each ring R with identity,*

$$\mathfrak{J}(R) = \{ a \in R : 1 + ar \text{ is right invertible, } \forall r \in R \}.$$

Proof. Since Proposition 7.3.4 is symmetric, it suffices to show that

$$\mathfrak{J}(R) \leq \{ a \in R : 1 + ar \text{ is right invertible, } \forall r \in R \}.$$

If $a \in \mathfrak{J}(R)$, then according to Theorem 7.3.1, and because $\mathfrak{J}(R)$ is a right ideal as well, $1 + ar$ is left invertible, for each $r \in R$. By Proposition 7.3.4, $1 + ar$ is also right invertible. ■

Definition 7.3.6. If R is a ring, for which $\mathfrak{J}(R) = 0$, then we say that R is *J -semisimple*. We use the unorthodox “ J -semisimple” to avoid confusion with the notion of semisimplicity introduced just prior to Theorem 7.1.7, for modules. Part of the Wedderburn–Artin Theorem (Theorem 7.4.4) says that every R -module is semisimple if and only if R is J -semisimple and has the descending chain condition on left ideals. (Notice the remark in [Hu74], on p. 429, about the possible confusion involving the term “semisimple”. No more will be said on the subject in this development.)

Hot Air 7.3.7. *Swirling!* Recall that a ring R is left primitive if it has a faithful, simple left R -module S . Since S is faithful, this means that there is a set of maximal left ideals $\{ M_i : i \in I \}$, so that each $R/M_i \cong S$ and $\bigcap_i M_i = 0$. Conversely, if there is a family of maximal left ideals $(M_i)_{i \in I}$ with the stated properties, then R/M_i (for any of the indices i) is a faithful, simple left R -module.

From the remarks made in 7.3.3 it now follows that:

Proposition 7.3.8. *A ring R is J -semisimple if and only if it is a subdirect product of left primitive rings, and also if and only if it is a subdirect product of right primitive rings.*

(But there are examples of left primitive rings which are not right primitive; see the reference in [Hu74], p. 418.)

Proof. The left oriented equivalence suffices. Recall, as well, that to say that R is a subdirect product of left primitive rings, is the same as saying that there is a family of ideals $\{K_i : i \in I\}$, with R/K_i primitive, and $\bigcap_i K_i = 0$.

In view of Theorem 7.3.1, if $\mathfrak{J}(R) = 0$, and $\{K_i : i \in I\}$ is the set of all annihilators of left simple R -modules, then $\bigcap_i K_i = 0$, and each R/K_i is left primitive. Conversely, if there is a family \mathcal{L} of ideals of R , with trivial intersection, and so that R/J is left primitive, for each $J \in \mathcal{L}$, then J is an intersection of maximal left ideals of R ; (see 7.3.3). By Theorem 7.3.1, $\mathfrak{J}(R) = 0$. ■

The following should also be clear:

Proposition 7.3.9. *If R is a ring, then $R/\mathfrak{J}(R)$ is J -semisimple.*

The following two exercises are 8 and 13, respectively, on p. 433, [Hu74].

Exercise 7.3.10. Let R be the ring of all upper triangular $n \times n$ matrices over a field F . Compute $\mathfrak{J}(R)$ and prove that $R/\mathfrak{J}(R)$ is isomorphic (as a ring) to the direct product F^n .

Exercise 7.3.11. For any ring R with identity,

$$\mathfrak{J}(\text{Mat}_n(R)) = \text{Mat}_n(\mathfrak{J}(R)).$$

($\text{Mat}_n(R)$ stands for the ring of all $n \times n$ matrices over the ring R . The result is valid for rings without an identity; have a look at [Hu74] for hints.)

For commutative rings we have the following relationship between the nil (or prime) radical and the Jacobson radical. First, it is possible, in the noncommutative case, to define an analogue of $n(A)$, but there are complications, with the definition of “prime” ideal, which we elect not to raise here.

Exercise 7.3.12. If R is a commutative ring with identity then $n(R) \leq \mathfrak{J}(R)$, but they need not be equal. (Give an example.)

Exercise 7.3.13. What is the formulation of Proposition 7.3.8 for commutative rings? (Hint: When is a commutative ring primitive?)

7.4 The Wedderburn–Artin Theorem.

Definition 7.4.1. Suppose that R is a ring with identity. We say that R is *left* (resp. *right*) *Artinian* if there is no descending sequence $J_1 > J_2 > \cdots$ of left (resp. right) ideals of R . Equivalently, R is left Artinian if and only if every nonempty set of left ideals has a minimal element. (Compare with Exercise 1.3.7, about the ascending chain condition vs. maximality.)

Although we shall have more to say about Artinian rings, the following are easily established.

Exercise 7.4.2. Suppose that R is a left Artinian ring with identity.

- (a) Show that if R is an integral domain, then it is a field.
- (b) If I is any ideal of R , then R/I is left Artinian.
- (c) A subring T (with identity) of R need not be left Artinian. Give an example.

We shall soon invoke the Jacobson Density Theorem (Theorem 7.2.7); the reader should review it, and Corollary 7.2.8. Indeed, the following lemma will be useful in the proof of part of the Wedderburn–Artin Theorem.

Lemma 7.4.3. *Suppose that R is a dense subring of $\text{End}_D(V)$, where D is a division ring, and V is a vector space over D . If R is left Artinian then D is finite dimensional over D , whence $R = \text{End}_D(V)$.*

Proof. The last conclusion follows from Corollary 7.2.8. So suppose that V has an infinite linearly independent set $\{v_1, v_2, \dots\}$. View V as a left R -module, and let L_k be the annihilator of $\{v_1, \dots, v_k\}$. It should be clear that $L_{n+1} \leq L_n$. By the density of R , there is an $r \in R$ such that $rv_i = 0$, ($1 \leq i \leq n$) and $rv_{n+1} \neq 0$, which proves that the L_n form an infinite descending chain of left ideals, a contradiction. ■

Though a few preliminaries are still needed, here is the theorem that we have been working for. The proof, intricate though ultimately straightforward in its details, is taken from a very nice text, that of James P. Jans ([Ja64]). It is unfortunate that this very readable introduction to the subject of ring theory and homological algebra is out of print.

Theorem 7.4.4. *The Wedderburn–Artin Theorem. For a ring R with identity, these are equivalent statements:*

- (a^l) *Each left R -module is projective.*
- (b^l) *Each short exact sequence of left R -modules splits.*
- (c^l) *Each left R -module is injective.*
- (d^l) *Each left R -module is a direct sum of simple submodules.*
- (e^l) *${}_R R$ is a direct sum of simple submodules.*
- (f^l) *${}_R R = L_1 \oplus L_2 \oplus \dots \oplus L_k$, where each L_i is a minimal left ideal and of the form $L_i = Re_i$, for a suitable idempotent e_i , so that $e_i e_j = 0$, whenever $i \neq j$, and $1 = e_1 + \dots + e_k$.*
- (g) *$R \cong R_1 \times R_2 \times \dots \times R_m$ (as rings) of simple rings R_i each of which is isomorphic to $\text{Mat}_{k(i)}(D_i)$, for some division ring D_i , and a suitable positive integer $k(i)$.*
- (h^l) *R is left Artinian and J -semisimple.*

Moreover, (g) is also equivalent to (a^r) through (f^r), and (h^r), the right oriented versions of (a^l) through (f^l) and (h^l).

Note that *minimal left* (resp. *right*) *ideal* means: minimal among the nonzero left (resp. right) ideals of the ring. Observe that a left (resp. right) ideal is minimal precisely when it is simple as a left (resp. right) submodule of ${}_R R$ (resp. R_R).

We record the following immediate corollary, for emphasis.

Corollary 7.4.5. *For a ring R with identity, R is left Artinian and J -semisimple if and only if it is right Artinian and J -semisimple.*

Definition 7.4.6. A ring R with identity is *left* (resp. *right*) *Noetherian* if there is no infinite ascending sequence of left (resp. right) ideals.

From (g) in Theorem 7.4.4, we also have the following corollary.

Corollary 7.4.7. *If R is left (or right) Artinian and J -semisimple, then it is both left and right Noetherian.*

Proof. Any full matrix ring $\text{Mat}_n(D)$ is evidently finite dimensional over the division ring D of entries. Moreover, each left ideal is a D -subspace, which makes it clear that $\text{Mat}_n(D)$ is Noetherian. It should be easy to see that, a finite product of rings which are left (resp. right) Noetherian, preserves this property. The corollary then follows from the Wedderburn–Artin Theorem. ■

The rest of this section concerns itself with the proof of Theorem 7.4.4, and it closes with a few exercises.

Recall that we have already proved the equivalence of the first four conditions (Theorem 7.1.7). It should also be obvious that (d^l) implies (e^l) . The final claim of the theorem, about the “right vs. left” issue, should also be self-evident.

Proof. Of the Wedderburn–Artin Theorem. $(e^l) \Rightarrow (f^l)$: Suppose that ${}_R R$ is a direct sum of left ideals $\{L_i : i \in I\}$, which are simple as left R -submodules. As has already been noted, each L_i is a minimal left ideal. Write

$$1 = e_1 + e_2 + \cdots + e_k,$$

with $e_j \in L_{i_j}$, uniquely, as prescribed by the direct sum. For each $j = 1, 2, \dots, k$, we have

$$e_i = e_i e_1 + \cdots + e_i^2 + \cdots + e_i e_k,$$

and $e_i e_j \in L_{i_j}$. The uniqueness of expressions in a direct sum then implies that $e_i = e_i^2$, and $e_i e_j = 0$, if $i \neq j$. By a similar argument, it also follows that

$$\{L_i : i \in I\} = \{L_{i_1}, L_{i_2}, \dots, L_{i_k}\};$$

in particular, there are only finitely many such left ideals.

Now relabel the set of minimal left ideals in the direct sum as $\{L_1, L_2, \dots, L_k\}$. Note that $L_i = Re_i$, for each $i = 1, 2, \dots, k$, by minimality. This proves that (e^l) implies (f^l) .

Note: an idempotent $e \neq 0$ which generates a minimal ideal is often referred to as a *primitive idempotent*.

$(f^l) \Rightarrow (d^l)$: It follows from (f^l) that any free left R -module is semisimple. Since every R -module can be obtained as a homomorphic image of a free one, one may then apply 7.1.8(ii) to conclude (d^l) .

To show that $(f^l) \Rightarrow (g)$, the following lemma is useful. We leave its proof as an exercise.

Exercise 7.4.8. If L is any minimal left ideal of the ring R , and M is any simple left R -module, then either L is R -isomorphic to M , or else $LM = 0$.

$(f^l) \Rightarrow (g)$: Assume the notation of (f^l) , as stated in the theorem. By the preceding exercise, if L is any minimal left ideal of R , then L must be R -isomorphic to one of the L_i in the decomposition. Conclusion: among the minimal left ideals there is a finite number of R -isomorphism classes, which we number 1 through n . Let R_i denote the left ideal generated by all the minimal left ideals belonging to the i -th isomorphism class. Let’s emphasize: from Exercise 7.4.8, if L and L' come from distinct

isomorphism classes, then $LL' = 0$. Thus, $R_iR_j = 0$, if i and j are distinct. It follows that each R_i is a twosided ideal, and that $R = R_1 \oplus R_2 \oplus \cdots \oplus R_m$, because the minimal left ideals generate ${}_R R$. Observe that, for each index i , the left ideals of R_i are just the R -left ideals contained in R_i , and so R_i has exactly one isomorphism class of minimal left ideals, and it is generated by them (as an R - or R_i -module).

For the remainder of the proof of this implication let $T = R_i$. As was done in the proof that (e^l) implies (f^l) , $T = L_1 \oplus \cdots \oplus L_m$, where each L_i is a minimal T -left ideal. Let $D = \text{Hom}_{T\text{-Mod}}(L_1, L_1)$; since L_1 is left T -simple, D is a division ring. ■

We shall give two proofs that T is a full matrix ring. The first, which appeals to the Jacobson Density Theorem (Theorem 7.2.7), is more elegant; some information (namely, that m in the above decomposition of T , that is, the number of minimal left ideals, is equal to the the dimension of L_1 over D) gets lost. The second proof is more tedious, but it preserves that information.

Proof. First proof that T is a Full Matrix Ring. Let's first use the Density Theorem. T is primitive; indeed, L_1 is a faithful, simple T -module. By the Density Theorem, T is isomorphic to a dense subring of $\text{End}_D(L_1)$, and if we can demonstrate that T is left Artinian, then we may apply Lemma 7.4.3 to show that $T = \text{End}_D(L_1)$. To that end, suppose that L is a left ideal of T ; then $L = LL_1 + LL_2 + \cdots + LL_m$, and each LL_i is a left ideal of T , contained in L_i , and therefore 0 or equal to L_i . This should suffice to convince the reader that T is left Artinian. ■

Proof. Second proof that T is a Full Matrix Ring. We will show that T is isomorphic to $\text{Mat}_m(D)$, and note well: m is the number of minimal left ideals of T . As in the proof that (e^l) implies (f^l) , we have $e_i \in L_i$, each e_i idempotent, with $1 = e_1 + \cdots + e_m$, and $e_i e_j = 0$, for distinct indices.

Suppose that $f \in \text{Hom}_{T\text{-Mod}}(L_i, L_j)$. As $f(e_i) \in Te_j$, we have that $f(e_i) = re_j$; note also that

$$f(e_i) = f(e_i^2) = e_i re_j.$$

On the other hand, if $x \in L_i$, then $x = xe_i$ (check this!), and so $f(x) = xe_i re_j$. To emphasize, as well as summarize, f acts by right multiplication, by a suitable element of L_j . It is then easy to verify that, as an abelian group

$$\text{Hom}_{T\text{-Mod}}(L_i, L_j) \cong e_i T e_j.$$

Next, suppose that $f : L_i \rightarrow L_j$ is a left T -homomorphism, $g : L_j \rightarrow L_k$ as well, effected by right multiplications $e_i re_j$ and $e_j se_k$, respectively. Then $g \cdot f$ is effected by right multiplication by $e_i re_j se_k$. (Note that this order of multiplication reverses composition.)

From now on, we identify a homomorphism $L_i \rightarrow L_j$ with the element from $e_i T e_j$ which effects it by right multiplication.

Now choose, for each $i = 1, 2, \dots, m$, a fixed isomorphism $g_i : L_1 \rightarrow L_i$. As an element of $e_1 T e_i$, it follows that $g_i^{-1} \in e_i T e_1$.

At last we define the (ring) isomorphism from T onto $\text{Mat}_m(D)$. For each $t \in T$, consider

$$g_i(e_i t e_j) g_j^{-1} \in e_1 T e_1 = L_1;$$

this is, in effect, an element in D . Define $M : T \rightarrow \text{Mat}_m(D)$ by

$$M(t) = (g_i(e_i t e_j) g_j^{-1})_{i,j}.$$

One has to prove that

- (i) M preserves addition and multiplication;
- (ii) M is one-to-one;
- (iii) M is onto $\text{Mat}_m(D)$.

We leave the additivity to the reader. As to the product, observe that, for $t_1, t_2 \in T$,

$$\begin{aligned} \sum_{k=1}^n (g_i e_i t_1 e_k g_k^{-1})(g_k e_k t_2 e_j g_j^{-1}) &= \sum_{k=1}^n g_i e_i (t_1 e_k t_2) e_j g_j^{-1} \\ &= g_i e_i t_1 \left(\sum_{k=1}^n e_k \right) t_2 e_j g_j^{-1} \\ &= g_i e_i t_1 t_2 e_j g_j^{-1}, \end{aligned}$$

which is the (i, j) -entry of $M(t_1 t_2)$. Thus,

$$M(t_1 t_2) = M(t_1)M(t_2),$$

and M is a ring homomorphism.

If $M(t) = 0$, then $g_i(e_i t e_j)g_j^{-1} = 0$, for each pair of indices; this means that $e_i t e_j = 0$. Summing over all e_i on the left, one gets that $t e_j = 0$, for each index j , and summing again, in the same way, on the right, $t = 0$, proving that M is one-to-one.

Finally, to show M is surjective, it suffices to prove, given $d \in D$, and a fixed position (i, j) , that there is a $t \in T$, so that $M(t) = 0$ in all positions, except at (i, j) , which has the entry d . Note that $g_i^{-1} d g_j : L_i \rightarrow L_j$ is a T -homomorphism, for which there is an element $t \in T$, so that

$$g_i^{-1} d g_j = e_i t e_j.$$

It follows that $d = g_i e_i t e_j g_j^{-1}$. Verify that $M(e_i t e_j)$ does what we want.

This proves that (f^l) implies (g). ■

(g) ⇒ (f^l): It suffices to show that every complete ring of $n \times n$ matrices over a division ring is a direct sum of minimal left ideals. We outline this in the following exercise:

Exercise 7.4.9. Let D be a division ring. In $\text{Mat}_n(D)$, let L_i be the set of all matrices which are zero everywhere, except in the i -th column. Show that each L_i is a left ideal, and minimal among left ideals.

Remark 7.4.10. *Progress Report.* At this point we have the equivalence of (a^l) through (f^l) and (g), as well as (a^r) through (f^r) and (g). To show the equivalence to (h^l) (resp. (h^r)), we will first work on Artinian J -semisimple rings.

Definition & Remarks 7.4.11. Recall that an element $a \in R$ is *nilpotent* if $a^k = 0$, for some positive integer k . A left (resp. right, resp. twosided) ideal I of R is *nil* if every element of I is nilpotent. I is said to be *nilpotent* if $I^k = 0$, for a suitable positive integer k .

Obviously, every nilpotent left (resp. right, resp. twosided) ideal is nil, but, in general, the converse is false. [Hu74] gives an example; it is Exercise 11, p. 433. However, the ring in question does not have an identity. Nevertheless, it can be modified to produce an example with identity.

With respect to the Jacobson radical, we have the following results concerning nilpotency.

Proposition 7.4.12. *Let R be a ring with identity.*

- (a) Every nil left (or right, or twosided) ideal is contained in $\mathfrak{J}(R)$.
- (b) If R is left (resp. right) Artinian, then $\mathfrak{J}(R)$ is nilpotent. It follows that every nil left (resp. right, resp. twosided) ideal is nilpotent.
- (c) If R is commutative and Artinian then

$$n(R) = \mathfrak{J}(R).$$

Proof. (a) Suppose that I is a nil left ideal and $a \in I$; then a is nilpotent, and we suppose that $a^k = 0$. It is easily seen that $1 + a$ is invertible, and

$$(1 + a)^{-1} = (1 - a + \cdots + (-1)^{k-1} a^{k-1}).$$

Furthermore, $ra \in I$, for each $r \in R$, and the above can then be applied to ra . Thus, $1 + ra$ is invertible for each $r \in R$, whence $a \in \mathfrak{J}(R)$.

(b) Denote $J = \mathfrak{J}(R)$, for purposes of this proof. Consider the sequence $J \geq J^2 \geq \cdots$; as R is left Artinian, $J^k = J^{k+1}$, for some positive integer k . We claim that $J^k = 0$. If not, then the set

$$\mathcal{P}(J) = \{I \leq R : I \text{ is a left ideal, } J^k I \neq 0\}$$

is nonempty. Since R is left Artinian, $\mathcal{P}(J)$ contains a minimal element S . As $J^k S \neq 0$, there is an element $a \in S$, such that $J^k a \neq 0$. Verify that $J^k a$ is a left ideal, and that $J^k a \in \mathcal{P}(J)$. This implies that $J^k a = J^k S$, by minimality. This says that $ra = a$, for some element $r \in J^k \leq J$, and so $-r + b + b(-r) = 0$, for some $b \in R$; (because $1 - r$ is left invertible.) Multiplying on the right by a one gets:

$$0 = -a + ba - ba = -a,$$

a contradiction.

It follows that $\mathfrak{J}(R)$ is nilpotent. The rest of (b) is obvious.

(c) This follows immediately, by the definition of $n(R)$, from (b) and from Exercise 7.3.12. ■
Putting together the above, we get the following corollary.

Corollary 7.4.13. *Suppose R is left Artinian. Then the following are equivalent.*

- (a) R is J -semisimple.
- (b) R has no nonzero nil left ideals.
- (c) R has no nonzero nilpotent left ideals.

Remark 7.4.14. *Caution!* Corollary 7.4.13 does *not* say that R has no non-zero nilpotent elements! $\text{Mat}_n(D)$, the ring of all $n \times n$ matrices over the division ring D is simple, both left and right Artinian (as Theorem 7.4.4 asserts, and, in any case, one can verify independently), but it certainly has nilpotent elements, for any $n \geq 2$.

Incidentally, [Hu74] states that $\text{Mat}_n(D)$ is both left and right Artinian, and proves it via the theory of composition series, which we shall omit here. (See Chapter VIII, Section 1, and, in particular, Corollary 1.12.)

Remark 7.4.15. *Preparing for the Finish ...* To show that (h^l) in Theorem 7.4.4 follows from the others, it suffices to show that (a^l) through (f^l) and (g) imply that R is left Artinian, and has no nonzero nilpotent left ideals, and conversely.

Proof. (a^l) through (g) ⇒ (h^l): By (g), and since each ring $\text{Mat}_n(D)$ (with D a division ring) is J -semisimple, it follows that R is J -semisimple, because a finite direct product of J -semisimple rings is J -semisimple. (Recall: the D -vector space D^n is a faithful and simple left $\text{Mat}_n(D)$ -module; see 7.2.2.)

Note that in the proof that (f^l) implies (g) it was already shown that a ring which is a finite direct sum of minimal left ideals must be left Artinian.

(h^l) ⇒ the Rest: We actually show (f^l) holds.

First, if J is any nonzero left ideal, then there is a minimal left ideal $L \leq J$, and an idempotent $e \in L$, so that $Re = L$. To see this, apply the left Artinian feature of R to the set of all nonzero left ideals of R contained in J , to get a minimal one, L . Since $L^2 \neq 0$, and so $L^2 = L$. This implies that $Lx \neq 0$, for a suitable $x \in L$, and also the existence of an $e \in L$ such that $ex = x$. Verify that e is idempotent – hint: the map $L \rightarrow Lx$ defined by $a \mapsto ax$ is an R -isomorphism – and that $Re = L$.

Thus, we have at least one minimal left ideal L_1 , and an idempotent e_1 , with $Re_1 = L_1$; letting $K_1 = R(1 - e_1)$, we see that $R = L_1 \oplus K_1$, with $K_1e_1 = 0$. Now, proceed by induction.

Suppose that

$$R = L_1 \oplus L_2 \oplus \cdots \oplus L_k \oplus K_k,$$

where $L_i = Re_i$, with e_i idempotent, so that $e_i e_j = 0$, for distinct indices i and j , and so that $K_k e_i = 0$, for each $i = 1, 2, \dots, k$. We may then find a minimal left ideal $L_{k+1} \leq K_k$, and an idempotent f , such that $L_{k+1} = Rf$ and $K_k = L_{k+1} \oplus K_{k+1}$, with $K_{k+1}f = 0$. Clearly, $f e_i = 0$, for each $i = 1, 2, \dots, k$.

Now, we can't say about the $e_i f$, so let's find a different idempotent: set

$$e_{k+1} = f - (e_1 + \cdots + e_k)f.$$

Then it is easily verified that

(i) $L_{k+1} = Re_{k+1}$, and

(ii)

$$R = L_1 \oplus \cdots \oplus L_{k+1} \oplus K_{k+1},$$

with $e_i e_j = 0$, for distinct i and j , while $K_{k+1} e_i = 0$, for each $i = 1, 2, \dots, k + 1$.

Since the sequence $K_1 > K_2 > \cdots$ must terminate, it follows that $K_j = 0$, for some j , whence $R = L_1 \oplus \cdots \oplus L_j$, and we're done.

The proof of Theorem 7.4.4 is now complete. ■

We finish this section, and the chapter, with an assortment of observations. We leave most verifications to the reader.

The first two items concern uniqueness of the decompositions in (f^l) and (g) in the Wedderburn–Artin Theorem.

Theorem 7.4.16. First Uniqueness Theorem. *Suppose that R is a ring with identity. Suppose also that A is a left R -module, and that*

$$A = S_1 \oplus S_2 \oplus \cdots \oplus S_m = T_1 \oplus T_2 \oplus \cdots \oplus T_n$$

are decompositions of A as direct sums of simple submodules. Then

(i) $m = n$, and

(ii) after a suitable permutation of the T_i we have, $S_i \cong T_i$.

Proof. By induction on m . Clearly, if $m = 1$, then by the simplicity of S_1 , n must be 1 as well. So suppose that $m > 1$. Consider the restriction p_j of the projection map $\pi_j : A \rightarrow T_j$. Each p_j is an R -homomorphism. Therefore, p_j is either zero or an isomorphism, and not all can be the zero map. So choose a $k = 1, 2, \dots, n$ so that p_k is an isomorphism onto T_k . Note that $\ker(\pi_k) \cap S_1 = 0$. Now show that

$$S_2 \oplus \cdots \oplus S_m \cong T_1 \oplus \cdots \oplus T_{k-1} \oplus T_{k+1} \oplus \cdots \oplus T_n,$$

by showing that both are R -isomorphic to A/S_1 . Then apply induction. ■

The matrix rings in (g) of Theorem 7.4.4 are simple; here then is what can be said about direct product decompositions of a ring into simple factors.

Theorem 7.4.17. Second Uniqueness Theorem. *Suppose that R is a ring with identity. If*

$$R = R_1 \times \cdots \times R_m = R'_1 \times \cdots \times R'_n$$

are two direct product decompositions of R , so that each R_i and each R'_j is a simple ring, then

- (i) $m = n$, and
- (ii) after a suitable permutation of the R'_j , we have $R_i = R'_i$.

There is no misprint in (ii); one does get equality of components, not just an isomorphism.

Proof. Observe that, for each $i = 1, 2, \dots, m$, $R_i = \sum_{j=1}^n R_i R'_j$. Each $R_i R'_j$ is a twosided ideal of R_i . By the simplicity of each component, it follows that $R_i = R_i R'_{k_i}$, for a suitable k_i . Moreover, this k_i is unique, as $R_i = R_i R'_k = R_i R'_l$ implies that R_i lies in both R'_k and R'_l , which is absurd. On the other hand, $R_i R'_{k_i}$ is also a twosided ideal of R'_{k_i} , and therefore equal to it. Thus, $R_i = R'_{k_i}$ and the map $i \mapsto k_i$ is a permutation. ■

Exercise 7.4.18. Give an independent proof that a finite direct product of left Artinian rings is left Artinian. Ditto with left Noetherian. (Direct products of two rings will suffice; then scream induction!)

Applying the Wedderburn–Artin Theorem to commutative rings, we have the following formulation. We shall revisit this situation in §12.1.

Theorem 7.4.19. *Suppose that R is a commutative ring with identity. Then the following are equivalent.*

- (a) Each R -module is projective.
- (b) Each short exact sequence of R -modules splits.
- (c) Each R -module is injective.
- (d) Each R -module is a direct sum of simple submodules.
- (e) ${}_R R = L_1 \oplus L_2 \oplus \cdots \oplus L_k$, where each L_i is a minimal ideal and of the form $L_i = Re_i$, for a suitable idempotent e_i , so that $e_i e_j = 0$, whenever $i \neq j$, and $1 = e_1 + \cdots + e_k$.
- (f) $R \cong F_1 \times F_2 \times \cdots \times F_m$ (as rings) of fields F_i .

(g) R is Artinian and J -semisimple.

(h) R is Artinian and has no nonzero nilpotent elements.

Proof. Nothing needs to be said about the equivalence of (a) through (e). They follow from Theorem 7.4.4. Now (f) here is (g) of Theorem 7.4.4, as soon as one realizes that, in view of commutativity, the matrix rings must be 1×1 and over fields. The equivalence of (g) to the preceding ones is also clear from Theorem 7.4.4, and (h) is equivalent to all others by Proposition 7.4.12(c). ■

As an application of the Wedderburn–Artin Theorem, here is an entertaining exercise. In Chapter 8, we shall consider a generalization of the rings in Exercise 7.4.20, namely, the von Neumann regular rings; (refer to 8.2.13).

Exercise 7.4.20. Let R be a finite ring with identity in which $r^3 = r$, for each $r \in R$. Prove that R is necessarily commutative, and a finite direct product of fields which are isomorphic to either \mathbb{F}_2 or \mathbb{F}_3 , the fields of 2 and 3 elements, respectively.

(Hint: First, the law $r^3 = r$ should spell out that R is J -semisimple and (left or right) Artinian; the point being that Theorem 7.4.4 applies. The issue turns around what the matrix rings in the decomposition of R look like. They should be 1×1 matrix rings (Why?). Thus, R is a finite product of division rings. Now Wedderburn’s so-called “Little Theorem” – see the next exercise – says that any finite division ring is a field. Finally, consider a finite field in which $r^3 = r$ holds.)

This “Little Theorem” is astonishingly nontrivial. See [Hu74], p. 462, for additional details.

Exercise 7.4.21. *Wedderburn’s Little Theorem.* Any finite division ring is a field.

Finally, we introduce the so-called group–algebra, to suggest ways in which the preceding material might be applied in the theory of groups.

Definition 7.4.22. Suppose that G is a group, and K is a field. We denote by $K[G]$ the following structure: the elements of $K[G]$ are finite formal sums $\sum_{g \in G} r_g g$, with each $r_g \in K$; one adds the sums term by term:

$$\left(\sum_{g \in G} r_g g\right) + \left(\sum_{g \in G} s_g g\right) = \sum_{g \in G} (r_g + s_g)g;$$

while the product is convolution:

$$\left(\sum_{g \in G} r_g g\right) \cdot \left(\sum_{g \in G} s_g g\right) = \sum_{g \in G} \left(\sum_{ef=g} r_e s_f\right)g.$$

One can think of $K[G]$ as a direct sum of copies of K , indexed over G , so that, additively, $K[G]$ is a K -vector space. $K[G]$ is the *group–algebra of G over K* .

The point of the exercise which follows is to prove Maschke’ Theorem.

Exercise 7.4.23. Suppose that G is a group, and K is a field.

(a) Show that $K[G]$ is a ring with identity. (A good candidate for the multiplicative identity is $\sum_{g \in G} r_g g$, where $r_g = 1$, for $g = e$, and $r_g = 0$, otherwise.)

Assume that G is a finite group. We try to answer the question of when the Wedderburn–Artin Theorem applies to $K[G]$. (It turns out that G is necessarily finite, but that is harder to prove. We do it in the next chapter.)

Consider a short exact sequence of left $K[G]$ -modules

$$(E_M) \quad 0 \longrightarrow A \longrightarrow B \xrightarrow{\beta} C \longrightarrow 0;$$

the question is: under what conditions (on G) does the above always split? Now (E_M) is a short exact sequence of K -vector spaces, and so it splits; that is, there is a K -linear transformation $\alpha : C \longrightarrow B$ so that $\beta \cdot \alpha = 1_C$. When is α a $K[G]$ -homomorphism?

Define $\alpha' : C \longrightarrow B$ by

$$\alpha'(x) = \sum_{g \in G} g^{-1} \alpha(gx);$$

remember: the group elements act as scalars on the modules.

- (b) Calculate the composite $\beta \cdot \alpha'$. Deduce from the calculation that if the characteristic of K does not divide $|G|$, then (E_M) splits.

Thus, the sufficiency in what is known as Maschke's Theorem:

Suppose that G is a finite group. Then $K[G]$ is J -semisimple and left Artinian if and only if the characteristic of K does not divide the order of G .

- (c) Prove the necessity: if the characteristic of K does divide $|G|$, show that $\sum_{g \in G} g$ lies in the Jacobson radical of $K[G]$.

The reader should also look at [Hu74], pp. 453–455. It is shown there that when the J -semisimple Artinian ring is also a vector space over an algebraically closed field K (that is, a K -algebra), then the matrix rings predicted by the Wedderburn–Artin Theorem have their entries in K .

8. Rings and Homology: Injectives and Flatness

In this chapter we prove the existence of injectives over an arbitrary ring with identity, and more: that any R -module can be embedded in an injective R -module in a minimal way. In a real sense the heart of the chapter is the theory of flat modules; they correct the defect in the way that tensor products preserve short exact sequences. A crucial role is played by a kind of duality between flat modules and injective ones. Von Neumann regular rings rear their heads and play a prominent role.

8.1 Injective Hulls. This section divides into two parts. First, we show that every module can be embedded in an injective one, and then that it can be done in a minimal way. The arguments use the characterization of injective abelian groups as the divisible ones (Theorem 6.3.14). Our main reference is the book by J. Lambek ([L86]).

Unless the contrary is mentioned, all modules in this section are right modules over an arbitrary ring with identity R .

Definition & Remarks 8.1.1. For the record, let us recall the definition of injective modules: an R -module J is said to be *injective* if for each one-to-one R -homomorphism $\alpha : A \rightarrow B$, and each R -homomorphism $f : A \rightarrow J$, there is an R -homomorphism $f^* : B \rightarrow J$ such that $f^* \cdot \alpha = f$. Since the α 's amount essentially to embedding an R -module as a submodule of another, the definition may be recast as follows: J is injective if, for each R -submodule A of B , and each R -homomorphism $f : A \rightarrow J$, there is an extension of f to B .

The reader should recall Proposition 6.3.13 stating that

- (a) a direct product of R -modules is injective if and only if each factor is injective, and
- (b) if J is an injective R -module then it is a summand in every R -module which contains it as a submodule.

We showed in Theorem 6.3.14 that, if $R = \mathbb{Z}$, then an abelian group is injective if and only if it is divisible. To arrive at the fact that, over a ring R , every module can be embedded in an injective one, it turns out to be useful to prove it first for abelian groups. So we pick up the thread with abelian groups.

We begin with two properties of divisible groups, which do not translate into general theorems for injectives.

Exercise 8.1.2. (a) Show that any direct sum of divisible abelian groups is divisible.

- (b) Show that if $f : G \rightarrow H$ is a surjective homomorphism of abelian groups, and G is divisible, then H is also divisible.

Proposition 8.1.3. *Every abelian group G can be embedded as a subgroup of a divisible abelian group.*

Proof. We begin with the simple observation that \mathbb{Z} itself, embedded in \mathbb{Q} , is embedded in a divisible abelian group. By taking direct sums, and applying 8.1.2(a), it follows that every free abelian group can be embedded in a divisible abelian group.

Now, suppose that G is an abelian group. Write G as a homomorphic image $h : F \rightarrow G$ of some free abelian group. By the first paragraph F is a subgroup of a divisible group D ; in forming $D/\ker(h)$ we get a divisible group, according to **8.1.2(b)**. Finally, G is isomorphic to $F/\ker(h)$, which is a subgroup of $D/\ker(h)$. ■

Now we use the preceding proposition to establish the embeddability, for arbitrary modules over a ring R . We need a lemma, however, which is a rather clever test for injectivity.

Lemma 8.1.4. *Suppose that J is an R -module. Then J is injective if and only if for each right ideal K of R , and each R -homomorphism $f : K \rightarrow J$, there is an extension of f to R .*

Proof. The necessity is trivial, so we move on to the sufficiency. The reader might recognize in the argument the proof of Theorem **6.3.14**.

Suppose that J satisfies the stated condition, with respect to right ideals of the ring. Let A be a submodule of B , and $f : A \rightarrow J$ be an R -homomorphism. We denote by $\text{Ex}(f)$ the set of all pairs (C, h) , in which C is a submodule of B containing A , and $h : C \rightarrow J$ is a homomorphism which extends f . Setting $(C, h) \leq (C', h')$ by $C \leq C'$, and so that h' , restricted to C is h , we get a partial order. We shall leave it to the reader to verify (as in the proof of **6.3.14**) that Zorn's Lemma applies to $\text{Ex}(f)$.

Now, let (B^*, g) be a maximal member of $\text{Ex}(f)$. We wish to show that $B^* = B$.

If not so, then pick $c \in B \setminus B^*$. Let $K = \{r \in R : cr \in B^*\}$; it is readily verified that K is a right ideal of R . Next, define $\phi : K \rightarrow J$ by $\phi(r) = g(cr)$, and check that this defines an R -homomorphism. By our assumption, there is an extension $\phi^* : R \rightarrow J$ of ϕ ; let $z = \phi^*(1)$ and set

$$f^*(b + cr) = g(b) + zr, \forall b \in B^*, r \in R.$$

If this is a well defined map, then it is easy to verify that it is an R -homomorphism extending g to the submodule $B^* + cR$, which is larger than B^* . This is a violation of the maximality of (B^*, g) , and the contradiction then implies that $B^* = B$, and finishes the proof.

So what remains is to show that f^* is well defined: if $b + cr = b' + cr'$, with $b, b' \in B^*$, and $r, r' \in R$, then $c(r - r') = b' - b \in B^*$, and therefore $r - r' \in K$. Thus,

$$g(b' - b) = \phi(r - r') = \phi^*(1)(r - r') = z(r - r');$$

that is to say, $g(b) + zr = g(b') + zr'$, proving that f^* is well defined. ■

Recall that if A is a left R -module and G is an abelian group, then $\text{Hom}_{\mathbb{Z}}(A, G)$ is a right R -module, defining the scalar multiplication by

$$(f \cdot r)(a) = f(ra), \forall r \in R, a \in A \text{ and } f \in \text{Hom}_{\mathbb{Z}}(A, G).$$

(See **5.4.7**.)

We should, at this juncture, also recall Proposition **6.1.3**. It should be underscored that the two isomorphisms are natural. For convenience we restate Proposition **6.1.3**:

Proposition 8.1.5. *For any ring R with identity we have:*

- (a) $\text{Hom}({}_R R, \)$ is naturally equivalent to $1_{\mathbf{RMod}}$.
- (b) $(\) \otimes_R R$ is naturally equivalent to $1_{\mathbf{Mod}_R}$.

Now we can, at last, produce injective modules.

Lemma 8.1.6. *Injective Producing Lemma.* Suppose that D is any divisible abelian group. Then $\text{Hom}_{\mathbb{Z}}(R, D)$ is (right) R -injective.

Proof. We use the adjunction of tensor products and Hom , and the natural isomorphism

$$\text{Hom}_{\text{Mod}_R}(A, \text{Hom}_{\mathbb{Z}}(R, D)) \cong \text{Hom}_{\mathbb{Z}}(A \otimes_R R, D)$$

of abelian groups, valid for any right R -module A . (Recall Exercise 5.4.11.)

Suppose now that K is a right ideal of R , and let j denote the inclusion map, $j : K \rightarrow R_R$. By Proposition 8.1.5(b), $j \otimes_R R$ is necessarily one-to-one, from which it follows (as D is a divisible, i.e., injective, abelian group) that

$$\text{Hom}_{\mathbb{Z}}(j \otimes_R R, D) : \text{Hom}_{\mathbb{Z}}(R_R \otimes_R R, D) \rightarrow \text{Hom}_{\mathbb{Z}}(K \otimes_R R, D)$$

is surjective. Now invoking the natural isomorphism of the first paragraph, we get that the group homomorphism $\text{Hom}_{\text{Mod}_R}(j, \text{Hom}_{\mathbb{Z}}(R, D))$,

$$\text{Hom}_{\text{Mod}_R}(R_R, \text{Hom}_{\mathbb{Z}}(R, D)) \rightarrow \text{Hom}_{\text{Mod}_R}(K, \text{Hom}_{\mathbb{Z}}(R, D)),$$

is onto. Lemma 8.1.4 then guarantees that $\text{Hom}_{\mathbb{Z}}(R, D)$ is (right) R -injective. ■

Theorem 8.1.7. *Every R -module A can be embedded as a submodule of an injective R -module.*

Proof. Viewed as an abelian group, A can be embedded as a subgroup of a divisible abelian group D . According to Proposition 6.2.5(a), the inclusion $i : A \rightarrow D$ leads to a one-to-one R -homomorphism $\text{Hom}_{\mathbb{Z}}(R, i) : \text{Hom}_{\mathbb{Z}}(R, A) \rightarrow \text{Hom}_{\mathbb{Z}}(R, D)$. Now, $\text{Hom}_{\text{Mod}_R}(R, A)$ is a right R -submodule of $\text{Hom}_{\mathbb{Z}}(R, A)$, which, in turn, is R -isomorphic to a submodule of $\text{Hom}_{\mathbb{Z}}(R, D)$. Applying 8.1.5, $A \cong \text{Hom}_{\text{Mod}_R}(R, A)$, and then invoking the Injective Producing Lemma, we're done. ■

The second part of the section has to do with the construction of the injective hull of an R -module.

Definition 8.1.8. Suppose that A is an R -module, and a submodule of the injective module A^* . A^* is an *injective hull* of A , if for each one-to-one R -homomorphism $f : A \rightarrow J$, with J injective, there is an extension $f^* : A^* \rightarrow J$ which is also one-to-one. (Note: the existence of f^* is not news, since J is injective; what is novel here is that the extension of the homomorphism should be one-to-one.)

From this definition there is nothing to lead one to suspect that a module has an injective hull, nor that it is unique (up to isomorphisms, that is.) We shall prove that both are true.

The necessary companion definition, in this context, is that of an essential extension.

Definition 8.1.9. Suppose that A is a submodule of the R -module B . It is said that B is an *essential extension* of A (or, sometimes, that A is an *essential* or *large submodule* of B) if each nonzero submodule of B intersects A nontrivially.

Note that \mathbb{Q} is an essential extension of \mathbb{Z} (as abelian groups), and, for that matter, of any of its nonzero subgroups. For if $x = m/n$ is any nonzero rational number, then $nx = m \neq 0$, proving that any nonzero subgroup of \mathbb{Q} must contain an integer. A similar argument shows that \mathbb{Q} is an essential extension of any of its nonzero subgroups.

The following exercise is the “torsion” companion of the preceding example.

Exercise 8.1.10. Let \mathbb{Z}_p^∞ denote the group of all p^n -th complex roots of 1 (for a fixed prime number p , and all natural numbers n). Show that \mathbb{Z}_p^∞ is an essential extension of any of its nontrivial subgroups. (Note: this group is the direct limit of its subgroups; recall Exercise 4.2.7.)

The following criterion for essential extensions is quite easy to check.

Exercise 8.1.11. Suppose that A is a submodule of the R -module B . Then a necessary and sufficient condition that B be an essential extension of A is that each R -homomorphism $f : B \rightarrow C$, for which the restriction to A is one-to-one, itself be one-to-one.

The relationship between essential extensions of a module and the injectives in which it may be embedded, is what we must investigate in detail. The following proposition is the first step along the way.

Proposition 8.1.12. (a) Suppose that $A \leq B \leq C$; then C is an essential extension of A if and only if C is an essential extension of B , and B is an essential extension of A .

(b) Suppose that B is an essential extension of A , and that J is any injective R -module which contains A . Then J contains a submodule B' , which is R -isomorphic to B .

Proof. (a) We prove sufficiency, leaving the necessity to the reader.

If C is an essential extension of B , and B is an essential extension of A , then, for each nonzero submodule M of C , $M \cap B$ is nonzero; from this it follows that

$$M \cap A = (M \cap B) \cap A \neq 0.$$

This shows that C is an essential extension of A .

(b) Because J is injective, the inclusion $i : A \rightarrow J$ can be extended to a homomorphism $\theta : B \rightarrow J$, which, in view of the assumption that B is an essential extension of A , must be one-to-one; for if not $\ker(\theta) \neq 0$, while $A \cap \ker(\theta) = 0$. The image $\theta(B)$ is the promised copy of B . ■

Hot Air 8.1.13. *With a purpose.* The importance of (b) in Proposition 8.1.12 is twofold. First, since every essential extension of a module A lies embedded in any injective module that contains A as a submodule, there is a cardinal bound on the size of all the essential extensions of A ; namely, that of the least cardinal number of an injective module which majorizes A . Secondly, the collection of all isomorphism classes of essential extensions of A is a set! It makes sense to attempt to apply Zorn's Lemma, to obtain maximal essential extensions. We are about to do exactly that.

The next step, however, is a crucial one, which is also quite revealing. First, let us make a convenient definition.

Definition 8.1.14. An R -module A is said to be *essentially closed* if it cannot be embedded properly in a module B which is an essential extension of A .

Theorem 8.1.15. *The R -module A is essentially closed if and only if it is injective.*

Proof. Suppose first that A is injective, and that B is an essential extension of A . According to Proposition 6.3.13(b), $B = A \oplus C$, for a suitable submodule C of B . Since $A \cap C = 0$, this means that C itself is 0. Thus $A = B$, which shows that A is essentially closed.

Conversely, suppose that A is essentially closed. Let J be an injective R -module, of which A is a submodule. Apply a Zorn's Lemma argument to obtain a submodule B of J , which is maximal

with respect to the condition $A \cap B = 0$. If we are able to show that J/B is an essential extension of $(A + B)/B$, then we are done, because, by the Third Isomorphism Theorem, $A \cong (A + B)/B$; since A is essentially closed, it follows that $A \oplus B = J$, and so $J \cong A \times B$, whence, by Proposition 6.3.13, A is injective.

So it remains only to show that J/B is an essential extension of $(A + B)/B$. Suppose, by way of contradiction, that K/B (with $B \leq K$) is a nonzero submodule of J/B , which intersects $(A + B)/B$ trivially. Then $B = K \cap (A + B) = (K \cap A) + B$, by the modular law for the lattice of submodules, from which we get that

$$K \cap A \leq B \cap A = 0.$$

However, since $B < K$, this is a contradiction to the maximality of B .

Thus, J/B is an essential extension of $(A + B)/B$, and the proof is done. ■

We are now able to settle the issue of existence and uniqueness of the injective hull.

Theorem 8.1.16. *Every R -module has an injective hull. Moreover, if J_1 and J_2 are injective hulls of A , then there is an R -isomorphism from J_1 onto J_2 which is the identity when restricted to A .*

Indeed, for an R -module A , which is a submodule of B , the following are equivalent:

- (a) B is the injective hull of A .
- (b) B is injective, and an essential extension of A .
- (c) B is an essential extension of A , which is essentially closed.

Proof. Suppose that A is an R -module. By Theorem 8.1.7, we may suppose that A is a submodule of an injective R -module M . Consider now the set $\text{Ess}(A)$ of all essential extensions of A in M , partially ordered by set theoretic inclusion. The reader will easily verify that the union of a chain of essential extensions of A in M is an essential extension of A , which means that Zorn's Lemma may be applied to produce a maximal element B in $\text{Ess}(A)$. To show that B is injective, it suffices, according to the preceding theorem, to prove that B is essentially closed. So suppose that C is an essential extension of B . By Proposition 8.1.12(a), C is an essential extension of A . Since M is injective, the inclusion of B in M can be extended to a one-to-one R -homomorphism $\alpha : C \rightarrow M$. But then $\alpha(C)$ is a member of $\text{Ess}(A)$; by maximality, this forces $\alpha(C) = B$, which means that $C = B$. Thus, B is essentially closed, and hence injective.

If J is any injective R -module, and $f : A \rightarrow J$ is a one-to-one homomorphism, then since B is an essential extension of A , there is a homomorphism $f^* : B \rightarrow J$ which is one-to-one, and extends f . By definition, this makes B an injective hull of A .

Now, suppose that J is any injective hull of A . Keeping to the notation of the previous paragraphs, there is a one-to-one homomorphism $g : J \rightarrow B$, which has the property that $g(a) = a$, for each $a \in A$. This says that $g(J)$ is a submodule of B containing A . But then $g(J)$ (being isomorphic to J) is injective, and, by Proposition 8.1.12(a), B is an essential extension of $g(J)$. Since injectives are essentially closed, according to Theorem 8.1.15, it follows that $g(J) = B$, and this proves that between any two injective hulls of A there is an isomorphism that restricts to 1_A .

As to the equivalence of (a), (b) and (c), a close reading of the preceding paragraph reveals that it has been shown that any injective hull is isomorphic to one which is an essential extension, which shows that (a) implies (b). The equivalence of (b) and (c) is Theorem 8.1.15, and the first two paragraphs of this proof show that (b) implies (a). ■

Remark 8.1.17. Let's have a look back at the examples in 8.1.9 and Exercise 8.1.10.

Since \mathbb{Q} is a divisible abelian group, and an essential extension of any of its nonzero subgroups, it follows from the previous theorem that \mathbb{Q} is the injective hull of any of its nonzero subgroups.

Likewise, for any prime number p , \mathbb{Z}_p^∞ is divisible, and an essential extension of any of its nonzero subgroups; thus, it too is the injective hull of any of its nonzero subgroups.

What makes the above tick is the subject for the next exercise. First, a definition.

Definition 8.1.18. An R -module A is said to be *indecomposable* if the only summands of A are 0 and A itself. Note that both \mathbb{Q} and \mathbb{Z}_p^∞ (for any prime number p) are indecomposable.

Incidentally, since \mathbb{Q} is indecomposable, it follows that it cannot be a free abelian group. Ditto for \mathbb{Z}_p^∞ , for any prime p .

Then we have the following result.

Exercise 8.1.19. For an R -module A the following are equivalent.

- (1) The injective hull A^* of A is indecomposable.
- (2) If B and C are any nonzero submodules of A , then $B \cap C \neq 0$.
- (3) If B and C are any nonzero submodules of A^* , then $B \cap C \neq 0$.
- (4) A^* is the injective hull of any of its nonzero submodules.

The next exercise concerns abelian groups. We will revisit it and related phenomena when we discuss rings of quotients.

Exercise 8.1.20. If G is a torsion free abelian group, then $G \otimes \mathbb{Q}$ is the injective (divisible) hull of G . Give an example to show that this is false if G is not torsion free.

We also have the following, anticlimactic corollary of Theorem 8.1.7; it is the converse of Proposition 6.3.13(b).

Exercise 8.1.21. If the module M is a summand of every R -module in which it is a submodule, then M is injective. (Hint: Embed M in an injective module, and then invoke Proposition 6.3.13(a).)

We conclude the section with the following exercise, which is also outlined in [Hu74], p. 198, Exercise 11. We refer the reader to the hints given there, or else the one below.

Exercise 8.1.22. Prove that a divisible abelian group can be expressed as a direct sum of copies of \mathbb{Q} and \mathbb{Z}_p^∞ 's, for various prime numbers p .

(Hint: The idea is to break up the group into a torsion and a torsion free component. Yes, the torsion part of a divisible group is divisible. Show that the torsion free part is a rational vector space; that should clear up the part involving \mathbb{Q} 's. The torsion subgroup is a direct sum of its primary components. Each one is divisible (Why?). Then see [Hu74], p. 198, Exercise 10.)

8.2 Flat Modules. As in previous sections, R is a ring with identity. Recall Proposition 6.2.7: suppose that the following sequence of left R -modules

$$(E) \quad 0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is short exact. Then for each right R -module M , the following sequence is exact:

$$(M \otimes (E)) \quad M \otimes A \xrightarrow{M \otimes f} M \otimes B \xrightarrow{M \otimes g} M \otimes C \longrightarrow 0.$$

As was shown in Example 6.2.8, exactness at $M \otimes A$ fails, in general.

Definition 8.2.1. A right R -module M is said to be (*right flat*) if for each short exact sequence (E) , as above, the sequence $(M \otimes (E))$ is short exact; (meaning, $M \otimes f$ is one-to-one). *Left flat* modules are defined analogously.

Note that it is a consequence of Proposition 8.1.5(b) that R_R is (right) flat. From this observation, and the fact that the functor $(\) \otimes B$ preserves direct sums (that is to say, coproducts), for each left R -module B , we will, shortly, obtain that every free module is flat.

First, let us record the following.

Proposition 8.2.2. *Let $M = \oplus_i M_i$ ($i \in I$) be a direct sum of right R -modules. Then M is flat if and only if each M_i is flat. An identical result holds for left R -modules.*

Proof. (Sufficiency) Recall Theorem 6.1.1(d): for any left R -module A ,

$$M \otimes A = \bigoplus_{i \in I} M_i \otimes A.$$

Indeed, this isomorphism is natural; let's be precise about this: let $d_{i,A} : M_i \otimes A \longrightarrow \oplus_i M_i \otimes A$ denote the i -th coprojection. Now, consider the functor T_M defined by setting $T_M(A) = \oplus_i M_i \otimes A$, and, for each left R -homomorphism $g : A \longrightarrow B$, $T_M(g) : T_M(A) \longrightarrow T_M(B)$ is the unique group homomorphism such that

$$T_M(g) \cdot d_{i,A} = d_{i,B} \cdot (M_i \otimes g).$$

Refer to the diagram which follows for the discussion below.

$$\begin{array}{ccc} M_i \otimes A & \xrightarrow{d_{i,A}} & T_M(A) \\ \downarrow M_i \otimes g & & \downarrow T_M(g) \\ M_i \otimes B & \xrightarrow{d_{i,B}} & T_M(B) \end{array}$$

The full force of Proposition 6.1.1(d) is that the functors $M \otimes (\)$ and T_M are naturally equivalent. We identify the natural equivalence, and otherwise leave the details to the reader. The isomorphism from $\oplus_i M_i \otimes A \longrightarrow M \otimes A$ is simply the unique group homomorphism t_A for which

$$t_A \cdot d_{i,A} = \delta_i \otimes A,$$

where $\delta_i : M_i \rightarrow M$ is the i -th coprojection. Now $t : T_M \rightarrow M \otimes ()$ is a natural equivalence.

It follows that, to prove that $M \otimes ()$ sends one-to-one maps to one-to-one maps, is equivalent to proving it for the functor T_M . But this uses a property of direct sums: if $g : A \rightarrow B$ is a left R -homomorphism which is one-to-one, then, by assumption, each $M_i \otimes g$ is one-to-one. From this it easily follows that $T_M(g)$ is one-to-one as well, because in a direct sum of groups G_i an element $g = \sum_i g_i = 0$, if and only if each $g_i = 0$.

(Necessity) What is true is this: if N is a summand of M , and M is flat, then so is N . The proof involves an argument upon a diagram like the one above. We leave the details to the reader. ■

An immediate corollary of Propositions 8.2.2, and 8.1.5(b) is this:

Proposition 8.2.3. *Every projective module is flat.*

(Note: the converse is false; as we shall see, every torsion free abelian group is flat.)

Proof. Every free module is a direct sum of copies of R , which, by 8.1.5(b), is flat. Now apply Proposition 8.2.2 and the fact that every projective module is a summand of a free one. ■

Now let's proceed to describe the duality between right injective modules and left flat modules.

Definition & Remarks 8.2.4. Suppose that M is a left R -module. We let $M^* = \text{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$, viewed as a right R -module, in the usual way. M^* is sometimes referred to as the *character module* of M ; we shall simply refer to it as the *dual* of M . Since the word "dual" has been used rather freely in these pages, this is patently risky. We shall, boldly, take the risk.

The dual of M has a curious, technical property. One could reasonably say that M^* *separates the elements of M* .

Lemma 8.2.5. *If $m \in M$ is nonzero, then there exists an $f \in M^*$ such that $f(m) \neq 0$.*

Proof. The proof hinges on the divisibility of \mathbb{Q}/\mathbb{Z} . To produce a homomorphism $f \in M^*$ for which $f(m) \neq 0$, it suffices to find one defined on $\mathbb{Z}m$, the cyclic subgroup generated by m . Then one uses the injectivity of \mathbb{Q}/\mathbb{Z} (over \mathbb{Z}) to extend it to M .

Now there are two cases to consider.

- I. *The order of m is infinite.* Then pick any natural number $n > 1$, and define $f(km) = \mathbb{Z} + k/n$. It is trivial, to show that f is a homomorphism.
- II. *The order of m is $t > 0$.* Define $f(km) = \mathbb{Z} + k/t$. One easily proves that f is well defined, and a homomorphism.

In either case f does the job. ■

With this lemma one can show that dualization both preserves and reflects short exactness. We leave it as an exercise for the reader.

Exercise 8.2.6. Suppose that $f : A \rightarrow B$ and $g : B \rightarrow C$ are homomorphisms of groups. Prove that

$$(E) \quad 0 \rightarrow A \xrightarrow{f} B \xrightarrow{g} C \rightarrow 0$$

is short exact if and only if

$$(E^*) \quad 0 \rightarrow C^* \xrightarrow{g^*} B^* \xrightarrow{f^*} A^* \rightarrow 0$$

is short exact. (Hint: Necessity follows from the divisibility of \mathbb{Q}/\mathbb{Z} . For the sufficiency use Lemma 8.2.5.)

Using the preceding lemma and exercise, one obtains:

Theorem 8.2.7. *The left R -module M is flat if and only if its dual M^* is a right injective module.*

Proof. We will use the natural group isomorphism

$$\text{Hom}_{\mathbb{Z}}(A \otimes M, \mathbb{Q}/\mathbb{Z}) \cong \text{Hom}_{\text{Mod}_R}(A, M^*).$$

If M is flat, and $\alpha : A \rightarrow B$ is an injective homomorphism, then $\alpha \otimes M : A \otimes M \rightarrow B \otimes M$ is also one-to-one; since \mathbb{Q}/\mathbb{Z} is divisible (injective), it follows that $\text{Hom}_{\mathbb{Z}}(\alpha \otimes M, \mathbb{Q}/\mathbb{Z})$ is onto. By the natural isomorphism recalled above, this means that $\text{Hom}_{\text{Mod}_R}(\alpha, M^*)$ is onto. By Proposition 6.3.2(b), it follows that M^* is injective.

Conversely, suppose that M^* is right injective, and $\alpha : A \rightarrow B$ is one-to-one. We reverse the steps in the preceding paragraph: $\text{Hom}_{\text{Mod}_R}(\alpha, M^*)$ is onto, and, therefore, so is the naturally equivalent $\text{Hom}_{\mathbb{Z}}(\alpha \otimes M, \mathbb{Q}/\mathbb{Z})$. By the previous exercise, $\alpha \otimes M$ is one-to-one, proving that M is flat. ■

Remark 8.2.8. *Look Back.* The reader should be aware that Proposition 8.2.2 can be derived from Theorem 8.2.7, using the fact that

$$\left(\bigoplus_{i \in I} M_i\right)^* = \prod_{i \in I} M_i^*,$$

which comes out of Theorem 6.1.1(c).

Hot Air 8.2.9. Suppose that M is a left R -module, and K is a right ideal of R . Let

$$KM \equiv \{r_1 m_1 + \cdots + r_k m_k : \forall m_i \in M, r_i \in K\}.$$

KM is a subgroup of M .

As the map $(r, m) \rightarrow rm$, from $K \times M$ into KM is clearly bilinear, there is a unique group homomorphism $\phi : K \otimes M \rightarrow KM$, for which $\phi(r \otimes m) = rm$. Since the tensors of the form $r \otimes m$ generate $K \otimes M$, it should be clear that ϕ is surjective.

Now, look at ϕ another way. The natural inclusion $j : K \rightarrow R$ induces $j \otimes M : K \otimes M \rightarrow R \otimes M$. Then we have the natural isomorphism (guaranteed by the dual of 8.1.5) $\tau_M : R \otimes M \rightarrow M$ (which satisfies $\tau_M(r \otimes m) = rm$, for each $m \in M$, and each $r \in R$). Observe now that $\phi = \tau_M \cdot (j \otimes M)$.

We are now set up to prove the following technical, but rather useful result. For convenience, let us refer to the map ϕ introduced two paragraphs ago as the *internal tensor map* associated with the right ideal K and the left R -module M .

Next, here is a technical result, which gets to the internal conditions that make a module flat.

Proposition 8.2.10. *If M is a left R -module, then the following are equivalent.*

- (a) M is a flat (left) R -module.
- (b) For each right ideal K of R , the internal tensor map associated with K and M is an isomorphism;
- (c) For each finitely generated right ideal K of R , the internal tensor map associated with K and M is an isomorphism.

Proof. It is clear that (b) implies (c).

(a) \Rightarrow (b): If M is flat, then

$$j \otimes M : K \otimes M \longrightarrow R \otimes M$$

is one-to-one, and since $\phi = \tau_M \cdot (j \otimes M)$, it follows that ϕ too is one-to-one. It has already been observed that it is onto, and therefore an isomorphism.

(c) \Rightarrow (b): Suppose that K is a right ideal. It suffices to show that if $r_1 m_1 + \cdots + r_k m_k = 0$ (with the scalars $r_i \in K$, and $m_i \in M$) then

$$(\#) \quad r_1 \otimes m_1 + \cdots + r_k \otimes m_k = 0,$$

computed in $K \otimes M$. However, if K_0 is the right ideal generated by r_1, r_2, \dots, r_k , then $(\#)$ holds, computed in $K_0 \otimes M$, by assumption. But we have the homomorphism $\theta : K_0 \otimes M \longrightarrow K \otimes M$ induced by the inclusion of K_0 in K . Apply θ to $(\#)$, calculated in $K_0 \otimes M$, to get $(\#)$, computed in $K \otimes M$.

(b) \Rightarrow (a): Invoking the identity

$$\phi = \tau_M \cdot (j \otimes M)$$

yet again, one sees that $j \otimes M$ is one-to-one, if ϕ is an isomorphism. Thus, $\text{Hom}_{\mathbb{Z}}(j \otimes M, \mathbb{Q}/\mathbb{Z})$ is onto (by the divisibility of \mathbb{Q}/\mathbb{Z}), and using the natural equivalence of functors one more time, we conclude that $\text{Hom}_{\text{Mod}_R}(j, M^*)$ is onto.

What this means is that, for each right ideal K of R , and any R -homomorphism $f : K \longrightarrow M^*$, there is an extension of f to R . According to Lemma 8.1.4, this suffices to conclude that M^* is injective. Then, by Theorem 8.2.7, M is flat. \blacksquare

Next, a characterization of the flat modules over certain integral domains; namely, the ones for which every finitely generated ideal is principal.

Definition 8.2.11. If A is an integral domain, and M is an A -module, we say that M is *torsion free* if $am = 0$ ($a \in A$, $m \in M$) implies that either $a = 0$ or $m = 0$.

The next exercise applies, particularly, to all PIDs. The integral domains having the property that every finitely generated ideal is principal are called *Bézout domains*.

Exercise 8.2.12. Suppose that A is a Bézout domain. Prove that a module M is flat if and only if it is torsion free.

(Hint: (Sufficiency) If M is torsion free then (c) in Proposition 8.2.10 holds, for every principal ideal K , hence, for every finitely generated ideal. By Proposition 8.2.10, M is flat.

(Necessity) If M is not torsion free, pick $a \in A$ and $m \in M$, both nonzero, such that $am = 0$. Now, the homomorphism $r \longrightarrow ra$, from $A \longrightarrow Aa$, is an isomorphism, since A is an integral domain. This means that the homomorphism induced by $r \otimes m \mapsto ra \otimes m$, from $A \otimes M \longrightarrow Aa \otimes M$ is also an isomorphism. Precede this with τ_M^{-1} (see the notation of 8.2.9) to obtain that the assignment $m \mapsto a \otimes m$, from $M \longrightarrow Aa \otimes M$ is an isomorphism. Complete the argument, by composing with ϕ in 8.2.10, with $K = Aa$.)

The major goal that remains in this discussion is to characterize the rings, over which every module is flat. Later, in the context of commutative rings, we shall have more to say about them.

Definition & Remarks 8.2.13. A ring R with identity is said to be *von Neumann regular* if for each $r \in R$ there is an $r' \in R$ so that $rr'r = r$.

The prototypes of von Neumann regular rings are (in the commutative case): all fields; all direct products of fields, and, in particular, taking the field to be \mathbb{Z}_2 , all boolean rings. In the noncommutative case, the ring $\text{Mat}_n(K)$, for any field K , is von Neumann regular.

To see the latter assertion, we should make it clear that more is true: if V is any vector space over the division ring D , then $\text{End}_D(V)$ is von Neumann regular. Here's a proof. Suppose that f is a linear transformation of V into itself. Choose a basis for V as follows: let \mathcal{B}_0 be any basis for $f(V)$, and extend this to a basis \mathcal{B} of V . Now define $g \in \text{End}_D(V)$ as follows: for $v \in \mathcal{B}_0$ let $g(v)$ be any preimage of v ; if $v \in \mathcal{B} \setminus \mathcal{B}_0$, let $g(v) = 0$. Extend to V by linearity, and then observe that $f \cdot g \cdot f = f$, by construction.

Here are some facts about von Neumann regular rings, to help the reader gain some intuition about them.

Exercise 8.2.14. Suppose that R is a ring with identity.

- (a) If R is von Neumann regular, show that it is J -semisimple.
- (b) Prove that R is von Neumann regular if and only if every principal left (resp. right) ideal is generated by an idempotent.
- (c) Any direct product of von Neumann regular rings is von Neumann regular.
- (d) If R is commutative and von Neumann regular, then it is a subdirect product of fields.
- (e) Any J -semisimple left Artinian ring is von Neumann regular. (Use the Wedderburn–Artin Theorem.)
- (f) If R is von Neumann regular, then every finitely generated left (resp. right) ideal is principal. (Do it for two generators and then scream induction! By (b), these generators are, without loss of generality, idempotent. The trick is to choose the correct idempotents, meaning, two that mutually annihilate one another.)

Now we have the following. Because of this theorem, von Neumann regular rings are often – especially in the French literature – referred to as *absolutely flat rings*.

Theorem 8.2.15. R is von Neumann regular if and only if every left (resp. right) R -module is flat.

Proof. Suppose that R is von Neumann regular. Then, by (f) in the preceding exercise, every finitely generated right ideal K is principal, and, indeed, $K = eR$, for some idempotent e . This means that $R = K \oplus (1 - e)R$. Thus, referring again to the notation of 8.2.9 and Proposition 8.2.10, the map $j \otimes M$ is one-to-one, for each left R -module M , because tensor products preserve direct sums. It follows that the internal tensor map associated with K and M is an isomorphism, proving (c) in 8.2.10. This says that M is flat.

Suppose that every left R -module is flat. Consider the diagram below, which is a commutative square, for each $a \in R$.

$$\begin{array}{ccc}
 aR & \xrightarrow{\beta} & aR \otimes R/Ra \\
 \downarrow \alpha & & \downarrow \theta' \\
 R & \xrightarrow{\theta} & R/Ra
 \end{array}$$

where α is the inclusion, θ the canonical homomorphism, and

$$\beta(ar) = ar \otimes (Ra + 1), \quad \text{and} \quad \theta'(ar \otimes (Ra + y)) = Ra + ary.$$

(Verify that

$$\theta \cdot \alpha = \theta' \cdot \beta = \phi,$$

the internal tensor map associated with aR and R/Ra .) Now, the kernel of $\theta \cdot \alpha$ is, obviously, $aR \cap Ra$, whereas $\ker(\theta' \cdot \beta) = \ker(\beta)$, because θ' is one-to-one, in view of the flatness of Ra .

However, appealing to Proposition 6.2.7, $\ker(\beta)$ is the image of the map $f : aR \otimes Ra \rightarrow aR$, defined by $f(ar \otimes sa) = arsa$. Since $aR \cap Ra = aRa$, and $a \in aR \cap Ra$, it follows that $a = ara$, for a suitable choice of $r \in R$. ■

The following exercises harken back to group-algebras and Maschke's Theorem (Exercise 7.4.23). The end result is a converse to that theorem.

If R is any ring with identity, and X is a subset of R , let $\text{Ann}_l(X)$ (resp. $\text{Ann}_r(X)$) denote the left (resp. right) annihilator of X .

Exercise 8.2.16. Suppose that G is a group and that K is a field. If H is any subgroup of G , let $j(H)$ denote the left ideal of $K[G]$ generated by all elements of the form $1 - h$, for all $h \in H$. Observe that $j(H)$ is always a *proper* left ideal of $K[G]$.

Prove the following:

- The mapping $H \mapsto j(H)$ is a one-to-one function which preserves lattice suprema.
- If H is normal, then $j(H)$ is a two-sided ideal.
- If H is an infinite subgroup of G , then $\text{Ann}_r(j(H)) = 0$.
- If H is a finite subgroup of G , then $\text{Ann}_r(j(H))$ is the right ideal generated by $\sum_{h \in H} h$.
- If $K[G]$ is Noetherian then G satisfies the ascending chain condition on its subgroups. (And this means that all its subgroups are finitely generated. See Exercise 1.3.7.)
- If $g \in G$ has finite order n , then $\text{Ann}_r(1 - g)$ is the right ideal generated by $1 + g + \cdots + g^{n-1}$.

Exercise 8.2.17. Prove that $K[G]$ is von Neumann regular if and only if every finitely generated subgroup of G is finite and the characteristic of K does not divide the order of any of the finite subgroups of G .

(Hint: (Necessity) Recall that in von Neumann regular rings each finitely generated left ideal is a summand. Now apply Exercise 8.2.16(a) to obtain that if H is a finitely generated subgroup of

G , then $\text{Ann}_r(j(H))$ is a nontrivial summand, and then use **8.2.16(d)** to conclude that H must be finite.

To prove that the characteristic of K does not divide the order of any finite subgroup of G , proceed as follows. If $g \in G$ of order n , use von Neumann regularity to produce an $r \in K[G]$ such that $(1-g)r(1-g) = 1-g$. Next, play with the latter identity, and apply **8.2.16(f)** to get an $s \in K[G]$ so that

$$(\dagger) \quad 1 - r(1 - g) = (1 + g + \cdots + g^{n-1})s.$$

Finally, let π be the map $K[G] \rightarrow K$, defined by

$$\pi\left(\sum_g r_g g\right) = \sum_g r_g.$$

This is a ring homomorphism; apply it to (\dagger) to conclude that the constant $n = n \cdot 1$ is nonzero in the field K .

(Sufficiency) Use Maschke's Theorem, applied as follows: for each $r \in K[G]$, let H be the subgroup generated by the elements of G bearing nonzero scalar coefficients, in the expression $r = \sum_g c_g g$. By assumption, H is finite and $\text{char}(K)$ does not divide $|H|$. So now apply Maschke's Theorem to $K[H]$, and then the fact that matrix rings are von Neumann regular.)

Exercise 8.2.18. Deduce from Exercise **8.2.17**, that if $K[G]$ is J -semisimple and left Artinian then G must be finite. (You'll also need **8.2.16(e)** and the Wedderburn–Artin Theorem.)

Remark 8.2.19. The result of Exercise **8.2.17** can be generalized. The reader should refer to the second appendix of [L86]. One finds (p. 155, Proposition 2, [L86]) the following for a group ring $R[G]$, where R is a ring with identity.

$R[G]$ is von Neumann regular if and only if

- (a) *R is von Neumann regular.*
- (b) *Every finitely generated subgroup of G is finite.*
- (c) *If H is a finite subgroup of G , and $n = |H|$, then $n \cdot 1$ is a multiplicative unit in R .*

We close with the following, which is quite general, and can be proved by imitating the proof of Proposition **8.2.2**.

Exercise 8.2.20. Prove that a direct limit of flat modules, with bonding maps which are one-to-one, is flat. (Hint: Use the isomorphism of Theorem **6.1.1(e)**.)

9. Commutative Algebra: Ideals and Spectra

This chapter initiates the study of commutative rings. The objective is to identify the principal tools that are used to analyze the structure of commutative rings. We review some material from earlier chapters, and introduce the prime spectrum. We go so far as to prove the Stone Duality. The main reference for this chapter will be [AM69].

All rings encountered in this chapter are commutative and possess an identity. Unless otherwise specified, categorical references should be applied to **CRn1**, the category of all commutative rings with identity, and all ring homomorphisms which preserve the identity.

9.1 Introduction. We begin with a brief review of some of the material already encountered in our discussions about rings, applying them to commutative rings. At the same time, the results being recalled will provide points of departure for new ideas.

We begin by summing up Exercise 2.4.4(b), on prime ideals.

Proposition 9.1.1. *Suppose that A is a ring. Then the set $n(A)$, consisting of all nilpotent elements, is the intersection of all prime ideals of A .*

This proposition prompts a definition. We shall use oldfashioned notation.

Definition 9.1.2. Suppose that \mathfrak{r} is an ideal of the ring A . Let $\sqrt{\mathfrak{r}}$ stand for the set

$$\{x \in A : \exists n \in \mathbb{N}, \text{ such that } x^n \in \mathfrak{r}\}.$$

We shall call $\sqrt{\mathfrak{r}}$ the *radical* of \mathfrak{r} . Evidently, $n(A) = \sqrt{0}$.

We first record the following extension of Proposition 9.1.1. Its proof depends on Zorn's Lemma, and is left to the reader.

Proposition 9.1.3. *Suppose that \mathfrak{r} is an ideal of the ring A . Then $\sqrt{\mathfrak{r}}$ is the intersection of all the prime ideals of A that contain \mathfrak{r} .*

On the heels of Proposition 9.1.3, now is as good a time as any to make the observation which follows. We record it as an exercise, with a definition.

Exercise 9.1.4. Suppose that \mathfrak{r} is an ideal of the ring A . Then $\sqrt{\mathfrak{r}} = \mathfrak{r}$ if and only if $a^2 \in \mathfrak{r}$ implies that $a \in \mathfrak{r}$.

An ideal satisfying this condition is called a *semiprime* ideal. When $\{0\}$ is semiprime (that is, when $n(A) = 0$) the ring A is also said to be *semiprime*. (Some texts still use the term *reduced ring* in place of semiprime ring.)

We also recall the Jacobson radical in its incarnation for commutative rings.

Definition 9.1.5. Let A be a ring. Recall that

$$\mathfrak{J}(A) = \{a \in A : 1 + ab \text{ is invertible, } \forall b \in A\},$$

which is the same as the intersection of all the maximal ideals of A .

To recall Exercise 7.3.12, the Jacobson radical contains $n(A)$, the *nil* or *prime* radical, but they need not agree. Here is an important example, which will rear its head again.

Exercise 9.1.6. Suppose that A is a ring. $A[[T]]$ denotes the *ring of formal power series*, $a_0 + a_1T + \cdots + a_nT^n + \cdots$, with the $a_n \in A$, and where the addition is termwise, while the product is “convolution”; that is,

$$\left(\sum_{n=0}^{\infty} a_nT^n\right)\left(\sum_{n=0}^{\infty} b_nT^n\right) = \sum_{n=0}^{\infty} c_nT^n,$$

so that

$$c_n = a_0b_n + a_1b_{n-1} + \cdots + a_nb_0,$$

for each $n \in \mathbb{N}$.

Then prove (see Exercise 5, Chapter 1, [AM69]):

- (a) $f = a_0 + a_1T + \cdots + a_nT^n + \cdots$ is invertible in $A[[T]]$ if and only if a_0 is invertible in A .
- (b) f (as in (a)) belongs to $\mathfrak{J}(A[[T]])$ precisely when $a_0 \in \mathfrak{J}(A)$. (In particular, if F is a field, then $\mathfrak{J}(F[[T]])$ consists of the nonunits of $F[[T]]$. Recall: a *unit* is simply an invertible element of the ring.)
- (c) If f is nilpotent, then so are all the a_n . (Is the converse true?) Thus, if A is semiprime, then so is $A[[T]]$.

Remark 9.1.7. *By Contrast.* (See [AM69], Ch. 1, Exercises. 2 & 4) For the polynomial ring $A[T]$ (for any ring A), the prime and Jacobson radical agree.

Here is a laundry list of properties of radicals of ideals. The beginproof is left to the reader.

Proposition 9.1.8. *Suppose that \mathfrak{r} and \mathfrak{s} are ideals of the ring A . Then*

- (a) $\sqrt{\mathfrak{r}} \geq \mathfrak{r}$ and $\sqrt{(\sqrt{\mathfrak{r}})} = \sqrt{\mathfrak{r}}$.
- (b) $\mathfrak{r} \leq \mathfrak{s}$ implies $\sqrt{\mathfrak{r}} \leq \sqrt{\mathfrak{s}}$.
- (c) $\sqrt{\mathfrak{r}\mathfrak{s}} = \sqrt{\mathfrak{r} \cap \mathfrak{s}} = \sqrt{\mathfrak{r}} \cap \sqrt{\mathfrak{s}}$.
- (d) $\sqrt{\mathfrak{r}} = A$ if and only if $\mathfrak{r} = A$.
- (e) If \mathfrak{r} is a prime ideal, then $\sqrt{\mathfrak{r}^n} = \mathfrak{r}$, for each $n \in \mathbb{N}$.
- (f) In general, $\sqrt{\mathfrak{r} + \mathfrak{s}} \geq \sqrt{\mathfrak{r}} + \sqrt{\mathfrak{s}}$, but equality need not hold.

Definition & Remarks 9.1.9. Suppose that $f : A \rightarrow B$ is a ring homomorphism which preserves the identity; (that is to say, a morphism in **Crn1**.) If \mathfrak{r} is an ideal of A , we denote by \mathfrak{r}^e the ideal of B generated by the image of \mathfrak{r} ; thus, $\mathfrak{r}^e = Bf(\mathfrak{r})$. (Note: $f(\mathfrak{r})$ itself need not be an ideal of B !) \mathfrak{r}^e is called the *extension of \mathfrak{r} in B* . An *extended ideal* of B is one of the form \mathfrak{r}^e , for some ideal \mathfrak{r} of A .

If \mathfrak{s} is an ideal of B , then $\mathfrak{s}^c = f^{-1}(\mathfrak{s})$ is an ideal of A , called the *contraction of \mathfrak{s} to A* . An ideal \mathfrak{r} of A is called a *contracted ideal* if $\mathfrak{r} = \mathfrak{s}^c$, for a suitable ideal \mathfrak{s} of B .

These terms are motivated by the situation in which A is a subring of B , and f is the inclusion. Then $\mathfrak{s}^c = \mathfrak{s} \cap A$, and \mathfrak{r}^e is just the ideal of B generated by \mathfrak{r} , $\mathfrak{r}^e = B\mathfrak{r}$.

More laundry lists! Some proofs can be found in [AM69], p. 10. In general, none of the inequalities in Proposition 9.1.10 can be improved to identities.

Proposition 9.1.10. *Suppose that \mathfrak{r}_1 and \mathfrak{r}_2 are ideals of the ring A and \mathfrak{s}_1 and \mathfrak{s}_2 are ideals of the ring B , while $f : A \rightarrow B$ is a ring homomorphism. Then*

- (i) $(\mathfrak{r}_1 + \mathfrak{r}_2)^e = \mathfrak{r}_1^e + \mathfrak{r}_2^e$, while $(\mathfrak{s}_1 + \mathfrak{s}_2)^c \geq \mathfrak{s}_1^c + \mathfrak{s}_2^c$;
- (ii) $(\mathfrak{r}_1 \cap \mathfrak{r}_2)^e \leq \mathfrak{r}_1^e \cap \mathfrak{r}_2^e$, while $(\mathfrak{s}_1 \cap \mathfrak{s}_2)^c = \mathfrak{s}_1^c \cap \mathfrak{s}_2^c$;
- (iii) $(\mathfrak{r}_1 \mathfrak{r}_2)^e = \mathfrak{r}_1^e \mathfrak{r}_2^e$, while $(\mathfrak{s}_1 \mathfrak{s}_2)^c \geq \mathfrak{s}_1^c \mathfrak{s}_2^c$;
- (iv) $\mathfrak{r} \leq \mathfrak{r}^{ec}$, for each ideal \mathfrak{r} of A , and $\mathfrak{s} \geq \mathfrak{s}^{ce}$, for each ideal \mathfrak{s} of B .
- (v) With \mathfrak{r} and \mathfrak{s} as in (iv), $\mathfrak{r}^e = \mathfrak{r}^{ece}$, and $\mathfrak{s}^c = \mathfrak{s}^{cec}$.
- (vi) $(\sqrt{\mathfrak{r}})^e \leq \sqrt{\mathfrak{r}^e}$, for each ideal \mathfrak{r} of A , while $(\sqrt{\mathfrak{s}})^c = \sqrt{\mathfrak{s}^c}$, for each ideal \mathfrak{s} of B .

Finally, $\mathfrak{r} \leq A$ is a contracted ideal of A if and only if $\mathfrak{r} = \mathfrak{r}^{ec}$, and $\mathfrak{s} \leq B$ is an extended ideal if and only if $\mathfrak{s} = \mathfrak{s}^{ce}$.

Finally, in this section, let us recall what was developed in Chapter 8 for von Neumann regular rings, and expand on that knowledge in the land of commutative rings.

Theorem 9.1.11. *Suppose that A is a ring. Then the following are equivalent.*

- (a) A is von Neumann regular; (meaning that, for each $a \in A$, there is an $x \in A$, such that $a^2x = a$.)
- (b) Every principal ideal of A is generated by an idempotent.
- (c) Every finitely generated ideal of A is a direct summand.
- (d) A is semiprime, and every prime ideal is maximal.
- (e) Every ideal is an intersection of maximal ideals.
- (f) Every ideal is semiprime.
- (g) Every A -module is flat.

Proof. In Exercise 8.2.14(b) the equivalence of (a) and (b) is stated. That (a) and (g) are equivalent is Theorem 8.2.15. Exercise 8.2.14 also gives that (a) implies (c), while the converse goes like this: if every finitely generated ideal of A is a summand, then so is every principal one; thus, for each $a \in A$, $1 = e + f$, where e and f are idempotent and $e \in Aa$, while $fa = 0$. It is then easy to check that $Aa = Ae$, proving (b).

It should be clear that a von Neumann regular ring has no nonzero nilpotent elements. If $\mathfrak{q} < \mathfrak{p}$ are prime ideals of A , then (without loss of generality) there is an idempotent $e \in \mathfrak{p} \setminus \mathfrak{q}$. But then $1 - e \in \mathfrak{q} \leq \mathfrak{p}$, which implies that $1 \in \mathfrak{p}$. This is absurd. We've shown that (a) implies (d), and since every homomorphic image of a von Neumann regular ring is von Neumann regular, we've actually shown that (a) implies (e).

Trivially, (f) follows from (e), and if every ideal is semiprime then, for each $a \in A$,

$$Aa = \sqrt{Aa} = \sqrt{Aa^2} = Aa^2,$$

meaning that $a^2x = a$ for a suitable $x \in A$, which shows that (f) implies (a).

This shows all, but that (d) implies the other statements. We leave this to the reader. Lemma 9.2.19 will do the trick nicely. ■

To close, three items; the first, again a repeated remark, which is an obvious consequence of all the above; the second, an exercise on boolean rings.

Corollary 9.1.12. (a) Any direct product of von Neumann regular rings is von Neumann regular.

(b) If A is von Neumann regular then $\mathfrak{J}(A) = n(A) = 0$.

Exercise 9.1.13. Suppose that A is a ring; then the following are equivalent.

(a) A is a boolean ring; (meaning that every element is idempotent.)

(b) A is semiprime, and for each prime ideal \mathfrak{p} , A/\mathfrak{p} is the field of two elements.

Finally, an important technical result, quite instrumental in some developments of commutative algebra. We won't use it until §12.4.

Exercise 9.1.14. *Nakayama's Lemma.* Suppose that A is a commutative ring with identity. Assume that M is a finitely generated A -module, \mathfrak{r} an ideal of A contained in $\mathfrak{J}(A)$, and that $\mathfrak{r}M = M$. Then $M = \{0\}$.

(Hint: Assume that $M \neq \{0\}$, and argue by contradiction, starting with a minimal generating set $\{x_1, \dots, x_k\}$. Since M is assumed to be nontrivial, we may take $x_1 \neq 0$. Now establish that if $r \in \mathfrak{r}$ then $1 - r$ must be a unit. Next, express x_1 as a combination

$$x_1 = \sum_{i=1}^k a_i x_i, \quad \text{each } a_i \in \mathfrak{r}.$$

Proceed to isolated multiples of x_1 on one side of the equation, and consider cases: either $k = 1$ or else it isn't.)

9.2 The Prime Spectrum. Let A be a ring. Denote by $\text{Spec}(A)$ the set of all prime ideals of A . In this section we will topologize $\text{Spec}(A)$, as well as a number of its subsets, in order to introduce, in a systematic manner, the interplay between commutative rings and topological spaces.

The reader is reminded that all rings in this chapter are commutative and possess an identity.

Definition & Remarks 9.2.1. *The Hull-kernel Topology.* The point here is to define a topology on $\text{Spec}(A)$. What is perhaps unusual about this is the notion that, in this setting, the prime ideals themselves are to be thought of as points.

This is probably a good time for the reader to review the examples discussed in Example 3.2.5 and Exercise 3.3.5. Let A be a ring, S be a subset of A . Let $V(S)$ stand for the set of all prime ideals of A containing S . Note that

$$V(S) = V(AS) = V(\sqrt{AS}),$$

where AS is the ideal generated by S .

In the following proposition we summarize the basic properties of the sets $V(S)$; in view of the comment just made, we shall assume that S is an ideal.

Proposition 9.2.2. Suppose that A is a ring, that $\{\mathfrak{s}_i : i \in I\}$ is a family of ideals of A , \mathfrak{s} and \mathfrak{t} are ideals of A . Then

(a) $\mathfrak{s} \leq \mathfrak{t}$ implies that $V(\mathfrak{t}) \subseteq V(\mathfrak{s})$.

(b) $V(\mathfrak{s} \cap \mathfrak{t}) = V(\mathfrak{s}\mathfrak{t}) = V(\mathfrak{s}) \cup V(\mathfrak{t})$.

- (c) $V(\sum_i \mathfrak{s}_i) = \cap_i V(\mathfrak{s}_i)$.
- (d) $V(\{0\}) = \text{Spec}(A)$ and $V(A) = \emptyset$.

Proof. We let the reader worry about (a), (c) and (d). Let's show that (b) holds.

Let \mathfrak{p} be a prime ideal of A . Then, if $a \in \mathfrak{s} \setminus \mathfrak{p}$ and $\mathfrak{st} \leq \mathfrak{p}$, we have that $\mathfrak{t} \leq \mathfrak{p}$. This immediately implies that $V(\mathfrak{st}) \subseteq V(\mathfrak{s}) \cup V(\mathfrak{t})$. From (a), one has the containments

$$V(\mathfrak{s}) \cup V(\mathfrak{t}) \subseteq V(\mathfrak{s} \cap \mathfrak{t}) \subseteq V(\mathfrak{st}),$$

whence the identities follow. ■

Definition & Remarks 9.2.3. Proposition 9.2.2 informs us that the collection

$$\{V(\mathfrak{s}) : \mathfrak{s} \text{ is an ideal of } A\}$$

is the collection of closed subsets of a topology, called the *Zariski* or *hull-kernel topology* on $\text{Spec}(A)$. The open sets are

$$\{U(\mathfrak{s}) : \mathfrak{s} \text{ is an ideal of } A\},$$

where $U(\mathfrak{s})$ is the set theoretic complement of $V(\mathfrak{s})$. In particular, letting $a \in A$, we have

$$V(a) = V(\{a\}) = V(Aa) = \{\mathfrak{p} \in \text{Spec}(A) : a \in \mathfrak{p}\}$$

and

$$U(a) = \{\mathfrak{p} \in \text{Spec}(A) : a \notin \mathfrak{p}\}.$$

The sets $U(a)$, for all $a \in A$, form a base for the open sets of the topology. With respect to this topology, one refers to $\text{Spec}(A)$ as the *prime spectrum* of A .

We record an exercise (a sort of companion to Proposition 9.2.2), the upshot of which is precisely that the $U(a)$ ($a \in A$) form a base of open sets. Recall that a family of subsets $\{U_i : i \in I\}$ of the set X form a base for the open sets of a given topology τ on X , if

- (op-i) for each point $p \in U_i \cap U_j$ there is a U_k so that $p \in U_k \subseteq U_i \cap U_j$;
- (op-ii) $X = \cup\{U_i : i \in I\}$;
- (op-iii) each open set is a union of some of the U_i .

Exercise 9.2.4. See Chapter 1, Exercise 17, [AM69]. Suppose that A is a ring. Then

- (a) $U(a) \cap U(b) = U(ab)$, for all $a, b \in A$;
- (b) $U(a) = \emptyset$ if and only if a is nilpotent;
- (c) $U(a) = \text{Spec}(A)$ if and only if a is a unit;
- (d) $U(a) = U(b)$ if and only if $\sqrt{Aa} = \sqrt{Ab}$;
- (e) For each ideal \mathfrak{s} of A ,

$$U(\mathfrak{s}) = \cup\{U(a) : a \in \mathfrak{s}\}.$$

Recall that a subset Y of a topological space is *compact* if every open cover of Y has a finite subcover. (Warning: compact, in these pages, does *not* signify "compact & Hausdorff"!!!)

- (f) Each $U(a)$ is compact; thus, $\text{Spec}(A) = U(1)$ is compact.

- (g) An open set in $\text{Spec}(A)$ is compact if and only if it is a finite union of members of the base $U(a)$ ($a \in A$).

Remark 9.2.5. *On disappointment!* In general $\text{Spec}(A)$ is not a Hausdorff space. Recall that a topological space is *Hausdorff* (or T_2) if for each pair of distinct points p and q there exist disjoint open sets U and V containing p and q , respectively.

For any ring A , observe that if U is an open set in $\text{Spec}(A)$, and $\mathfrak{p} < \mathfrak{q}$ are prime ideals, with $\mathfrak{q} \in U$, then $\mathfrak{p} \in U$. Now, compute $\text{Spec}(\mathbb{Z})$: it consists of the $p\mathbb{Z}$, for all prime numbers p , plus the zero ideal. If p and q are any two distinct positive primes, then if U and V are open sets containing $p\mathbb{Z}$ and $q\mathbb{Z}$, respectively, then $\{0\} \in U \cap V$.

In brief, we will address (and solve) the question of when $\text{Spec}(A)$ is a Hausdorff space in its Zariski topology. However, let us first mention what is (in a sense) the antithesis of being Hausdorff.

Definition 9.2.6. Let X be a topological space. It is said to be *irreducible* if any two nonempty open subsets intersect. Equivalently, X is irreducible if it is not the union of two proper closed subsets. The next exercise says that what happened in \mathbb{Z} is typical of integral domains, and also more generally.

Exercise 9.2.7. Show that $\text{Spec}(A)$ is irreducible if and only if $n(A)$ is a prime ideal.

Definition & Remarks 9.2.8. More generally, letting X be a topological space, and Y be a subset of X , call Y *irreducible* if every pair of open sets U and V which intersect Y nontrivially, intersect each other nontrivially in Y . (This is equivalent to saying that, whenever Y is contained in the union of two closed sets, it is actually a subset of one of them.) Observe, by the definition of “base for a topology” that we may, in this definition, replace “open set” with “basic open set”, for any base we might have in mind.

Suppose that Y is an irreducible subset of X , and U and V are open subsets of X , which intersect $\text{cl}_X Y$, the closure of Y . Then U and V intersect Y , which means that they intersect each other in Y , and hence $\text{cl}_X Y$. This proves the first part of the following.

Proposition 9.2.9. *Suppose that X is a topological space. Then*

- (a) *if Y is an irreducible subset, so is $\text{cl}_X Y$;*
- (b) *if $x \in X$, then $\text{cl}_X \{x\}$ is irreducible;*
- (c) *each irreducible subset is contained in a maximal irreducible subset of X ;*
- (d) *the maximal irreducible subsets of X are closed and cover X ;*
- (e) *if $X = \text{Spec}(A)$, where A is a ring, then the maximal irreducible subsets of $\text{Spec}(A)$ are the ones of the form $V(\mathfrak{p})$, where \mathfrak{p} is a minimal prime ideal of A ;*
- (f) *if \mathfrak{s} is a semiprime ideal of A , then $V(\mathfrak{s})$ is irreducible if and only if \mathfrak{s} is a prime ideal.*

Proof. Having done (a), we brush (c) aside with Zorn’s Lemma. Then (d) follows from an application of (a) and (c).

As to (b), notice that if an open set U intersects $\text{cl}_X \{x\}$, then, in fact, $x \in U$. Then it is clear that $\text{cl}_X \{x\}$ is irreducible.

Now (e): by (b), $V(\mathfrak{p}) = \text{cl}_{\text{Spec}(A)} \{\mathfrak{p}\}$ is irreducible, for each prime ideal \mathfrak{p} . Now, suppose that \mathfrak{p} is a minimal prime ideal, but that $V(\mathfrak{p})$ is not maximal; according to (a), there is a semiprime

ideal \mathfrak{s} of A , so that $V(\mathfrak{s})$ is irreducible and properly contains $V(\mathfrak{p})$. Then $\mathfrak{s} \subseteq \mathfrak{p}$. On the other hand, \mathfrak{s} must be prime: for if $ab \in \mathfrak{s}$, then it is easy to see that $V(\mathfrak{s}) \subseteq V(a) \cup V(b)$, and since $V(\mathfrak{s})$ is irreducible, we may (without loss of generality) assume that $V(\mathfrak{s}) \subseteq V(a)$, which in turn implies that $a \in \sqrt{Aa} \leq \mathfrak{s}$. Since \mathfrak{p} is minimal among prime ideals, we have a contradiction.

Now, since any maximal irreducible subset is closed, it must be of the form $V(\mathfrak{t})$, for some ideal \mathfrak{t} , which must be prime, by the argument in the preceding paragraph. Since we proved that $V(\mathfrak{q})$ is irreducible, for every prime ideal \mathfrak{q} , it follows that if $V(\mathfrak{t})$ is a maximal irreducible subset, then \mathfrak{t} is a minimal prime ideal.

Finally (f): Note that, if \mathfrak{s} is a semiprime ideal, then the map $\mathfrak{p} \mapsto \mathfrak{p}/\mathfrak{s}$ is a homeomorphism from $V(\mathfrak{s})$ onto $\text{Spec}(A/\mathfrak{s})$. (That it is a bijection is well known; the reader will easily check that it is continuous and open.) By Exercise 9.2.7, the latter is irreducible if and only if $n(A/\mathfrak{s}) = \{0\}$ is a prime ideal; i.e., precisely when \mathfrak{s} itself is prime. ■

Exercise 9.2.10. Determine the maximal irreducible subsets of a Hausdorff space.

Putting all the above together we obtain the following proposition, the proof of which we leave as an exercise. Have another look at Theorem 9.1.11 as well.

Proposition 9.2.11. *For a ring A , the following are equivalent:*

- (a) $\text{Spec}(A)$ is Hausdorff.
- (b) Every prime ideal is maximal.
- (c) $A/n(A)$ is von Neumann regular.
- (d) $\text{Spec}(A)$ is T_1 ; (which is to say that each singleton set is closed.)

Remarks 9.2.12. The reader should know that any Hausdorff topological space is T_1 . In general, the reverse is not true. On the set \mathbb{N} of natural numbers, declare a subset closed if it is finite. This defines a T_1 space which is not Hausdorff. (So, in view of Proposition 9.2.11, this space cannot be $\text{Spec}(A)$, for any ring A !)

T_1 spaces can also be regarded as having the following kind of separation: if x and y are distinct points, then there is an open set which contains x but not y . There is a still weaker separation property, which $\text{Spec}(A)$ possesses, for each commutative ring A . The topological space X is T_0 if for any two distinct points x and y , there is an open set containing one of the points but not the other.

It is high time to abstract the topological flavor of a prime spectrum.

Definition 9.2.13. A topological space X is said to be *spectral* if it satisfies the following conditions:

- (spec-a) X has T_0 separation.
- (spec-b) Every closed irreducible subset of X is the closure of a singleton set.
- (spec-c) X has a base of compact and open subsets.

Since in $\text{Spec}(A)$ each closed set is of the form $V(\mathfrak{s})$, for a suitable semiprime ideal \mathfrak{s} , and $V(\mathfrak{p}) = \text{cl}_{\text{Spec}(A)}\{\mathfrak{p}\}$, for each prime ideal \mathfrak{p} , Proposition 9.2.9(f) guarantees that (spec-b) holds in $\text{Spec}(A)$. That (spec-c) is satisfied is given by Exercise 9.2.4(f), and we shall let the reader verify that (spec-a) holds in $\text{Spec}(A)$.

Let's turn to the functorial properties of Spec .

Remark 9.2.14. We suppose that $f : A \rightarrow B$ is a ring homomorphism, which preserves the identity. Since the inverse image of a prime ideal is prime, we have that, for each $\mathfrak{q} \in \text{Spec}(B)$, $f^{-1}(\mathfrak{q})$ is a prime ideal. Previously called the contraction \mathfrak{q}^c of \mathfrak{q} , we denote it here by $\text{Spec}(f)(\mathfrak{q})$. Thus, $\text{Spec}(f)$ defines a map from $\text{Spec}(B)$ to $\text{Spec}(A)$.

We record the basic properties of $\text{Spec}(\)$ in an exercise. See Chapter 1, Exercise 21, in [AM69].

Exercise 9.2.15. Suppose that $f : A \rightarrow B$ is a ring homomorphism, which preserves the identity. Then the following hold.

- (a) For each $a \in A$, $\text{Spec}(f)^{-1}(U(a)) = U(f(a))$; therefore, $\text{Spec}(f)$ is continuous, relative to the respective Zariski topologies.
- (b) For each ideal \mathfrak{s} of B , $\text{cl}_{\text{Spec}(A)}(\text{Spec}(f)(V(\mathfrak{s}))) = V(\mathfrak{s}^c)$.
- (c) If f is surjective, then $\text{Spec}(f)$ is a homeomorphism of $\text{Spec}(B)$ onto its image, $V(\ker(f))$.
- (d) If f is one-to-one, then the image of $\text{Spec}(f)$ is dense in $\text{Spec}(A)$.
- (e) Spec is a contravariant functor from **Crn1** to the subcategory **Spec** of all spectral spaces and all continuous maps between them.
- (f) Suppose that A is an integral domain having just one nonzero prime ideal \mathfrak{p} , and let K be its field of fractions. Let $B = A/\mathfrak{p} \times K$. Define $f : A \rightarrow B$ by $f(a) = (P + a, a)$. Prove that $\text{Spec}(f)$ is a bijection, but not a homeomorphism.

Hot Air 9.2.16. *Looking Back.* As we have seen in 5.2.4(b)(3), the map $A \rightarrow A/n(A)$ is a reflection, which is adjoint to the inclusion functor of categories, embedding the subcategory **SP** of semiprime rings in **Crn1**. We use the notation of 5.2.4(b), denoting the reflection functor by $\hat{A} = A/n(A)$.

In this connection, Exercise 9.2.15(c) implies that $\text{Spec}(\)$ and $\text{Spec}(\hat{\ })$ are naturally equivalent functors. Put differently: for each ring A , $\text{Spec}(A)$ is naturally homeomorphic to $\text{Spec}(A/n(A))$.

In a sense this says – although one should not push this point too far, as it rarely surfaces with practical significance – that in dealing with spaces of prime ideals, one can just as well deal with semiprime rings.

Definition & Remarks 9.2.17. The time has come to consider some important subspaces of $\text{Spec}(A)$ that come up from time to time. First, $\text{Max}(A)$ stands for the subset of $\text{Spec}(A)$ consisting of all maximal ideals of A . We endow $\text{Max}(A)$ with the subspace topology; explicitly, the closed sets are the ones of the form

$$V_M(\mathfrak{s}) = \{ \mathfrak{m} \in \text{Max}(A) : \mathfrak{m} \geq \mathfrak{s} \},$$

where \mathfrak{s} ranges over all the ideals of A . This is also referred to as the Zariski topology. The term “hull–kernel topology” is used as well for $\text{Max}(A)$. $\text{Max}(A)$ is also commonly called the *max-spectrum* of A .

Since, in general, the inverse image of a maximal ideal, under a ring homomorphism, is not maximal, Max is not functorial.

Let $\text{Min}(A)$ denote the subset of $\text{Spec}(A)$, consisting of all minimal prime ideals of A . As with Max , one endows $\text{Min}(A)$ with the subspace topology, in which the closed sets are of the form

$$V_m(\mathfrak{s}) = \{ \mathfrak{m} \in \text{Min}(A) : \mathfrak{m} \supseteq \mathfrak{s} \}.$$

Min fails to be functorial as well, as the inverse image of a minimal prime need not be minimal.

In $\text{Min}(A)$ (resp. $\text{Max}(A)$) denote the basic open sets by $U^m(a)$ (resp. $U^M(a)$), for all $a \in A$. Lemmas 9.2.18 and 9.2.19 highlight some of the fundamental properties of minimal prime ideals. The Proof of the first involves Zorn's Lemma, and is left to the reader.

First recall that

$$\text{Ann}(x) = \{ a \in A : xa = 0 \};$$

this ideal is the *annihilator* of a .

A subset T of nonzero elements of A is said to be a *multiplicative system* if T is closed under multiplication.

Lemma 9.2.18. *Let A be a ring, and T be a multiplicative system.*

- (a) *There exists a prime ideal \mathfrak{q} which is maximal with respect to the condition $\mathfrak{q} \cap T = \emptyset$.*
- (b) *T is contained in a maximal multiplicative system.*
- (c) *Every prime ideal contains a minimal prime ideal.*

Proof. All three parts are proved by Zorn's Lemma; in (c) one applies it to the set of all prime ideals contained in the given one, partially ordered by *reverse* inclusion. ■

Lemma 9.2.19. *Suppose that A is a semiprime ring and $a \in A$. If \mathfrak{p} is a minimal prime ideal, then either $a \in \mathfrak{p}$ or $\text{Ann}(a) \leq \mathfrak{p}$, but not both.*

More generally, if \mathfrak{p} is any minimal prime ideal, then $A \setminus \mathfrak{p}$ is a maximal multiplicative system. Conversely, if T is a maximal multiplicative system of A , let

$$\nu(T) = \cup \{ \text{Ann}(t) : t \in T \};$$

then $\nu(T)$ is a minimal prime ideal. Moreover, the maps $\mathfrak{p} \mapsto A \setminus \mathfrak{p}$ and $T \mapsto \nu(T)$ are mutually inverse bijections between $\text{Min}(A)$ and the set of all maximal multiplicative systems of A .

Proof. The first claim follows from the assertions in the second paragraph. For if, $a \notin \mathfrak{p}$, then $\text{Ann}(a) \leq \mathfrak{p}$, by definition of a prime ideal. If $a \in \mathfrak{p}$, then, since $A \setminus \mathfrak{p}$ is a maximal multiplicative system, $ax = 0$ must hold for some $x \notin \mathfrak{p}$, otherwise the set of all products ay ($y \notin \mathfrak{p}$) generates a multiplicative system which properly contains $A \setminus \mathfrak{p}$. Thus, $\text{Ann}(a) \not\leq \mathfrak{p}$.

Now to the claims in the second paragraph. First $A \setminus \mathfrak{p}$ is a multiplicative system, for any prime ideal \mathfrak{p} , by definition.

If T is a maximal multiplicative system, consider $\nu(T)$. It is easy to check that this is an ideal. (Note that the $\text{Ann}(t)$, with $t \in T$, form a directed union.) If $ab \in \nu(T)$, but $a \notin \nu(T)$, then $abx = 0$, for some $x \in T$, while at is nonzero, for each $t \in T$. But, owing to the maximality of T , this means that $a \in T$. Thus, $b(ax) = 0$, which says that $b \in \nu(T)$, and shows that $\nu(T)$ is a prime ideal.

Next, observe that

$$A \setminus \nu(T) = \{ a \in A : at \neq 0, \forall t \in T \} = T,$$

because of the maximality of T . So if \mathfrak{q} is a prime ideal of A , with $\mathfrak{q} < \nu(T)$, then T is properly contained in $A \setminus \mathfrak{q}$, violating maximality. This shows that $\nu(T)$ is a minimal prime ideal, for each maximal multiplicative system T .

Now, if \mathfrak{p} is a minimal prime ideal, we choose (by Lemma 9.2.18(b)) a maximal multiplicative system T , containing $A \setminus \mathfrak{p}$; then $\nu(T) \leq \mathfrak{p}$, for if $at = 0$, for some $t \in T$, then $a \notin T$, which implies that $a \in \mathfrak{p}$. By what has already been demonstrated, $\nu(T) = \mathfrak{p}$, whence $T = A \setminus \mathfrak{p}$, proving that if \mathfrak{p} is a minimal prime, then $A \setminus \mathfrak{p}$ is a maximal multiplicative system.

Summarizing: we have shown that

- (i) if \mathfrak{p} is a minimal prime ideal, then $A \setminus \mathfrak{p}$ is a maximal multiplicative system;
- (ii) if T is a maximal multiplicative system, then $\nu(T)$ is a minimal prime ideal, and $A \setminus \nu(T) = T$.

What remains to be shown is that if \mathfrak{p} is a minimal prime ideal, then $\nu(A \setminus \mathfrak{p}) = \mathfrak{p}$; we leave this to the reader. ■

We can now see that $\text{Min}(A)$ is always a zero-dimensional space in the hull-kernel topology; see Example 3.2.5.

Corollary 9.2.20. *Suppose that A is a ring. Then, for each $a \in A$, $\text{Min}(A) \setminus U^m(a) = U^m(\text{Ann}(a))$. Thus, each basic open set $U^m(a)$ is closed and $\text{Min}(A)$ has a base of clopen sets.*

Putting together Corollary 9.2.20 and Proposition 9.2.11, we get a very satisfying conclusion to this section.

Theorem 9.2.21. *For each von Neumann regular ring A , $\text{Spec}(A) = \text{Max}(A) = \text{Min}(A)$ is a compact, Hausdorff and zero-dimensional space.*

9.3 Stone Duality.

Hot Air 9.3.1. Most Refined. In this section we prove the famous “Stone Duality”, between **KZero**, the category of compact, Hausdorff, zero-dimensional spaces, together with all continuous maps, and the category **Bool**, of all boolean algebras and boolean homomorphisms.

The reader would do well, at this point, to review Exercise 3.3.5 and the notion of equivalence of categories. A look at Exercise 1.5.7, on the relationship between boolean rings and boolean algebras is also highly recommended. Specifically, two things should be highlighted about the connection between boolean algebras and boolean rings. Suppose that B is a boolean ring. Let \vee, \wedge and $()'$ denote the associated boolean operations on B . (Recall: $a \wedge b = ab$, $a \vee b = a + b + ab$, and $a' = 1 + a$; recall that a boolean ring has characteristic 2.) Then

- (i) a subset \mathfrak{r} is an ideal of the ring structure if and only if it is a boolean ideal; (this is part of Exercise 1.5.7;)
- (ii) if $f : B \rightarrow C$ is a **Crn1**-morphism between boolean rings, then it is a **Bool**-morphism between the associated boolean algebra structures, and conversely. (This was not mentioned earlier; it is easy to verify.)

What all this says is summarized in the following proposition, the details of which we leave to the reader.

Proposition 9.3.2. ***Bool** and the full subcategory **BRng** of **Crn1**, consisting of all boolean rings, are naturally equivalent categories.*

We also have the following categorical “rephrasing” of Theorem 9.2.21.

Proposition 9.3.3. *Spec is a contravariant functor from \mathbf{BRng} to \mathbf{KZero} .*

The point of Stone Duality is that, in the above context, Spec is a duality. Most of the rest of this section is devoted to proving that.

Let us recover the functor which accompanies Spec in this duality.

Remarks 9.3.4. In Example 3.2.5 it was shown that \mathfrak{B} , the functor which assigns, to each topological space X , its boolean algebra of clopen sets $\mathfrak{B}(X)$, is a contravariant functor from \mathbf{Top} to \mathbf{Bool} . Here we think of $\mathfrak{B}(X)$, instead, as a boolean ring (relative to $+$ as symmetric difference, and \cdot as set theoretic intersection). It easy to verify that \mathfrak{B} then becomes a contravariant functor from \mathbf{Top} to \mathbf{BRng} .

Next, we now restrict \mathfrak{B} to \mathbf{KZero} . To show that Spec is a duality, as promised, it suffices to show that, for each boolean ring B , B is naturally isomorphic to $\mathfrak{B}(\text{Spec}(B))$, and, for each compact, zero-dimensional Hausdorff space X , X and $\text{Spec}(\mathfrak{B}(X))$ are naturally homeomorphic.

For each boolean ring B , and $x \in B$, let $U_B(x) = U(x)$, the basic open set determined by x in $\text{Spec}(B)$. Exercise 3.3.5(g) asserts that the map U_B is a \mathbf{Bool} -isomorphism, hence also a \mathbf{BRng} -isomorphism. Part (h) of 3.3.5 assures us that

$$U : \mathbf{1BRng} \longrightarrow \mathfrak{B} \cdot \text{Spec}$$

is natural. That's half the proof of duality.

Next, suppose that X is a compact, zero-dimensional Hausdorff space; we consider $\text{Spec}(\mathfrak{B}(X))$. For each $x \in X$, let

$$\mathfrak{m}_x = \{ C \in \mathfrak{B}(X) : x \notin C \}.$$

In the following proposition we summarize the main features of the association $x \mapsto \mathfrak{m}_x$.

Proposition 9.3.5. *Suppose that X is compact, Hausdorff and zero-dimensional. Then we have the following items.*

- (a) *For each $x \in X$, \mathfrak{m}_x is a maximal ideal of $\mathfrak{B}(X)$.*
- (b) *The map $x \mapsto \mathfrak{m}_x$ is a homeomorphism of X onto $\text{Spec}(\mathfrak{B}(X))$.*
- (c) *$\mathfrak{m} : \mathbf{1KZero} \longrightarrow \text{Spec} \cdot \mathfrak{B}$ is a natural transformation.*

Proof. (a) We let the reader verify that \mathfrak{m}_x is closed under symmetric difference; it should then be clear that it is an ideal. As to the maximality, if $C \in \mathfrak{B}(X) \setminus \mathfrak{m}_x$, then $x \in C$, and so $X \setminus C \in \mathfrak{m}_x$ and $X = C + (X \setminus C)$.

(b) We first show that $x \mapsto \mathfrak{m}_x$ is onto $\text{Spec}(\mathfrak{B}(X))$. From Theorem 9.2.21, every prime ideal is maximal. Suppose that \mathfrak{m} is a maximal ideal of $\mathfrak{B}(X)$; its complement \mathfrak{m}^* is a multiplicative system. This means that \mathfrak{m}^* is a collection of clopen sets in a compact space, such that any finite number of them have finite intersection. By compactness, $\bigcap \mathfrak{m}^*$ is nonvoid; we argue that it contains a single point!

Theorem 9.2.21 also asserts that all prime ideals are minimal; by Lemma 9.2.19, \mathfrak{m}^* is a maximal multiplicative system. Therefore, if $p \in \bigcap \mathfrak{m}^*$, and C is a clopen set containing p , then $C \in \mathfrak{m}^*$. This says that if $q \in \bigcap \mathfrak{m}^*$, then any clopen set containing p also contains q . This contradicts that X is Hausdorff. Thus, $\bigcap \mathfrak{m}^*$ contains a single element x , and it is easy to see that $\mathfrak{m} = \mathfrak{m}_x$.

The Hausdorff feature plus zero-dimensionality imply that the map $x \mapsto \mathfrak{m}_x$ is one-to-one.

Now, if C is any clopen set, then

$$C = \{ x \in X : C \notin \mathfrak{m}_x \} = U(C);$$

this makes sense; in $U(C)$, C is viewed as an element of $\mathfrak{B}(X)$. We've then proved that the map $x \mapsto \mathfrak{m}_x$ is both open and continuous; this completes the proof of (b).

(c) is straightforward, and is left to the reader. ■

All of the above complete the proof of

Theorem 9.3.6. Stone Duality. $\text{Spec} : \mathbf{Bool} \longleftrightarrow \mathbf{KZero} : \mathfrak{B}$ is a duality.

We turn to a few applications of Theorem 9.3.6. For that, remember what a duality entails. In this case, it implies that the functors Spec and \mathfrak{B} take products to coproducts and vice versa. There are many other consequences, in particular, those involving projectives in \mathbf{KZero} and injectives in \mathbf{Bool} . We shall not pursue those important results here.

Instead, we limit the discussion to highlighting the role of the Cantor set in this context, in the exercises which follow. For this we shall need some theorems from topology. We refer the reader to [Wi70], Theorem 30.3 and the Urysohn Metrization Theorem (23.1) for details. We cite any additional reference further on.

The first exercise follows from Stone Duality, for boolean algebras, but is true more generally. By the *disjoint union* of two spaces X_1 and X_2 we mean just that, set theoretically, so that a set is open in the union if and only if its intersection with each X_i is open.

Exercise 9.3.7. Suppose A_1 and A_2 are rings. Then $\text{Spec}(A_1 \times A_2)$ is homeomorphic to the disjoint union of $\text{Spec}(A_1)$ and $\text{Spec}(A_2)$.

Exercise 9.3.8. Let X denote any infinite set. The *one-point compactification* of X , denoted αX , is the topological space, containing X , plus one additional point α , for which the open sets are

- (i) the subsets of X , and
- (ii) the subsets containing α , which have finite complement in X .

Prove that αX is a compact, zero-dimensional Hausdorff space, whose Stone dual is the boolean algebra $B_f(X)$ of all subsets of X which are either finite or else have finite complement. (Hint: Show that $B_f(X)$ has two types of maximal ideals, the \mathfrak{m}_x , (for each $x \in X$) consisting of all subsets in $B_f(X)$ which do not contain x , plus one additional ideal \mathfrak{m}_∞ , containing all the finite subsets.)

Exercise 9.3.9. *Stone-Čech compactification.* Let X be any infinite set, and define

$$\beta X \equiv \text{Spec}(\mathcal{P}(X)),$$

where $\mathcal{P}(X)$ denotes the power set of X . Prove that βX has the following property: if $g : X \rightarrow K$ is any mapping into a compact, zero-dimensional Hausdorff space K , then there is a unique continuous map $g^* : \beta X \rightarrow K$ such that $g^* \cdot M = g$, where $M : X \rightarrow \beta X$ stands for the map

$$M(x) = \mathfrak{m}_x = \{ Y \subseteq X : x \notin Y \}.$$

(Note: actually, this universal property is valid for all compact Hausdorff spaces K . Considerable (!!!) additional material is required to prove that! βX is called the **Stone-Čech compactification** of X , and as the property suggests, it is the “largest” or “freest” of all compactifications of X . A *compactification* of X (technically speaking) is any compact Hausdorff space which contains X as a dense subspace.)

The reader should also note that β is, in this context, the free functor from \mathbf{Set} to \mathbf{KZero} , and left adjoint to the underlying-set functor which forgets that a compact, Hausdorff and zero-dimensional space has any structure at all.

Remarks 9.3.10. We assume that the reader is familiar with the traditional description of the Cantor set: from the unit interval one excludes the middle third open interval; then the middle thirds from the two closed intervals that remain, etc. ... ending up by taking the (countable) intersection of the sets left at each stage of the process.

The thing to remember is that the Cantor set is compact, second countable (it has a countable base, as do all compact metric spaces,) Hausdorff, zero-dimensional, having no isolated points. A very nice theorem from topology asserts that any topological space which is compact, second countable, Hausdorff, zero-dimensional and has no isolated points, is homeomorphic to the Cantor set. (Theorems 30.3 and 23.1 in [Wi70])

Here's another way to look at this matter; the reader should refer to 17.9 in [Wi70]. Let I be a set. Denote by $\mathbf{2}^I$ the product of I copies of the set $\mathbf{2} = \{0, 1\}$, with the trivial topology. Topologize $\mathbf{2}^I$ with a Zariski-like topology; the basic open sets are the following: for each finite subset F of I , and each $g \in \mathbf{2}^I$, let

$$U(F, g) = \{ f \in \mathbf{2}^I : f(i) = g(i), \forall i \in F \}.$$

The reader can easily verify (see Exercise 9.3.11) that the conditions are satisfied to make of the collection

$$\{ U(F, g) : F \subseteq I \text{ finite, } g \in \mathbf{2}^I \},$$

a base for a topology. Under this topology $\mathbf{2}^I$ is compact, Hausdorff and zero-dimensional. If I is infinite then $\mathbf{2}^I$ has no isolated points.

The connection with the Cantor set is this: if I is countably infinite, then $\mathbf{2}^I$ is second countable. Applying the characterization just quoted, it follows that $\mathbf{2}^I$ is homeomorphic to the Cantor set.

These spaces $\mathbf{2}^I$ are, generically called *Cantor spaces*. Notice, by the way, that if I is finite, then $\mathbf{2}^I$ is discrete.

Exercise 9.3.11. Show that if $\mathbf{2}^I$ is topologized as indicated in 9.3.10, then it is indeed a compact, Hausdorff, zero-dimensional space. $\mathbf{2}^I$ has no isolated points if and only if I is infinite.

Exercise 9.3.12. Suppose that I is a set. Prove that the dual of the Cantor space $\mathbf{2}^I$ is the free boolean algebra on the set I .

(Hint: Categorically speaking, the free boolean algebra on I is the coproduct of I copies of the boolean algebra of four elements, which is the free boolean algebra on a singleton. On the other hand, the Cantor space $\mathbf{2}^I$ is the product of I copies of the space $\mathbf{2}$ in **KZero**.)

For the next exercise the reader might have a look at Exercise 1.6.9.

Exercise 9.3.13. Suppose that B is a boolean algebra. Prove that B has no atoms if and only if $\text{Spec}(B)$ has no isolated points. (Recall that an *atom* is an element $a > 0$, which is minimal with respect to " > 0 ".)

(Hint: Show that $0 < a \in B$ is an atom if and only if $\text{Ann}(a) = \{ b \in B : b \leq a' \}$ is a maximal ideal, if and only if $U(a)$ is a singleton.)

Putting the previous three exercises together we get:

Exercise 9.3.14. (a) The free boolean algebra on an infinite set has no atoms.

(b) The free boolean algebra on a countably infinite alphabet is (Stone) dual to the Cantor set.

And recalling an old exercise from §1.5:

Exercise 9.3.15. The free boolean algebra on the finite set I is the power set of $\mathbf{2}^I$.

We conclude the chapter with an exercise on spaces of maximal ideals.

Definition 9.3.16. Let X be a Hausdorff space. $C(X)$ stands for the ring of all continuous, real valued functions defined on X . $C(X)$ is a commutative ring with identity, with respect to pointwise addition and multiplication. For each point $x \in X$, let

$$\mathfrak{m}_x \equiv \{ f \in C(X) : f(x) = 0 \},$$

and for each $g \in C(X)$ denote by

$$\text{coz}(g) = \{ x \in X : g(x) \neq 0 \}.$$

Note that, since g is continuous, $\text{coz}(g)$ is an open set of X .

Exercise 9.3.17. Suppose that X is a compact, Hausdorff space. Show the following:

- (a) \mathfrak{m}_x is a maximal ideal, for each $x \in X$.
- (b) The map $\mu(x) = \mathfrak{m}_x$ is onto $\text{Max}(C(X))$.
- (c) μ is one-to-one. (You need to show that the sets $\text{coz}(g)$ (all $g \in C(X)$) is a base for the topology.)
- (d) $\mu(\text{coz}(g)) = U^m(g)$, for each $g \in C(X)$, which proves that μ is a continuous and open map, and hence a homeomorphism.

Conclusion: $X \cong \text{Max}(C(X))$, for each compact Hausdorff space X .

With appropriate conditions placed on the category of rings, one can make Max functorial, and obtain a rather useful adjoint situation between Max and the contravariant $C(\cdot)$. We refer the reader to [LZ71] for more information.

10. Commutative Algebra: Extensions of Rings

In this chapter we consider rings of fractions, localization and the concept of integral extensions. As in the preceding chapter, all rings will be commutative with identity. All categorical considerations will (unless the contrary is specified) refer to **Crn1**.

As in the previous chapter, our main reference will be [AM69].

10.1 Rings of Fractions and Localization. In the upcoming discussion A is a ring, and S is a multiplicative system of A . Consider the set of all pairs $[a, s]$, where $a \in A$, $s \in S$, and define $[a, s] \sim [b, t]$, provided there is a $u \in S$ so that $(at - bs)u = 0$. We leave it to the reader to verify that \sim is an equivalence relation. (Else, see [AM69], p. 36.)

Note: in [AM69] multiplicative systems are allowed to contain 0; we do not consider that case, because then the ring of fractions we are about to define is trivial.

Definition 10.1.1. $S^{-1}A$ denotes the set of all equivalence classes modulo \sim . We denote the equivalence class of $[a, s]$ by a/s . Define addition and multiplication on $S^{-1}A$ in the expected manner:

$$(a/s) + (b/t) = (at + bs)/st \quad \text{and} \quad (a/s)(b/t) = ab/st.$$

It is left as an exercise to show that these definitions are independent of the representatives used, and that they make of $S^{-1}A$ a commutative ring with identity s/s (for any $s \in S$). Unlike some sources, we do not require the identity of the ring to belong to S .

There is a canonical homomorphism of A into $S^{-1}A$, denoted by $\phi(a) = as/s$ (again, independent of the choice of $s \in S$).

The following proposition gives the essential properties of the ring $S^{-1}A$, called *the ring of fractions of A with respect to S* .

Proposition 10.1.2. *Suppose that S is a multiplicative system of A . Let $\phi : A \longrightarrow S^{-1}A$ be given by $\phi(a) = as/s$. Then*

- (a) *if $s \in S$, then $\phi(s)$ is a unit of $S^{-1}A$;*
- (b) $\ker(\phi) = \{ a \in A : \exists s \in S, as = 0 \}$.
- (c) *If $f : A \longrightarrow B$ is a homomorphism such that $f(s)$ is a unit of B , for each $s \in S$, then there is a unique homomorphism $f^* : S^{-1}A \longrightarrow B$, such that $f = f^* \phi$.*

Proof. (a) $(ts/s)(s/ts) = u/u$, for all $s, t, u \in S$.

(b) If $as = 0$, for some $s \in S$, then $\phi(a) = as/s = 0/s = 0$. Conversely, $at/t = 0$ implies that $at/t = 0/s$, for any $s \in S$, which means that there is a $u \in S$, such that $(ats - 0)u = 0$.

(c) Define $f^*(a/s) = f(a)/f(s)$; this makes sense, since $f(s)$ is invertible. And if $a/s = b/t$, then $(at - bs)u = 0$, for some $u \in S$, whence $(f(a)f(t) - f(b)f(s))f(u) = 0$. Since $f(u)$ is invertible, we may cancel it; this gives that $f(a)/f(s) = f(b)/f(t)$, proving that the function f^* is well defined. We let the reader convince himself that f^* is a homomorphism of rings, preserving the identity. ■

With an additional observation we obtain the following characterization of $S^{-1}A$. See Corollary 3.2 of [AM69].

Corollary 10.1.3. *Suppose that B is a ring, with a homomorphism $g : A \longrightarrow B$ so that*

- (i) $g(s)$ is a unit, for each $s \in S$;
- (ii) $\ker(g) = \{a \in A : \exists s \in S, as = 0\}$; and
- (iii) each $b \in B$ can be expressed as $b = g(a)/g(s)$, for some $a \in A$ and $s \in S$.

Then there is a unique isomorphism $h : S^{-1}A \longrightarrow B$, such that $h \cdot \phi = g$.

Hot Air 10.1.4. *For Category Lovers.* Needless to say, the previous results broadly hint at the presence of a functor. Let us explain, but only in brief.

Let **CRnS** be the category whose objects are pairs (A, S) , with A a commutative ring with 1 and S a multiplicative system of A , with morphisms defined as follows: $f : (A, S) \longrightarrow (B, T)$ is a morphism if f is a morphism of **Crn1** and $f(S) \subseteq T$. With this setup, the passage from (A, S) to $S^{-1}A$ is functorial, by Proposition 10.1.2. But having come this far, let us be more specific.

The functor $\text{Frac} : \mathbf{CRnS} \longrightarrow \mathbf{Crn1}$ is defined by $\text{Frac}(A, S) = S^{-1}A$, on objects. Incidentally, **Crn1** is to be thought of as a full subcategory of **CRnS**: each commutative ring A has a canonically associated multiplicative system, namely $U(A)$ the group of units. Morphisms of **Crn1** automatically preserve units. We let E denote the embedding functor of **Crn1** in **CRnS**.

Let $\phi_{(A,S)} : A \longrightarrow E(S^{-1}A)$ be the map $\phi_{(A,S)}(x) = xs/s$, for an arbitrary $s \in S$. Continuing, if $f : (A, S) \longrightarrow (B, T)$ is a morphism of **CRnS**, then $\phi_{(B,T)} \cdot f$ maps S to units of $T^{-1}B$, so that there is (by 10.1.2) a unique homomorphism $\text{Frac}(f) : S^{-1}A \longrightarrow T^{-1}B$ such that

$$\text{Frac}(f) \cdot \phi_{(A,S)} = \phi_{(B,T)} \cdot f.$$

This observation shows that Frac is a functor, indeed a reflection, with $\phi : 1_{\mathbf{CRnS}} \longrightarrow E \cdot \text{Frac}$ as natural transformation.

Of course, now Corollary 10.1.3 is a triviality.

Examples 10.1.5. Here are some important types of rings of fractions by multiplicative systems.

- (a) If $S = A \setminus \mathfrak{p}$, where \mathfrak{p} is a prime ideal, then we denote $S^{-1}A = A_{\mathfrak{p}}$, and call it *the localization of A at \mathfrak{p}* . We shall have plenty of occasion to consider properties of localizations in the next section, but for now shall say no more.
- (b) If $s \in A$, then $S = \{s^n : n \in \mathbb{N}\}$ is a multiplicative system, provided s is *regular*; that is to say, a non-zero-divisor. We use the notation $S^{-1}A = A_s$ in this event.
- (c) Suppose that $A = F[T_1, T_2, \dots, T_n]$, the polynomial ring over a field F , in n indeterminates. Let \mathfrak{p} be a prime ideal of A . The localization $A_{\mathfrak{p}}$ at \mathfrak{p} can be viewed as follows: its elements are ratios f/g of polynomials, with $g \notin \mathfrak{p}$. So suppose that $V(\mathfrak{p})$ stands for the set

$$V(\mathfrak{p}) = \{(x_1, \dots, x_n) \in F^n : \forall f \in \mathfrak{p}, f(x_1, \dots, x_n) = 0\}.$$

Then we may (assuming F is infinite) think of the ratios of $A_{\mathfrak{p}}$ as functions which (to quote [AM69]) “are defined at almost every point of $V(\mathfrak{p})$ ”. No more will be said about this in these pages, except in the example that follows. For a more thorough (and classical) treatment of this subject we refer the reader to [Sh77].

Here’s an illustration: consider $A = \mathbb{R}[T_1, T_2]$, and \mathfrak{p} , the ideal generated by $T_2 - T_1^2$. Then $A_{\mathfrak{p}}$ can be thought of as the set of rational functions defined at all but a finite number of points of the curve $T_2 = T_1^2$. This type of example also identifies where the name “localization” comes from: by passing from A to $A_{\mathfrak{p}}$ one “localizes” to the curve in question.

- (d) If A is an integral domain, then any ring of fractions $S^{-1}A$ is also an integral domain, and, indeed, a subring of the field of fractions, which we shall denote qA . By the way, $qA = S^{-1}A$, when S consists of all the nonzero elements of A .

The notion of ring of fractions can be extended to the modules over that ring.

Definition 10.1.6. Let M be an A -module, and S be a multiplicative system of A . We denote, as in the case of rings of fractions, by $S^{-1}M$, the set of all equivalence classes m/s , with $m \in M$ and $s \in S$, subject to the identity: $m/s = n/t$ if and only if $u(tm - sn) = 0$, for a suitable $u \in S$.

With addition again defined as

$$m/s + n/t = (tm + sn)/st,$$

and scalar multiplication defined by

$$(a/s)(m/t) = am/st,$$

one makes $S^{-1}M$ into an $S^{-1}A$ -module. $S^{-1}M$ is called *the module of fractions with respect to S* .

Note that $S^{-1}M = 0$ precisely when there exists an $s \in S$ for which $sM = 0$; i.e., if and only if $\text{Ann}(M) \cap S \neq \emptyset$.

$S^{-1}(\)$ becomes a functor, from $\mathbf{A}\mathbf{Mod}$ to $\mathbf{S}^{-1}\mathbf{A}\mathbf{Mod}$. We leave the verification of this, as well as the remaining details concerning the definition of modules of fractions, to the reader. (If $f : M \rightarrow N$ is an $\mathbf{A}\mathbf{Mod}$ -morphism, then define $S^{-1}f(m/s) = f(m)/s$.)

The next proposition tells us that S^{-1} preserves exactness.

Proposition 10.1.7. *If the sequence of A -modules*

$$M \xrightarrow{f} M' \xrightarrow{g} M''$$

is exact at M' , then the sequence

$$S^{-1}M \xrightarrow{S^{-1}f} S^{-1}M' \xrightarrow{S^{-1}g} S^{-1}M''$$

is exact at $S^{-1}M'$.

Proof. Functoriality implies that the image of $S^{-1}f$ is contained in $\ker(S^{-1}g)$. As to the reverse, if $S^{-1}g(m/s) = 0$, then $g(m)/s = 0$, which implies that $g(tm) = tg(m) = 0$, for some $t \in S$; that is, $tm \in \ker(g)$. This means that $tm = f(x)$, for some $x \in M$; that is, $m/s = f(x)/st = S^{-1}f(x/st)$. ■

As a corollary of Proposition 10.1.7 we get the next item; we shall let the reader check [AM69], to see the proof; there it is Corollary 3.4. One needs Proposition 10.1.7, so that the corollary makes sense to begin with.

Corollary 10.1.8. *Suppose that N and P are A -submodules of the A -module M . Then*

- (a) $S^{-1}(N \cap P) = S^{-1}N \cap S^{-1}P$;
- (b) $S^{-1}(N + P) = S^{-1}N + S^{-1}P$;
- (c) as $S^{-1}A$ -modules $S^{-1}(M/N) \cong S^{-1}M/S^{-1}N$.

And now the piece that cements rings and modules of fractions together:

Theorem 10.1.9. *Let S be a multiplicative system of A , and M be an A -module. Then the functors S^{-1} and $S^{-1}A \otimes_A (\)$ are naturally equivalent. More specifically, the map $\alpha_M : S^{-1}A \otimes_A M \rightarrow S^{-1}M$ defined by*

$$\alpha_M(a/s \otimes m) = am/s,$$

for all $a \in A$, $s \in S$ and $m \in M$, defines a natural equivalence, which, in particular, is an $S^{-1}A$ -isomorphism.

Proof. First of all, $S^{-1}A \otimes_A M$ is an $S^{-1}A$ -module, as explained in earlier chapters, because $S^{-1}A$ is an $S^{-1}A$ -module over itself. Observe that the map $\theta(a/s, m) = am/s$ is A -bilinear. This right away implies the existence of α_M , unique with respect to the stated condition. It should be clear that α_M is surjective.

Next let's observe the following. A typical element of $S^{-1}A \otimes_A M$ is of the form $\sum_i a_i/s_i \otimes m_i$; let s be the product of all the s_i and t_i be the product of all s_j ($j \neq i$). Then

$$\sum_i a_i/s_i \otimes m_i = \sum_i (a_i t_i)/s \otimes m_i = 1/s \otimes \left(\sum_i a_i t_i m_i \right);$$

the point of which is that each element in this tensor product may be written as $1/s \otimes m$. Now, if $\alpha_M(1/s \otimes m) = 0$, then $m/s = 0$, so that $tm = 0$, for some $t \in S$. Thus,

$$1/s \otimes m = t/st \otimes m = 1/st \otimes tm = 0,$$

proving that α_M is one-to-one.

We leave the verification that α is natural to the reader. ■

As an immediate corollary of Proposition 10.1.7 and Theorem 10.1.9 we have:

Corollary 10.1.10. *For each multiplicative system S , $S^{-1}A$ is a flat A -module.*

Also, in this vein, we have the following exercises.

Exercise 10.1.11. Suppose that M and N are A -modules, and that S is a multiplicative system of A . There is a unique $S^{-1}A$ -isomorphism $\theta : S^{-1}M \otimes_{S^{-1}A} S^{-1}N \rightarrow S^{-1}(M \otimes_A N)$, such that $\theta(x/s \otimes y/t) = (x \otimes y)/st$.

Exercise 10.1.12. Suppose that S and T are multiplicative systems and that U is the image of T under the natural map $\phi : A \rightarrow S^{-1}A$. Prove that $(ST)^{-1}A$ is isomorphic to $U^{-1}(S^{-1}A)$. (Note: if $st = 0$ for some $s \in S$ and $t \in T$, then both of these rings are trivial. Deal with the nontrivial case. In that case, U is a multiplicative system of $S^{-1}A$.)

We close this section with a mention of the *classical* (or *total*) ring of fractions, which generalizes fields of fractions for integral domains.

Exercise 10.1.13. Let $\rho(A)$ denote the set of all regular elements of A . Then this is a multiplicative system. Let qA stand for $\rho(A)^{-1}A$. Prove the following:

- (a) The canonical map $A \rightarrow qA$ (by $a \mapsto a/1$) is one-to-one.
- (b) $\rho(A)$ is the largest multiplicative system S , for which the map $a \mapsto as/s$ is one-to-one.
- (c) In qA every element is either a unit or else a zero-divisor.
- (d) If B is a ring with the property that every element is either a unit or a zero-divisor, then $qB \cong B$, under the canonical $a \mapsto a/1$.

10.2 Properties of Localizations and Local Properties. The title of this section is, of course, a play on words. The first part deals with features of localizations, in general, and especially with regard to how the prime spectrum is affected. The second part of the section is devoted to so-called “local properties”, the kind that hold in a ring A if and only if each of the localizations of A has that property.

We begin by observing that, if \mathfrak{r} is an ideal of A , and S is a multiplicative system, then the extension \mathfrak{r}^e of \mathfrak{r} , as induced by the homomorphism $a \mapsto as/s$, is none other than $S^{-1}\mathfrak{r}$. Then here is a proposition with the barest essentials concerning extension and contraction of ideals, in the passage from A to $S^{-1}A$.

Proposition 10.2.1. (a) *Every ideal of $S^{-1}A$ is an extended ideal; that is, each ideal of $S^{-1}A$ is of the form $S^{-1}\mathfrak{r}$, for some ideal \mathfrak{r} of A .*

(b) *The ideal \mathfrak{r} of A is a contracted ideal if and only if every element of S is regular in A/\mathfrak{r} .*

Proof. (a) Suppose that \mathfrak{s} is an ideal of $S^{-1}A$; we’re to show that $\mathfrak{s} \leq \mathfrak{s}^{ce}$, since according to Proposition 9.1.10(iv), the reverse inclusion always holds. If $x/s \in \mathfrak{s}$ then $xs/s \in \mathfrak{s}$, which means that $x \in \mathfrak{s}^c$, from which it follows that $x/s \in \mathfrak{s}^{ce}$.

We leave (b) to the reader; (see pp. 41-42 of [AM69]). ■

Next, we have a general lemma, about the relationship between the prime spectra of a ring and one of its rings of fractions. We state it with some redundancy, on purpose.

Lemma 10.2.2. *Suppose that S is a multiplicative system of the ring A . The map $\mathfrak{p} \mapsto S^{-1}\mathfrak{p}$ defines a one-to-one correspondence from*

$$\{\mathfrak{p} \in \text{Spec}(A) : \mathfrak{p} \cap S = \emptyset\}$$

onto $\text{Spec}(S^{-1}A)$. Its inverse, the contraction, is a homeomorphism from $\text{Spec}(S^{-1}A)$ onto its image. In particular, contraction maps $\text{Min}(S^{-1}A)$ homeomorphically onto

$$\{\mathfrak{p} \in \text{Min}(A) : \mathfrak{p} \cap S = \emptyset\},$$

and $\text{Max}(S^{-1}A)$ homeomorphically onto

$$\{\mathfrak{p} \in \text{Spec}(A) : \mathfrak{p} \text{ is maximal with respect to } \mathfrak{p} \cap S = \emptyset\}.$$

Proof. It should be clear that, if \mathfrak{q} is a prime ideal of $S^{-1}A$, then

$$\mathfrak{q}^c = \{a \in A : as/s \in \mathfrak{q}, \text{ for some } s \in S\}$$

is a prime ideal of A , which does not meet S . (Remember: prime ideals are proper ideals!) In addition, since each ideal of $S^{-1}A$ is an extended ideal, $\mathfrak{q} = \mathfrak{q}^{ce}$, which implies that contraction is one-to-one. If $\mathfrak{p} \in \text{Spec}(A)$, and $\mathfrak{p} \cap S = \emptyset$, then, if $x \in \mathfrak{p}^{ec}$, we have that $xs/s \in \mathfrak{p}^e$ (for any $s \in S$), meaning that $xs/s = a/t$, for some $a \in \mathfrak{p}$ and $t \in S$. Then $u(xst - as) = 0$, for some $u \in S$, whence $xstu = asu \in \mathfrak{p}$. Since $stu \notin \mathfrak{p}$, it follows that $x \in \mathfrak{p}$, proving that $\mathfrak{p} = \mathfrak{p}^{ec}$, and, therefore, that contraction is onto the set of prime ideals of A which do not intersect S .

As to the topological matters, contraction is continuous, because $\text{Spec}(f)$ is continuous, for each **Crn1**-morphism f . Plus, if $\mathfrak{q} \in \text{Spec}(S^{-1}A)$, then $a/s \in \mathfrak{q}$ if and only if $a \in \mathfrak{q}^c$, which says exactly that the image of the basic closed set $V(a/s)$, under contraction, is

$$V(a) \cap \{\mathfrak{p} \in \text{Spec}(A) : \mathfrak{p} \cap S = \emptyset\}.$$

This completes the proof of the first paragraph.

We leave the rest to the reader. ■

Two corollaries, the proofs of which are left as exercises. For the first, the reader is referred to p. 42 of [AM69]. Part (b) should be straightforward.

Corollary 10.2.3. *Suppose that S is a multiplicative system of A . Then*

$$(a) \quad S^{-1}n(A) = n(S^{-1}A).$$

$$(b) \quad \text{For any two ideals } \mathfrak{r} \text{ and } \mathfrak{s} \text{ of } A, S^{-1}(\mathfrak{r}\mathfrak{s}) = (S^{-1}\mathfrak{r})(S^{-1}\mathfrak{s}).$$

Corollary 10.2.4. *Suppose \mathfrak{p} is a prime ideal of A . Then $\text{Spec}(A_{\mathfrak{p}})$ is homeomorphic to the subspace of $\text{Spec}(A)$ of all prime ideals that are contained in \mathfrak{p} .*

Proof. In Lemma 10.2.2 let $S = A \setminus \mathfrak{p}$. ■

At last, a definition which begs to be recorded.

Definition 10.2.5. A ring A with exactly one maximal ideal is said to be $\widehat{\text{local}}$. Corollary 10.2.4 tells us that each localization $A_{\mathfrak{p}}$ is a local ring.

Hot Air 10.2.6. Typical Play with Localization. Suppose that \mathfrak{p} and \mathfrak{q} are ideals of the ring A , with $\mathfrak{p} \geq \mathfrak{q}$. By passing to A/\mathfrak{q} , one eliminates all the ideals of A , except those that contain \mathfrak{q} . Form $(A/\mathfrak{q})_{\mathfrak{p}/\mathfrak{q}}$; this is isomorphic to $A_{\mathfrak{p}}/\mathfrak{q}_{\mathfrak{p}}$, by Corollary 10.1.8(c), as $A_{\mathfrak{p}}$ -modules, in the first instance. The reader should check that it is also a ring isomorphism. Then one is left with only the primes between $\mathfrak{p}_{\mathfrak{p}}$ and $\mathfrak{q}_{\mathfrak{p}}$.

Also note the following: A/\mathfrak{p} is an integral domain, with field of fractions $q(A/\mathfrak{p})$. Observe that $\mathfrak{p} + a \neq 0$ precisely when $a \notin \mathfrak{p}$. Therefore if $T = \rho(A/\mathfrak{p})$ – recall the notation of Exercise 10.1.13 – then

$$q(A/\mathfrak{p}) = T^{-1}(A/\mathfrak{p}) \cong (A/\mathfrak{p})_{\mathfrak{p}},$$

and invoking Corollary 10.1.8(c) once more, the latter is isomorphic to the field obtained by factoring out the unique maximal ideal of $A_{\mathfrak{p}}$. This field is called the *residue field of A at \mathfrak{p}* .

Examples 10.2.7. As explained above, any localization is local. In particular, for each prime number p , the ring $\mathbb{Q}_{(p)}$ of all rational numbers with denominator which is relatively prime to p , is a local ring.

Consider next the ring $F[[T]]$ of power series over a field F ; (see Exercise 9.1.6.) We leave it to the reader to check that it is a local ring. In fact ...

Exercise 10.2.8. For any field F , let $F[[T]]$ be the ring of power series over F . Show that $F[[T]]$ is an integral domain in which the only nonzero ideals are, for each nonnegative integer m ,

$$\mathfrak{t}_m = \left\{ \sum_n a_n T^n : a_n = 0, \forall n < m \right\}.$$

Note that $F[[T]] = \mathfrak{t}_0$. (In particular, the lattice of ideals forms a chain. We shall return to rings with this property in the next chapter.)

Exercise 10.2.9. Let A be a ring. If it is local, then the maximal ideal is the set of nonunits. Conversely, if the set \mathfrak{m} of nonunits of A forms an ideal, then it is maximal, and A is local with \mathfrak{m} as its maximal ideal.

We now turn to local properties. They play a prominent role in commutative algebra.

Definition 10.2.10. Let \mathfrak{P} be a property of rings. We say that it is a *local property* if \mathfrak{P} holds for a ring A if and only if it holds for each localization $A_{\mathfrak{q}}$ (all $\mathfrak{q} \in \text{Spec}(A)$).

Moreover, if $f : M \rightarrow N$ is a $\mathbf{A}\mathbf{Mod}$ -morphism, then we have the induced $f_{\mathfrak{q}} : M_{\mathfrak{q}} \rightarrow N_{\mathfrak{q}}$, which is a $\mathbf{A}_{\mathfrak{q}}\mathbf{Mod}$ -morphism. If \mathfrak{P} is a property of module homomorphisms, we say that \mathfrak{P} is a *local property* if f (an $\mathbf{A}\mathbf{Mod}$ -morphism) satisfies \mathfrak{P} precisely when $f_{\mathfrak{q}}$ does, for each prime ideal \mathfrak{q} of A .

One defines *local property* for modules in a similar manner.

Here is a starter for a list of local properties:

Proposition 10.2.11. *Being a trivial module is a local property. More precisely, the following are equivalent for an A -module M :*

- (a) $M = 0$;
- (b) $M_{\mathfrak{p}} = 0$, for each $\mathfrak{p} \in \text{Spec}(A)$;
- (c) $M_{\mathfrak{q}} = 0$, for each maximal ideal \mathfrak{q} of A .

Proof. Exercise, or else see [AM69], p. 40. ■

Proposition 10.2.12. *Being one-to-one is a local property of module homomorphisms. Again, more to the point, the following are equivalent:*

- (a) $f : X \rightarrow Y$ is a one-to-one $\mathbf{A}\mathbf{Mod}$ -morphism;
- (b) $f_{\mathfrak{p}}$ is one-to-one, for each prime ideal \mathfrak{p} of A ;
- (c) $f_{\mathfrak{m}}$ is one-to-one, for each maximal ideal \mathfrak{m} of A .

Likewise, being surjective is also local.

Proof. (a) implies (b) on account of Proposition 10.1.7, and (c) is obviously a consequence of (b).

Now assume (c), and let $X' = \ker(f)$; then

$$0 \rightarrow X' \rightarrow X \xrightarrow{f} Y$$

(where $X' \rightarrow X$ is inclusion) is exact, whence

$$0 \rightarrow X'_{\mathfrak{m}} \rightarrow X_{\mathfrak{m}} \xrightarrow{f_{\mathfrak{m}}} Y_{\mathfrak{m}}$$

is exact (Proposition 10.1.7 again). This means that $X'_{\mathfrak{m}} = 0$ (for each maximal ideal \mathfrak{m}), and so, by the preceding proposition, $X' = 0$. ■

Proposition 10.2.13. *Flatness of modules is a local property. These are equivalent statements:*

- (a) The A -module M is flat (over A).
- (b) $M_{\mathfrak{p}}$ is $A_{\mathfrak{p}}$ -flat, for each $\mathfrak{p} \in \text{Spec}(A)$.
- (c) $M_{\mathfrak{q}}$ is $A_{\mathfrak{q}}$ -flat, for each maximal ideal \mathfrak{q} of A .

Proof. Exercise, or else see [AM69], p. 41. ■

We close this section with an assortment of exercises.

Exercise 10.2.14. (a) If A is a von Neumann regular ring then so is $S^{-1}A$, for each multiplicative system S of A .

(b) A is von Neumann regular if and only if $A_{\mathfrak{m}}$ is a field, for each maximal ideal \mathfrak{m} of A . Thus, being von Neumann regular is a local property.

The next exercise links up once more with properties of spectra.

Exercise 10.2.15. First, a definition; if $f : A \rightarrow B$ is a ring homomorphism, B is an A -module (recall, by setting $a \cdot b = f(a)b$). If it is a flat A -module we say it is an A -flat algebra.

Now suppose that $f : A \rightarrow B$ is an A -flat algebra. Then the following are equivalent:

- (a) Each ideal of A is a contracted ideal.
- (b) $\text{Spec}(f) : \text{Spec}(B) \rightarrow \text{Spec}(A)$, that is to say, contraction, is surjective.
- (c) For every maximal ideal \mathfrak{m} of A , \mathfrak{m}^e is a proper ideal.
- (d) If X is a nonzero A -module, then $B \otimes_A X$ is nonzero.
- (e) For every A -module X , the homomorphism $y \mapsto 1 \otimes y$ into $B \otimes_A X$ is one-to-one.

If one (and therefore all) of these conditions hold, we say that B is *faithfully flat* over A . (For hints, see p. 46 of [AM69].)

10.3 Integral Extensions. The concept of being integral over a ring is the ring-theoretic cousin of being algebraic over a field in field theory. There are a number of subtle differences, however, which present rather special difficulties. That is the subject of this section.

Definition 10.3.1. Suppose that A is a subring of B , and that $y \in B$. We say that y is *integral over A* , if there exists a monic polynomial $f(T) \in A[T]$, such that $f(y) = 0$. If every element of B is integral over A , we say that B is *integral over A* , or an *integral extension* of A . Clearly, every element of A is integral over A .

Remarks 10.3.2. By the rational roots test for polynomials with integer coefficients it is seen that the only rational numbers which are integral over \mathbb{Z} are the integers themselves.

On the other hand, consider \mathbb{Z} as a subring of

$$\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}.$$

Then i is obviously integral over \mathbb{Z} , and then it is also easy to see that every Gaussian integer is integral over \mathbb{Z} . (In a moment we shall establish, anyway, that the set of all integral elements in an extension form a subring.) As we shall see later, the set of integral elements of $\mathbb{Q}[i]$ is in fact $\mathbb{Z}[i]$.

For field extensions, the notions of “algebraic over” and “integral over” (obviously) coincide. However, in general, the two concepts are different. Indeed, historically, the kind of integral extension given above arose by looking at a finite algebraic extension E of \mathbb{Q} , and identifying those elements which are integral over \mathbb{Z} . These are traditionally referred to as the *algebraic integers* of E . We shall return to these in the next chapter, when we discuss Dedekind domains, and again (with vigor) in the chapter on Galois theory.

The following lemma is technical, but has a number of practical consequences. The reader ought to keep in mind that, from this point on, we shall use the terms “ A is subring of B ” and “ B is an extension of A ” interchangeably.

Lemma 10.3.3. *Suppose that B is an extension of A , and $y \in B$. Then the following are equivalent.*

- (a) y is integral over A .
- (b) The subring of B generated by A and y , namely,

$$A[y] = \{ f(y) : f(T) \in A[T] \},$$

is a finitely generated A -module.

- (c) $A[y]$ is contained in a subring C of B , such that C is a finitely generated A -module.
- (d) There is a faithful $A[y]$ -module M which is finitely generated as an A -module.

Proof. (a) \Rightarrow (b): To say that y is integral over A , is to say that there are $a_0, a_1, \dots, a_{k-1} \in A$, so that

$$y^k + a_{k-1}y^{k-1} + \dots + a_1y + a_0 = 0.$$

Thus, y^k is an A -linear combination of the powers $1, y, \dots, y^{k-1}$. By induction (and substitution) it can then be shown that all the higher powers of y are A -linear combinations of $1, y, \dots, y^{k-1}$.

(b) \Rightarrow (c): Obvious; take $C = A[y]$.

(c) \Rightarrow (d): Let $M = C$, which is faithful over itself.

(d) \Rightarrow (a): Multiplication by y is an $\mathbf{A}\mathbf{Mod}$ -morphism of M into itself. Now let $\{x_1, x_2, \dots, x_m\}$ be a set of generators of M over A . Thus, each $yx_i = a_{i1}x_1 + \dots + a_{im}x_m$, for suitable $a_{ij} \in A$. Putting in another way,

$$\sum_{j=1}^m (\delta_{ij}y - a_{ij})x_j = 0, \quad \forall i = 1, \dots, m,$$

where δ_{ij} is the Kronecker-delta. Now multiply on the left by the adjoint of the matrix $yI - \alpha$ (with $\alpha = (a_{ij})$); this yields that

$$\det(yI - \alpha)x_i = 0, \quad \forall i = 1, \dots, m.$$

The determinant in question is a (monic) element in $A[y]$; therefore, since M is faithful over $A[y]$, it follows that $\det(yI - \alpha) = 0$, proving that y is integral over A . ■

Proposition 10.3.4. *Suppose that B is an extension of A . Then*

- (i) If $y_1, y_2, \dots, y_m \in B$ are all integral over A , then the ring $A[y_1, y_2, \dots, y_m]$ is a finitely generated A -module.
- (ii) The subset C of B consisting of all the elements which are integral over A , is a subring of B .
- (iii) If B is an integral extension of A and B_1 is an integral extension of B , then B_1 is an integral extension of A .

Proof. (i) By induction. The case $m = 1$ is covered by Lemma 10.3.3. Denote $A_r = A[y_1, y_2, \dots, y_r]$; by the inductive hypothesis A_{m-1} is a finitely generated A -module. Since y_m is integral over A_{m-1} , we have that A_m is a finitely generated A_{m-1} -module. But then it is easily verified that A_m is finitely generated over A .

(ii) Suppose that y and z are integral over A . As $y \pm z$ and yz belong to $A[y, z]$, they are all integral over A , by (i) and (c) of Lemma 10.3.3.

(iii) Suppose that B is integral over A , and B_1 is integral over B . If $y \in B_1$ then we have an equation $y^n + b_1y^{n-1} + \dots + b_n = 0$, for suitable $b_1, b_2, \dots, b_n \in B$. $A[b_1, \dots, b_n]$ is finitely generated over A , by (i), and $A[b_1, \dots, b_n, y]$ is finitely generated over $A[b_1, \dots, b_n]$, whence $A[b_1, \dots, b_n, y]$ is finitely generated over A . Once more by (c) of Lemma 10.3.3, it follows that y is integral over A . ■

Hot Air 10.3.5. *Finitary Properties.* Since various notions of finitary generation will begin to pop up in the material that follows here and in later chapters, let us make a cautionary distinction.

First, if $f : A \rightarrow B$ is any homomorphism (**Crn1**-morphism), then B may be regarded as an A -module, via the scalar multiplication $a \cdot b = f(a)b$. Any ring B which is a module over A is referred to as an A -algebra, provided that $a(bc) = (ab)c$, for all $a \in A$ and $b, c \in B$. This gives rise to possible consideration of the category of A -algebras and all A -homomorphisms (maps which simultaneously are ring homomorphisms and preserve the A -scalar structure).

With this setting in mind there are two notions of “finitely generated”, which ought to be clearly distinguished. Let B be an A -algebra. To say that B is *finitely generated as an A -module* means that there is a finite set $x_1, \dots, x_n \in B$, so that each $b \in B$ can be written as a A -linear combination of the x_i . In terms of universal algebra this means that B is a homomorphic image of A^n (for some n) in the category **A**Mod.

To say that B is *finitely generated as an A -algebra* means that there is a finite set $x_1, \dots, x_n \in B$, so that each $b \in B$ is a polynomial in the x_i . In categorical terms this means that B is a homomorphic image of the polynomial algebra $A[T_1, \dots, T_n]$, for some n , in the category of A -algebras.

Next, here is the notion of integral closure.

Definition 10.3.6. If B is an extension of A , and the only elements of B which are integral over A are the elements of A , we say that A is *integrally closed* in B . A is said to be *integrally closed* (without qualification) if it is integrally closed in its classical ring of fractions qA .

In view of Proposition 10.3.4, we may speak of the *integral closure of A in the extension B* , as the set of all elements of B , that are integral over A . We denote this subring of B by $i_B(A)$. Owing to Proposition 10.3.4(iii), $i_B(A)$ is integrally closed in B .

Example 10.3.7. Consider \mathbb{Z} as a subring of $\mathbb{Q}[i]$. By Proposition 10.3.4, the integral closure of \mathbb{Z} in $\mathbb{Q}[i]$ must contain $\mathbb{Z}[i]$. One may well wonder if there are any more integral elements in $\mathbb{Q}[i]$. If we knew that $\mathbb{Z}[i]$ were integrally closed ($\mathbb{Q}[i]$ is its field of fractions), we would know that $\mathbb{Z}[i]$ was the integral closure of \mathbb{Z} in $\mathbb{Q}[i]$.

Much of what lies ahead in this section deals with integral domains and how one recognizes the ones that are integrally closed.

As a preliminary, here’s an exercise, the solution of which is, in essence, the proof of the rational roots test for a polynomial with integer coefficients.

Exercise 10.3.8. If A is a unique factorization domain then it is integrally closed.

(Hint: if there is a fraction a/b which is integral over A , which is not in A , then some irreducible element p divides b , but not a .)

The following proposition is useful in the sequel. We leave the proof as an exercise to the reader. (Or else, see [AM69], p. 61.)

Proposition 10.3.9. *Suppose that B is an integral extension of A . Then*

- (a) *If \mathfrak{s} is an ideal of B and \mathfrak{r} is its contraction – that is, $\mathfrak{r} = \mathfrak{s} \cap A$, then B/\mathfrak{s} is integral over A/\mathfrak{r} .*
- (b) *If S is a multiplicative system of A , then $S^{-1}B$ is integral over $S^{-1}A$.*

Our first goal now is the so-called Going Up Theorem. We reach that goal through a number of preliminaries.

Proposition 10.3.10. *Suppose that A is a subring of the integral domain B , so that B is an integral extension of A . Then B is a field if and only if A is one.*

Proof. Suppose that A is a field. Suppose that $y \in B$, $y \neq 0$, and that

$$y^n + a_1y^{n-1} + \dots + a_{n-1}y + a_n = 0$$

is a polynomial equation of least degree, which witnesses that y is integral over A . Since B is an integral domain $a_n \neq 0$. Then one readily verifies that

$$y^{-1} = -a_n^{-1}(y^{n-1} + a_1y^{n-2} + \dots + a_{n-1}),$$

which is in B . Thus, B is a field.

We sketch the converse; the reader should look at [AM69], p. 61. Suppose that B is a field. Then, if $x \in A$, with $x \neq 0$, we have that $x^{-1} \in B$, and therefore satisfies some monic polynomial over A . After a bit of computation, similar to that of the first part, one can observe that $x^{-1} \in A$. ■

The Going Up Theorem will say something about how primes contract when an extension is integral. The next proposition says that under contraction maximal ideals correspond (part (a)), and that you cannot collapse chains by contraction (part (b)).

Proposition 10.3.11. *Suppose that B is integral over A .*

- (a) *Suppose that $\mathfrak{q} \in \text{Spec}(B)$ and $\mathfrak{p} = \mathfrak{q}^c$. Then \mathfrak{q} is maximal if and only if \mathfrak{p} is maximal (in A).*
- (b) *Suppose that $\mathfrak{q} \leq \mathfrak{q}'$ are prime ideals of B , such that $\mathfrak{p} = \mathfrak{q}^c = \mathfrak{q}'^c$. Then $\mathfrak{q} = \mathfrak{q}'$.*

Proof. By Proposition 10.3.9(a), B/\mathfrak{q} is integral over A/\mathfrak{p} , and, of course, both are integral domains. Now apply Proposition 10.3.10. This proves (a).

According to Proposition 10.3.9(b), $B_{\mathfrak{p}}$ is integral over $A_{\mathfrak{p}}$. Denote the extension of \mathfrak{p} in $A_{\mathfrak{p}}$ by \mathfrak{m} , and the extensions of \mathfrak{q} and \mathfrak{q}' in $B_{\mathfrak{p}}$ by \mathfrak{n} and \mathfrak{n}' , respectively. Then $\mathfrak{n} \leq \mathfrak{n}'$, and both contract to \mathfrak{m} (by Corollary 10.1.8). However, \mathfrak{m} is maximal in $A_{\mathfrak{p}}$, which implies that $\mathfrak{n} = \mathfrak{n}'$. By Lemma 10.2.2, it follows that $\mathfrak{q} = \mathfrak{q}'$. ■

Next, we have this preliminary version of the Going Up Theorem.

Proposition 10.3.12. *Suppose that B is an integral extension of A . If $\mathfrak{p} \in \text{Spec}(A)$, then there is a prime ideal \mathfrak{q} of B which contracts to \mathfrak{p} ($\mathfrak{q}^c = \mathfrak{p}$).*

Proof. First (by 10.1.8(b)) $B_{\mathfrak{p}}$ is integral over $A_{\mathfrak{p}}$. Keep in mind the following commutative diagram, where j is the inclusion; recall that forming rings of fractions preserves exactness, whence $j_{\mathfrak{p}}$ is one-to-one. The vertical maps are the canonical homomorphisms $x \mapsto x/1$.

$$\begin{array}{ccc}
 A & \xrightarrow{j} & B \\
 \downarrow & & \downarrow \\
 A_{\mathfrak{p}} & \xrightarrow{j_{\mathfrak{p}}} & B_{\mathfrak{p}}
 \end{array}$$

If \mathfrak{n} is a maximal ideal of $B_{\mathfrak{p}}$, then $\mathfrak{n} \cap A_{\mathfrak{p}}$ is maximal, owing to Proposition 10.3.11(a). Thus $\mathfrak{m} = \mathfrak{n} \cap A_{\mathfrak{p}}$ is the unique maximal ideal of $A_{\mathfrak{p}}$. This means that the contraction of \mathfrak{n} to A is \mathfrak{p} . But note as well that if \mathfrak{q} denotes the contraction of \mathfrak{n} to B , then $\mathfrak{q} \cap A = \mathfrak{p}$. ■

Theorem 10.3.13. The Going Up Theorem. *Suppose that B is an integral extension of A . Suppose that*

$$\mathfrak{p}_1 < \mathfrak{p}_2 < \cdots < \mathfrak{p}_n$$

is a chain of prime ideals of A , and that

$$\mathfrak{q}_1 < \cdots < \mathfrak{q}_m \quad (m < n)$$

are prime ideals of B , so that $\mathfrak{q}_i^c = \mathfrak{p}_i$, for each $i = 1, \dots, m$. Then the chain $\mathfrak{q}_1 < \cdots < \mathfrak{q}_m$ can be extended to a chain $\mathfrak{q}_1 < \cdots < \mathfrak{q}_n$, so that each \mathfrak{q}_i contracts to \mathfrak{p}_i .

Proof. Leaving the induction arguments involved to the reader, we shall do the case $m = 1$ and $n = 2$.

So suppose that $\mathfrak{p} = \mathfrak{q}^c$, and that $\mathfrak{p} < \mathfrak{p}'$, with $\mathfrak{q} \in \text{Spec}(B)$ and $\mathfrak{p}, \mathfrak{p}' \in \text{Spec}(A)$. By Proposition 10.3.11(a), B/\mathfrak{q} is integral over A/\mathfrak{p} , and $\mathfrak{p}'/\mathfrak{p}$ is a prime ideal of A/\mathfrak{p} . Thus, by Proposition 10.3.12, there is a prime ideal of B/\mathfrak{q} (necessarily of the form $\mathfrak{q}'/\mathfrak{q}$, where \mathfrak{q}' is prime in B , such that $\mathfrak{q}'/\mathfrak{q} \cap A/\mathfrak{p} = \mathfrak{p}'/\mathfrak{p}$. This means that $\mathfrak{q}' \cap A = \mathfrak{p}'$, and we're done. ■

Next, for integral domains, we have the reverse (so to speak) of Theorem 10.3.13, namely the Going Down Theorem. First, we have the following improvement on Proposition 10.3.11(b). We leave the proof to the reader, or else one can consult [AM69], p. 62.

Proposition 10.3.14. *Suppose that B is an extension of A , and that C is the integral closure of A in B . For each multiplicative system S of A , $S^{-1}C$ is the integral closure of $S^{-1}A$ in $S^{-1}B$.*

As a consequence of Proposition 10.3.14 we obtain the following.

Theorem 10.3.15. *Being integrally closed is a local property, for integral domains. More precisely, if A is an integral domain, then the following are equivalent.*

- (a) A is integrally closed.
- (b) $A_{\mathfrak{p}}$ is integrally closed, for each prime ideal \mathfrak{p} of A .
- (c) $A_{\mathfrak{m}}$ is integrally closed, for each $\mathfrak{m} \in \text{Max}(A)$.

Proof. Recall that qA denotes the field of fractions of A . Suppose that iA is the integral closure of A in qA . The statement “ A is integrally closed” is equivalent to “the inclusion $j : A \rightarrow iA$ is surjective”. Therefore (by Proposition 10.3.14), $A_{\mathfrak{p}}$ (resp. $A_{\mathfrak{m}}$) with $\mathfrak{p} \in \text{Spec}(A)$ (resp. $\mathfrak{m} \in \text{Max}(A)$) is integrally closed if and only if the inclusion $A_{\mathfrak{p}} \leq (iA)_{\mathfrak{p}}$ (resp. $A_{\mathfrak{m}} \leq (iA)_{\mathfrak{m}}$) is surjective. Now apply Proposition 10.2.12. ■

Definition 10.3.16. Suppose that B is an extension of A , and \mathfrak{r} is ideal of A . An element $b \in B$ is \mathfrak{r} -integral if b satisfies a monic polynomial $f(T)$ whose lower-term coefficients belong to \mathfrak{r} . The \mathfrak{r} -integral closure in B is the set of all \mathfrak{r} -integral elements of B .

The next result should not be terribly surprising.

Lemma 10.3.17. Assume that B is an integral domain, and suppose that B is an extension of A , and C is the integral closure of A in B . Let \mathfrak{r} be an ideal of A and \mathfrak{r}^e be its extension in C . The \mathfrak{r} -integral closure in B is $\sqrt{\mathfrak{r}^e}$.

Proof. Suppose that $y \in B$ is \mathfrak{r} -integral, and choose $a_1, a_2, \dots, a_k \in \mathfrak{r}$, so that $y^k + a_1y^{k-1} + \dots + a_k = 0$. Then $y \in C$, and it should be easy to see that $y^k \in \mathfrak{r}^e$, whence $y \in \sqrt{\mathfrak{r}^e}$.

Conversely, if $z \in \sqrt{\mathfrak{r}^e}$, then $z^n = c_1a_1 + \dots + c_ma_m$, with each $c_i \in C$ and $a_i \in \mathfrak{r}$. According to Proposition 10.3.4(i), we have that $M = A[c_1, \dots, c_m]$ is a finitely generated A -module. Observe that $z^nM \leq \mathfrak{r}M$. Now, by imitating the proof that (d) of Lemma 10.3.3 implies (a), we can show that z^n , and therefore z , is \mathfrak{r} -integral.

(Note: no need to worry about the “faithful” part of the argument in Lemma 10.3.3, as we are dealing with integral domains.) ■

Hot Air 10.3.18. *Galois Theory: an Early Sighting.* In the following consequence of Lemma 10.3.17, “Galois groups” rear their heads, albeit innocently. We shall return to the Galois Theory in Chapter 14.

Recall this: if K is a field, and a subfield of L , with $t \in L$ algebraic over K , let $p(T) \in K[T]$ be the minimum polynomial of t over K . Then under any automorphism α of any field extension E of K , which fixes the elements of K , $\alpha(t)$ must again be a root of $p(T)$. Therefore, under any such automorphism α , there are finitely many images of t possible.

You should also recall from elementary field theory, that if s is any root of $p(T)$, then there is an automorphism of the splitting field L of $p(T)$, which maps t to s . Point is: all the roots of $p(T)$ are images under some automorphism of its splitting field.

If you want a fancy description: the *Galois group* $G(L/K)$, of all automorphisms of L that fix every element of K , acts as a transitive permutation group on the roots of the minimum polynomial of t .

Proposition 10.3.19. Suppose that B is an integral domain, and an extension of A , which is integrally closed. Suppose that \mathfrak{r} is an ideal of A , and $y \in B$ is \mathfrak{r} -integral. Then y is algebraic over qA , and if $p(T) \in qA[T]$ is the minimum polynomial of y over qA , then all the lower-term coefficients of $p(T)$ belong to $\sqrt{\mathfrak{r}}$.

Proof. As y is \mathfrak{r} -integral, it satisfies an equation of the form

$$(*) \quad y^n + a_1y^{n-1} + \dots + a_n = 0,$$

with each $a_i \in \mathfrak{r}$. Let L be a field extension of qA , which contains all the roots y_1, \dots, y_k of the minimum polynomial of y over qA . Then each y_j satisfies (*), and is therefore \mathfrak{r} -integral. Since the coefficients of the minimum polynomial of y over qA are polynomials in the roots, it follows that each such coefficient is \mathfrak{r} -integral, and, by Lemma 10.3.17, an element of $\sqrt{\mathfrak{r}}$. ■

Theorem 10.3.20. The Going Down Theorem. Suppose that B is an integral domain, which is integrally closed in its field of fractions and an integral extension of A . Suppose that

$$\mathfrak{p}_1 > \mathfrak{p}_2 > \cdots > \mathfrak{p}_n$$

is a chain of prime ideals of A , and that

$$\mathfrak{q}_1 > \mathfrak{q}_2 > \cdots > \mathfrak{q}_m$$

are prime ideals of B ($m < n$), so that $\mathfrak{p}_i = \mathfrak{q}_i^c$, for each $i = 1, 2, \dots, m$. Then the chain $\mathfrak{q}_1 > \cdots > \mathfrak{q}_m$ can be extended to $\mathfrak{q}_1 > \cdots > \mathfrak{q}_n$, in $\text{Spec}(B)$, so that each \mathfrak{q}_i contracts to \mathfrak{p}_i .

Proof. As with the Going Up Theorem, we stick to the $n = 2$ and $m = 1$ case, leaving the induction to the reader. So, we relabel: $\mathfrak{p} > \mathfrak{p}'$ are prime ideals of A , $\mathfrak{q} \in \text{Spec}(B)$, and $\mathfrak{q} \cap A = \mathfrak{p}$. We localize B to $B_{\mathfrak{q}}$, and it is sufficient to show that $B_{\mathfrak{q}}\mathfrak{p}'$ contracts to \mathfrak{p}' , by Lemma 10.2.2.

Now pick $x \in B_{\mathfrak{q}}\mathfrak{p}'$; it is of the form $x = a/s$, with $a \in B_{\mathfrak{q}}\mathfrak{p}'$, and $s \notin B_{\mathfrak{q}} \setminus \mathfrak{q}$. According to Lemma 10.3.17 (letting $B = C$), a is \mathfrak{p}' -integral, and so there exist $c_1, c_2, \dots, c_k \in \mathfrak{p}'$, such that $a^k + c_1 a^{k-1} + \cdots + c_k = 0$. By Proposition 10.3.19, we may assume, without loss of generality, that the polynomial $p(T) = T^k + c_1 T^{k-1} + \cdots + c_k$ is indeed the minimum polynomial of a over qA .

If $x \in B_{\mathfrak{q}}\mathfrak{p}' \cap A$, then $s = ax^{-1} \in qA$. This means that the minimum polynomial of s over qA is of the form

$$q(T) = T^k + d_1 T^{k-1} + \cdots + d_k,$$

where $d_i = c_i/x^i$. Thus, $c_i = d_i x^i \in \mathfrak{p}'$ ($i = 1, \dots, k$). But s is integral over A , and so, by Proposition 10.3.19 (taking $I = A$), each $d_i \in A$. If $x \notin \mathfrak{p}'$ then $d_i \in \mathfrak{p}'$, and (using 10.3.19 again) $s \in B_{\mathfrak{q}}\mathfrak{p}' \leq B_{\mathfrak{q}}\mathfrak{p} \leq \mathfrak{q}$, a contradiction. Hence $x \in \mathfrak{p}'$, showing that $B_{\mathfrak{q}}\mathfrak{p}' \cap A = \mathfrak{p}'$. ■

We conclude with an assortment of comments and exercises. The first introduces Krull dimension.

Definition & Remarks 10.3.21. Let A be a ring. If there exists a nonnegative integer n and a chain of prime ideals

$$\mathfrak{p}_1 < \mathfrak{p}_2 < \cdots < \mathfrak{p}_{n+1},$$

but no longer chains of prime ideals, we say that A has *Krull dimension* n . If no such n exists, we say that A has *infinite Krull dimension*. We denote the Krull dimension of A by $\text{Kd}(A)$.

Note that the semiprime rings of Krull dimension 0 are the von Neumann regular rings; it is for this reason that the von Neumann regular rings are referred to by some authors as *zero-dimensional rings*. Recall that in a principal ideal domain (PID) every nonzero prime ideal is maximal; this says that a PID A has $\text{Kd}(A) = 1$.

The significance of the Going Up and Going Down Theorems is that if A is an integrally closed integral domain, and B is an integral domain, which is an integral extension of A , then $\text{Kd}(A) = \text{Kd}(B)$.

Exercise 10.3.22. Suppose that B is an integral extension of A .

- (a) Show that if $a \in A$ is invertible in B , then a is a unit in A .
- (b) Show that $\mathfrak{J}(B)^c = \mathfrak{J}(A)$.

The result in the following exercise could be interpreted by saying that in the category **Crn1**, the algebraically closed fields are injective relative to integral extension.

Exercise 10.3.23. Suppose that B is an integral extension of A . If L is an algebraically closed field, and $f : A \rightarrow L$ is a homomorphism then there is an extension of f to B .

We end with a famous result of Noether's, a standard preliminary piece in any preparation to do algebraic geometry. We restrict ourselves to infinite fields. As pointed out in [AM69], for finite fields a different proof is needed. We outline the proof; it is the same as the sketch given in [AM69], on p. 69.

Theorem 10.3.24. Noether's Normalization Lemma. *Suppose that K is an infinite field and $A \neq 0$ is a finitely generated K -algebra. Then there exist $y_1, y_2, \dots, y_r \in A$, which are algebraically independent over K , and such that A is an integral extension of $K[y_1, \dots, y_r]$.*

Proof. (And do supply the details!) Suppose that $\{x_1, \dots, x_n\}$ generates A as a K -algebra. Arrange these such that $\{x_1, \dots, x_r\}$ are algebraically independent over K , and each x_{r+1}, \dots, x_n is algebraic over $K[x_1, \dots, x_r]$. (Note: r could be 1.) We do induction on n , noting that if $n = r$ there is nothing to prove. So suppose $n > r$, and the theorem is true for $n - 1$ generators.

Now x_n is algebraic over $K[x_1, \dots, x_{n-1}]$, and so a polynomial $f \neq 0$ exists in n variables such that $f(x_1, \dots, x_{n-1}, x_n) = 0$. Let g denote the homogeneous part of f of highest degree. Since K is infinite, there exist $z_1, \dots, z_{n-1} \in K$ such that $g(z_1, \dots, z_{n-1}, 1) \neq 0$. Define $x'_i = x_i - z_i x_n$, for $i = 1, 2, \dots, n - 1$. Then show that x_n is integral over $A' = K[x'_1, \dots, x'_{n-1}]$, and therefore integral over A .

Then apply the induction hypothesis. ■

In the next chapter we turn to valuation ring theory, an indispensable tool in modern commutative algebra, and a beautiful theory besides.

11. Commutative Algebra: Valuation Theory

This chapter presents a general account of valuation rings, with two particular highlights: a discussion of valuations and value groups; and a brief account of the p -adic integers. All rings are assumed to be commutative, with identity. If there's no reference to the contrary, all categorical matters are assumed to pertain to **Crn1**. Our main reference remains [AM69], specifically Chapter 5. In the discussion of p -adic integers, the reader will see additional references.

11.1 Valuation Rings. Note that all rings in this section (unless the contrary is expressly signalled) will be integral domains. If A is an integral domain, recall that qA denotes its field of fractions. The term "overring", which we shall use from time to time, here and elsewhere, refers to an extension of A inside qA .

Definition 11.1.1. Let A be an integral domain. A is a *valuation ring* if the lattice of ideals of A is a chain. The name "valuation ring" comes from the theory of valuations and will be explained in due course. The main examples of valuation rings are the localizations of \mathbb{Z} at any prime p , $\mathbb{Q}_{(p)}$, and the ring of formal power series $F[[T]]$ over a field F . (See Exercise **10.2.8**.)

The following captures the main features of valuation rings.

Proposition 11.1.2. *Let A be an integral domain.*

- (a) *Any valuation ring is a local ring.*
- (b) *The following are equivalent:*
 - (i) *A is a valuation ring;*
 - (ii) *for each pair $x, y \in A$, either y is a multiple of x , or the reverse;*
 - (iii) *if $y \in qA$, then either y or y^{-1} belong to A .*
- (c) *Any overring of a valuation ring is a valuation ring.*
- (d) *A valuation ring is integrally closed.*

Proof. (a) is obvious from the definition. As to (b), suppose that A is a valuation ring; if $x, y \in A$, apply the definition to the principal ideals Ax and Ay ; thus, (i) implies (ii). If (ii) holds and $y = a/b$ belongs to qA , then if b divides a , $y \in A$, if a divides b , the reverse. So (ii) implies (iii). Finally, if (iii) is true, and \mathfrak{r} and \mathfrak{s} are ideals of A , with $\mathfrak{r} \not\leq \mathfrak{s}$, then pick $y \in \mathfrak{r} \setminus \mathfrak{s}$, and let $a \in \mathfrak{s}$. If $y/a \in A$ then $y \in \mathfrak{s}$ is forced, a contradiction. Therefore, $a/y \in A$, which means that a is a multiple of y , whence $a \in \mathfrak{r}$. Hence, $\mathfrak{s} \leq \mathfrak{r}$, and we're done with (b).

(c) follows immediately from (iii) in (b).

(d) Suppose that A is a valuation ring, and that $x \in qA$ is integral over A . If $x \in A$, we're happy. If $x^{-1} \in A$ then we have an identity: $x^n + a_1x^{n-1} + \cdots + a_n = 0$, with suitable $a_i \in A$. Multiply the equation by the inverse of x^{n-1} and get:

$$x = -(a_1 + a_2x^{-1} \cdots + a_nx^{-n+1}),$$

which lies in A . Thus, A is integrally closed. ■

One of our goals in this section is to show that the integral closure of an integral domain A is the intersection of all its overrings which are valuation rings. To that end, we need some preliminaries.

Hot Air 11.1.3. *More Zorn's Lemma.* Suppose that K is any field, and L any algebraically closed field. Let \mathcal{S} be the set of all pairs (A, f) , where A is a subring of K , and $f : A \rightarrow L$ is a homomorphism. We declare $(A, f) \leq (A', f')$ to mean that $A \leq A'$ and that the map f' extends f .

We shall let the reader verify that \mathcal{S} satisfies the conditions for Zorn's Lemma. What we wish to show is that if (B, g) is a maximal element, then B is a valuation ring.

Lemma 11.1.4. *With the notation of 11.1.3, if (B, g) is a maximal member of \mathcal{S} , then B is local and $\mathfrak{m} = \ker(g)$ is its maximal ideal.*

Proof. Since g maps into a field it should be clear that \mathfrak{m} is a prime ideal of B . Moreover, g can be extended to $B_{\mathfrak{m}}$ by defining $\tilde{g} : B_{\mathfrak{m}} \rightarrow L$ by $\tilde{g}(a/s) = g(a)/g(s)$, for all $a \in A$, $s \in A \setminus \mathfrak{m}$. Maximality of the pair (B, g) in \mathcal{S} then implies that $B = B_{\mathfrak{m}}$, whence B is local and \mathfrak{m} is its maximal ideal. ■

More technical stuff, crucial to the proceedings. For any ring R , we use $R^{\#}$ for the set of nonzero elements.

Lemma 11.1.5. *Suppose (B, g) is as in Lemma 11.1.4. If $x \in K^{\#}$ and $B[x]$ denotes the subring of K generated by B and x , then either $\mathfrak{m}[x] \neq B[x]$ or $\mathfrak{m}[x^{-1}] \neq B[x^{-1}]$, where $\mathfrak{m}[x]$ stands for the extension of \mathfrak{m} in $B[x]$.*

Proof. By way of contradiction, suppose that \mathfrak{m} generates the whole ring in both $B[x]$ and $B[x^{-1}]$. Then we have polynomial expressions, with coefficients in \mathfrak{m} ,

$$(\dagger) \quad u_0 + u_1x + \cdots + u_mx^m = 1,$$

and

$$(\ddagger) \quad v_0 + v_1x^{-1} + \cdots + v_nx^{-n} = 1.$$

We assume that the degrees of the above equations are as small as possible, and without loss of generality assume that $m \geq n$. Now multiply (\ddagger) by x^n ; this gets you:

$$(1 - v_0)x^n = v_1x^{n-1} + \cdots + v_n.$$

Since B is local and $v_0 \in \mathfrak{m}$, $1 - v_0$ is a unit in B , and we may rewrite the preceding equation (with suitable $w_1, \dots, w_n \in \mathfrak{m}$) as

$$x^n = w_1x^{n-1} + \cdots + w_n.$$

Replacing x^m in (\dagger) by $w_1x^{m-1} + \cdots + w_nx^{m-n}$, one obtains a polynomial identity in x with coefficients in \mathfrak{m} of lesser degree. ■

Proposition 11.1.6. *Let A be an integral domain and a subring of the field K . In \mathcal{S} any maximal element is a valuation ring, and K is its field of fractions.*

(If A is a valuation ring and K is its field of fractions, then we say that A is a *valuation ring* of K .)

Proof. Suppose that (B, g) is a maximal element of \mathcal{S} . We prove that if $x \in K$, then either x or x^{-1} belongs to B . That is sufficient, according to Proposition 11.1.2(b)(iii).

By Lemma 11.1.5, we may assume that the extension of \mathfrak{m} in $B[x]$ is a proper ideal; it is then contained in a maximal ideal \mathfrak{m}' of $B[x]$. Note that \mathfrak{m}' must contract to \mathfrak{m} , as $\mathfrak{m}' \cap B$ is a proper ideal containing \mathfrak{m} . Let $F = B/\mathfrak{m}$ and $F' = B[x]/\mathfrak{m}'$; observe that F is (isomorphic to) a subfield of L , and that F' is an extension of F . In fact, $F' = F[\mathfrak{m}' + x]$. Since \mathfrak{m}' contains the extension of \mathfrak{m} in

$B[x]$, $\mathfrak{m}' + x$ is algebraic over F . Then the inclusion of F in L can be extended to a homomorphism of F' into L , which is necessarily one-to-one. Let's call this embedding $j : F' \rightarrow L$. Then the homomorphism $g' : B[x] \rightarrow L$, defined by $g'(b) = j(\mathfrak{m}' + b)$, in the pair $(B[x], g')$, violates the maximality of (B, g) , unless $B = B[x]$; that is, $x \in B$. ■

We've actually proved more than we set out to do:

Theorem 11.1.7. *Suppose that A is a subring of the field K . Then the integral closure of A in K , $i_K(A)$, is the intersection of all the valuation subrings of K which contain A .*

Proof. Suppose that $A \leq V$ is a valuation subring of K . Since V is integrally closed (Proposition 11.1.2(d)), $i_K(A) \leq V$.

Conversely, suppose that $x \in K$ is not integral over A . This means that $x \notin A[x^{-1}]$, and therefore x^{-1} is a nonunit in $A[x^{-1}]$. Thus, x^{-1} is contained in a maximal ideal \mathfrak{m} of $A[x^{-1}]$. Let L be the algebraic closure of $A[x^{-1}]/\mathfrak{m}$, and $\theta : A[x^{-1}] \rightarrow L$ be the canonical homomorphism. Extend $(A[x^{-1}], \theta)$ to a maximal pair (B, g) , as in done in earlier results of this section; Proposition 11.1.6 assures us that B is a valuation ring of K , and since x^{-1} lies in $\ker(g)$, x cannot be in B .

This completes the proof. ■

Although this mention is a little out of place here, one can, with a little extra work, get at a preliminary form of the Hilbert Nullstellensatz. We set the steps up in two exercises. In mulling over the first one, refer perhaps to Exercise 10.3.23.

Exercise 11.1.8. Suppose that B is an integral domain and an extension of A , so that B is finitely generated (as an A -algebra) over A . Suppose that $v \neq 0$ is in B . Then there exists an element $u \neq 0$ in A with the following property: if $f : A \rightarrow L$ is a homomorphism into an algebraically closed field L , with $f(u) \neq 0$, then there exists an extension $g : B \rightarrow L$ of f , such that $g(v) \neq 0$.

(Hint: By induction one can reduce to the case $B = A[x]$. Next, treat the cases “ x is transcendental over qA ” and “ x is algebraic over qA ” separately. In the latter case, if x is algebraic over qA , then so is v^{-1} , because v is a polynomial in x over A . This leads to two equations (each of least possible degree)

$$(*) \quad a_0x^m + a_1x^{m-1} + \cdots + a_m = 0 \quad (\text{with each } a_i \in A),$$

and

$$(**) \quad c_0v^{-n} + c_1v^{1-n} + \cdots + c_n = 0 \quad (\text{with each } c_j \in A).$$

Now let $u = a_0c_0$. If $f : A \rightarrow L$ (with L an algebraically closed field) is a homomorphism, and $f(u) \neq 0$, then f can be extended to a homomorphism $f_1 : A[u^{-1}] \rightarrow L$ by setting

$$f_1(u^{-1}) = f(u)^{-1}$$

(because u is not in $\ker(f)$). Then, by Proposition 11.1.6, extend f_1 to a homomorphism $g : C \rightarrow L$ where C is a valuation ring containing $A[u^{-1}]$. From $(*)$, x is integral over $A[u^{-1}]$, which eventually says (why?) that $v \in C$. But v^{-1} is also integral over $A[u^{-1}]$, whence $v^{-1} \in C$. Conclude that $g(v) \neq 0$. Since $B \leq C$ (why?), we may restrict g to B , and the proof is complete.)

Exercise 11.1.9. Suppose that the field L is an extension of the field K , and a finitely generated K -algebra. Then L is a finite algebraic extension of K .

Here's a weak form of the Hilbert Nullstellensatz.

Exercise 11.1.10. Suppose that L is an algebraically closed field. Let A be a finitely generated L -algebra and \mathfrak{m} be a maximal ideal of A . Then $A/\mathfrak{m} \cong L$.

11.2 Valuations and Value Groups. We now direct our attention to the “valuation” aspect of valuation theory. The reader should check out the exercises on p.72, [AM69].

Historically, the valuation maps came first; indeed, the motivation is the ordinary absolute value of reals, or of normed vector spaces. These valuations simply have very striking logarithmic properties. Still, we get in by the back door; let’s first look at the so-called value groups, which are the targets of the valuation maps.

Definition & Remarks 11.2.1. Suppose that A is an integral domain and that U is its (multiplicative) group of units. Let $\text{VG}(A)$ stand for the factor group $qA^\# / U$. $\text{VG}(A)$ is called the *value group* of A . We place a partial order on $\text{VG}(A)$, and to study properties of that ordering, when A is a valuation ring.

Define $Ux \geq Uy$ to mean that $xy^{-1} \in A$; that is to say, y is a factor of x in A . Verify that this definition is independent of the representatives. This relation is evidently reflexive and transitive. If $Ux \geq Uy$ and $Ux \leq Uy$, then both $xy^{-1} \in A$ and $yx^{-1} \in A$, which makes xy^{-1} a unit of A ; that is to say, $Ux = Uy$. So we have a partial order indeed. Verify that \geq has the additional compatibility feature: if $Ux \geq Uy$ then $Uxz \geq Uyz$, for all $z \in qA^\#$.

A group G which is partially ordered, and such that $a \geq b$ implies that $ag \geq bg$ and $ga \geq gb$, for all $g \in G$, is called a *partially ordered group* (or *po-group*, for short).

Now let $v : qA^\# \rightarrow \text{VG}(A)$ be the canonical homomorphism. The following properties are satisfied:

- (a) $v(xy) = v(x)v(y)$, for all $x, y \in qA^\#$;
- (b) $v(x + y) \geq Ua$, for all $Ua \leq Ux, Uy$.

Let’s prove (b), (a) being the homomorphic property. If $Ux \geq Ua$ and $Uy \geq Ua$, then both xa^{-1} and ya^{-1} lie in A , which means that $(x + y)a^{-1} \in A$, whence $U(x + y) \geq Ua$.

For valuation rings we have

Proposition 11.2.2. *Let A be an integral domain. Then $\text{VG}(A)$ is totally ordered if and only if A is a valuation ring. In this event, (b) can be focussed to*

- (b’) $v(x + y) \geq \min\{v(x), v(y)\}$, for all $x, y \in qA^\#$.

Proof. Suppose that $x, y \in qA^\#$. If A is a valuation ring, then either xy^{-1} or yx^{-1} is in A , whence $Ux \geq Uy$ or else $Uy \geq Ux$ holds. The converse should be clear. ■

The next proposition gives us an idea about how to manufacture valuation rings. All groups appearing in the rest of this section are assumed to be abelian.

Proposition 11.2.3. *Suppose that K is any field and G any additive totally ordered group. Suppose that $w : K^\# \rightarrow G$ is a function which satisfies*

- (i) $w(xy) = w(x) + w(y)$, $\forall x, y \in K^\#$, and
- (ii) $w(x + y) \geq \min\{w(x), w(y)\}$, $\forall x, y \in K^\#$.

Then

$$A \equiv \{x \in K^\# : w(x) \geq 0\} \cup \{0\}$$

is a valuation ring of K , and there is an order-preserving isomorphism $\psi : \text{VG}(A) \rightarrow w(K^\#)$. The maximal ideal of A is $\mathfrak{m}_w = \{x \in A : w(x) > 0\} \cup \{0\}$.

Proof. (i) says that w is a homomorphism onto its image $w(K^\#)$, which means that $0 = w(1) = w((-1)^2) = 2w(-1)$. Observe now that a totally ordered group cannot have elements of finite order: for if $g > 0$, then $ng > \cdots > g > 0$, and likewise if $g < 0$. Thus, in our present case, $w(-1) = 0$. This implies that $w(-x) = w(x)$, for each $x \in K^\#$, and therefore, A is closed under forming additive inverses.

(ii) clearly implies that A is closed under addition, and (i) that A is closed under multiplication. So A is a subring of K , with identity. Further, for each $x \in K^\#$, either $w(x) \geq 0$ or $w(x) \leq 0$, which, in turn, means that $x \in A$ or $x^{-1} \in A$. So A is a valuation ring of K .

By an application of the First Isomorphism Theorem, $w(K^\#) = K^\# / \ker(w)$, and $\ker(w) = \{x \in K^\# : w(x) = 0\}$. We shall let the reader verify that $\ker(w) = U$, the groups of units of A .

Finally, there is the matter of preservation of order; note that $Ux \geq Uy$ precisely when $xy^{-1} \in A$, which, in turn happens if and only if $0 \leq w(xy^{-1}) = w(x) - w(y)$. This is so if and only if $w(x) \geq w(y)$. ■

Definition 11.2.4. A mapping $w : K^\# \rightarrow G$ which satisfies (i) and (ii) in Proposition 11.2.3 is called a *valuation* and, if $w(K^\#) = G$, we say that G is the *value group* of w . Clearly, there is no ambiguity in the notions “value group of a valuation ring” and “value group of a valuation”.

The valuation ring A of the proposition is referred to as the *valuation ring of the valuation* w .

Now we go through a number of exercises which are designed to exhibit the elegant relationship between prime ideals of a valuation ring and certain subgroups of its value group. This correspondence, in turn leads to a Galois connexion.

Exercise 11.2.5. Suppose that $w : K^\# \rightarrow G$ is a valuation with value group G . A subgroup H of G is said to be *convex* if $a \leq c \leq b$, with $a, b \in H$ imply that $c \in H$.

Now, if A is the valuation ring of w , and $\mathfrak{p} \in \text{Spec}(A)$, define

$$\gamma(\mathfrak{p}) = \langle \{w(s) : s \in A \setminus \mathfrak{p}\} \rangle.$$

($\langle S \rangle$ denotes the subgroup of G generated by S .) Conversely, if H is a convex subgroup of G , let

$$P(H) = \{x \in A : w(x) \notin H\}.$$

Prove:

- $\gamma(\mathfrak{p})$ is a convex subgroup of G , for each $\mathfrak{p} \in \text{Spec}(A)$.
- $P(H)$ is a prime ideal of A , for each convex subgroup H of G .
- The mappings γ and P are mutually inverse, order-reversing bijections of $\text{Spec}(A)$ onto the set $\mathcal{C}(G)$ of all convex subgroups of G .

And now fodder for the topic of Galois connexions, upon which we shall perhaps elaborate, if time permits. The next exercise also falls into the if-time-permits category.

Exercise 11.2.6. Suppose that $w : K^\# \rightarrow G$ is a valuation with value group G . Suppose that H is a convex subgroup of H .

- Partially order G/H by defining $H + g \geq H + h$ to mean that $g - h \geq x$, for some $x \in H$. Prove that this definition yields a well defined partial order, and that it makes G/H into a totally ordered group, so that the canonical map $g \mapsto H + g$ is order-preserving. (This is a fact about ordered groups; it has nothing to do with valuations.)

- (b) Define $w^* : K^\# \rightarrow G/H$ by $w^*(x) = H + w(x)$. Prove that w^* is a valuation onto G/H (i.e., with value group G/H) and that the valuation ring of w^* is $A_{P(H)}$.
- (c) Prove that there is a valuation $v : q(A/P(H))^\# \rightarrow H$, with H as value group, so that the valuation ring of v is exactly $A/P(H)$.

To conclude this discussion, let us show how to use these ideas to construct plenty of valuation rings. The next exercise shows, in fact, that every totally ordered group is the value group of some valuation.

Exercise 11.2.7. Suppose that G is a totally ordered group, this time written multiplicatively. Consider the group algebra $K[G]$, over any field K . (Review the notion of the group algebra, from Chapter 7. It is convenient here, in view of the ordering on G , to view the nonzero elements of $K[G]$ as expressions $r_1g_1 + r_2g_2 + \cdots + r_kg_k$, where each $r_i \in K$ is nonzero, and $g_1 < g_2 < \cdots < g_k$. Each element of $K[G]^\#$ has a unique expression of that type.)

- (a) Prove that $K[G]$ is an integral domain.

Define $v_o : K[G]^\# \rightarrow G$ as follows:

$$v_o(r_1g_1 + \cdots + r_kg_k) = g_1.$$

(It is understood that the elements of $K[G]^\#$ are represented as stipulated above.)

- (b) Prove that v_o satisfies conditions (i) and (ii) of Proposition 11.2.3.
- (c) Let L be the field of fractions of $K[G]$. Show that there is a valuation v of L , extending v_o , the value group of which is exactly G .

11.3 p -adic Valuations and p -adic Integers. In this section, we introduce the concept of a discrete valuation, which serves as introduction to the discussion of the classical rings of commutative algebra. Our first target is the ring of p -adic integers, and then, in the next chapter, we take up Dedekind domains.

Definition & Remarks 11.3.1. A valuation v with cyclic value group (i.e., with \mathbb{Z} as value group, in its natural ordering) is called a *discrete valuation*. The valuation ring of a discrete valuation is also called a *discrete valuation ring*.

Note that if A is a discrete valuation ring, then, since \mathbb{Z} has no nonzero proper convex subgroups, it follows from Exercise 11.2.5, that the maximal ideal of A is the only nonzero prime ideal. The converse is false: letting $G = \mathbb{R}$ in Exercise 11.2.7, one constructs a valuation v with value group \mathbb{R} . \mathbb{R} also has no nonzero proper convex subgroups. So the valuation ring v has the feature that its maximal ideal is the only nonzero prime ideal.

We shall now quote a classical result from the theory of ordered groups, which gives us a handle on when a valuation ring has a value group which is a subgroup of the additive reals. A proof of this theorem may be found in [Fu63] and also in [D95], Theorem 24.16. We give an outline of the proof.

Hölder's Theorem dates back to 1902.

Theorem 11.3.2. (Hölder's Theorem) *Suppose that G is a totally ordered group; (not even necessarily abelian!) Then the following are equivalent:*

- (a) G is order-isomorphic to a subgroup of the additive group of real numbers in the usual ordering.
 (b) G has no nonzero, proper convex subgroups.

Proof. That (a) \Rightarrow (b) is clear from the archimedean property of the real numbers.

Now assume (b), and suppose that $0 < a \in G$; this element is fixed in the proof. For each $b \in G$ let

$$b^* = \left\{ \frac{m}{n} \in \mathbb{Q} : ma \leq nb \right\}.$$

The reader will easily establish the following facts:

1. $0 \in b^*$, so that $b^* \neq \emptyset$.
2. $\frac{p}{q} \leq \frac{m}{n} \in b^*$ implies that $\frac{p}{q} \in b^*$. This says that each $b \in G$ defines a “cut” in the rationals.
3. Define $\tau : G \rightarrow \mathbb{R}$, first on positive elements of G :

$$\tau(b) = \bigvee b^*.$$

Prove that τ is a homomorphism and extend it to G in the obvious manner: if $g \leq 0$ in G , define $\tau(g) = -\tau(-g)$. Show that this extension is still a homomorphism. Note that $\tau(a) = 1$.

4. τ is one-to-one and it preserves order.

Notice that the commutativity of G is hereby forced. ■

Hölder’s Theorem immediately implies the following.

Proposition 11.3.3. *Suppose that A is a valuation ring. Then the value group of A is order-isomorphic to a subgroup of the additive reals if and only if the Krull dimension of A is 1.*

Example 11.3.4. Let p be a fixed prime number, and $\mathbb{Q}_{(p)}$ be the local ring of all rational numbers which can be written $\frac{a}{b}$, where b is relatively prime to p . We’ve already observed it, but $\mathbb{Q}_{(p)}$ is a valuation ring. The reader should verify that the only ideals of $\mathbb{Q}_{(p)}$, aside from $\mathbb{Q}_{(p)}$ and $\{0\}$, are in the sequence

$$p\mathbb{Q}_{(p)} > p^2\mathbb{Q}_{(p)} > \cdots > p^n\mathbb{Q}_{(p)} > \cdots.$$

$\mathbb{Q}_{(p)}$ is also a principal ideal domain, and its maximal ideal is the only nonzero prime ideal.

Now if $x \in \mathbb{Q}_{(p)}^\#$, then x can be written uniquely as follows: $x = \frac{a}{b}p^k$, where $k \in \mathbb{Z}$, and a and b are integers, both relatively prime to p . Define $v_p(x) = k$.

The reader should verify that v_p is a valuation. Once that is done, it is clear that its value group is \mathbb{Z} . Moreover, the valuation ring of v_p is $\mathbb{Q}_{(p)}$. Thus, $\mathbb{Q}_{(p)}$ is a discrete valuation ring. v_p is called the *p -adic valuation on \mathbb{Q}* .

Here is the by-now-standard companion to the preceding example.

Exercise 11.3.5. Let K be a field, and $K[[T]]$ be the formal ring of power series with coefficients in K . We have already seen that this is valuation ring; refer to Exercise 10.2.8.

Define a discrete valuation on the field of fractions of $K[[T]]$ with $K[[T]]$ as valuation ring. In order to do this, it is first worth deciding what $v_K[[T]]$ is.

To get at the p -adic integers, one needs to introduce a metric, and then consider when this metric is complete, in the sense that every Cauchy sequence converges. The reader should at this time have a working understanding of metric topology.

Definition & Remarks 11.3.6. Suppose that $v : K^\# \rightarrow G$ is a valuation with value group G , and assume that G is an additive subgroup of \mathbb{R} . Thus, it is assumed that A has Krull dimension 1.

Define $\exp(v)$ to mean the obvious: $\exp(v)(x) = 0$ if $x = 0$, and $\exp(v)(x) = e^{-v(x)}$, otherwise. (Well, perhaps one of the obvious possibilities!) Verify that $\exp(v)$ is a function from $K \rightarrow \mathbb{R}^+$, so that

- (a) $\exp(v)(x) = 0$ if and only if $x = 0$.
- (b) $\exp(v)(xy) = \exp(v)(x)\exp(v)(y)$, for all $x, y \in K$.
- (c) $\exp(v)(-x) = \exp(v)(x)$, for all $x \in K$.
- (d) $\exp(v)(x+y) \leq \max\{\exp(v)(x), \exp(v)(y)\}$, $\forall x, y \in K$, and this maximum is bounded by the sum $\exp(v)(x) + \exp(v)(y)$.

Then, letting $d_v(x, y) = \exp(v)(x - y)$, a metric is defined on the field K . A metric defines a canonical topology, in which a base of neighborhoods around $x \in K$ consists of the open disks centered at x . More precisely, the base of open neighborhoods of x consists of the set of all $\rho(x, n)$ ($n \in \mathbb{N}$) defined by

$$\begin{aligned} \rho(x, n) &= \{y \in K : d_v(x, y) < e^{-n}\} \\ &= \{x\} \cup \{y \in K : y \neq x, \text{ and } v(x - y) > n\}. \end{aligned}$$

Although the next exercise will be generalized almost immediately, it is still worth thinking about.

Exercise 11.3.7. If v_p is the p -adic valuation on \mathbb{Q} , then show that $\rho(0, n) = p^{n+1}\mathbb{Q}_{(p)}$ – that’s right! – the ideal of $\mathbb{Q}_{(p)}$ generated by p^{n+1} .

The situation with the classical p -adic valuation can be generalized.

Hot Air 11.3.8. *Generalizing the p -adic case.* Suppose now that A is any discrete valuation ring, and let $v : qA^\# \rightarrow \mathbb{Z}$ be a discrete valuation with valuation ring A . Let $q \in A$, such that $v(q) = 1$. Then the reader can easily verify that $v(a) \geq 1$ precisely when $a \in Aq$, and that $v(a) = 1$ if and only if $a = uq$, where u is a unit of A .

More generally, the reader can check (by induction) that $v(a) = n > 0$ if and only if $a = uq^n$. Now, since it is true that $v(a) > 0$, for any nonunit in A (Proposition 11.2.3), we’ve shown that each element of A is either a unit, or zero, or of the form $a = uq^n$, for a suitable unit u and a positive integer n . This proves half of the following theorem.

Theorem 11.3.9. *Suppose that A is a valuation ring. Then A is discrete if and only if it is a principal ideal domain. If so, then in A any two irreducible elements are unit multiples of one another.*

Proof. The discussion leading up to the theorem proves the necessity; in fact, we showed that the nonzero ideals are

$$A > Aq > \cdots > Aq^n > \cdots.$$

As to sufficiency, if A is a PID and also a valuation ring, then the maximal ideal \mathfrak{m} of A is generated by an irreducible q . Since every irreducible in a PID generates a maximal ideal, it follows that any two irreducibles are unit multiples of one another.

Recall that any PID is a unique factorization domain. Thus each $a \in A$ which is nonzero and not a unit is of the form $a = uq^n$, with u a unit, and $n \in \mathbb{N}$, unique up to units. Hence each $x \in qA^\#$

can be expressed as $x = uq^n$, where $n \in \mathbb{Z}$, and u is a unit in A . Now define $v : qA^\# \rightarrow \mathbb{Z}$ by $v(uq^n) = n$. As in the case with the p -adic valuation, v turns out to be a valuation with valuation ring A . We leave the verification of this part to the reader. ■

Here's a generalization of Exercise 11.3.7.

Exercise 11.3.10. Suppose that A is a discrete valuation ring, and $a \in A$. Then

$$\rho(a, n) = Aq^{n+1} + a;$$

that is, the coset of a , modulo the ideal generated by the $n + 1$ -st power of the irreducible element q .

Hot Air 11.3.11. Metric Stuff. Any metric space (X, d) , where d is the metric, can be completed. We shall not outline the process, but refer the reader to any standard text on the subject. [Wi70] does the job well enough.

Let's at least review the concept of completeness. In the metric space (X, d) a sequence $(s_n)_n$ converges to $s \in X$, if for each $\varepsilon > 0$ there is a positive integer k , so that $n \geq k$ implies that $d(s_n, s) < \varepsilon$. The sequence $(s_n)_n$ is *Cauchy* if for each $\varepsilon > 0$ there is a positive integer k , so that for all $n, m \geq k$, it follows that $d(s_n, s_m) < \varepsilon$.

The metric d is *complete* if every Cauchy sequence converges. Now, every metric space (X, d) admits a *completion*; that is to say, there is a complete metric space (X^*, d^*) , with $X \subseteq X^*$, and so that the metric d^* restricts to d on X , and X is *dense* in X^* , which means that for each $x \in X^*$ and each $\varepsilon > 0$, there is a point $p \in X$, so that $d^*(x, p) < \varepsilon$. It is well known that the density condition is equivalent to saying that each $x \in X^*$ is the limit of a Cauchy sequence $(x_n)_n$ of points in X .

We shall conclude the discussion of valuation rings by giving a characterization of complete discrete valuation rings.

Definition 11.3.12. A *complete discrete valuation ring* A is a discrete valuation ring which arises as the valuation ring of a valuation v , for which the associated metric d_v is complete.

Suppose, in the discussion which follows that A is a complete discrete valuation ring, that v is a discrete valuation with valuation ring A , so that d_v is complete. Fix an irreducible generator q for the maximal ideal \mathfrak{m} of A . Let R denote a complete set of representatives of the cosets in the residue field A/\mathfrak{m} .

Note: *Without loss of generality, $v(uq^n) = n$, for each unit u of A , and each integer n .* (Check this! It follows from the fact that $v(q) = 1$, necessarily.)

Theorem 11.3.13. *Suppose that A, M, q, R, v and d_v are as outlined in 11.3.12. Then each $a \in A$ can be expressed uniquely as a series*

$$a = \sum_{n=0}^{\infty} a_n q^n,$$

where the $a_n \in R$.

Proof. By induction. First, consider $\mathfrak{m} + a$; as $\mathfrak{m} + a = \mathfrak{m} + a_0$, for some $a_0 \in R$, we have that $a - a_0 \in \mathfrak{m} = Aq$. Now, suppose we have found $a_0, a_1, \dots, a_k \in R$ so that

$$a - (a_0 + a_1q + \dots + a_kq^k) \in Aq^{k+1}.$$

Then, $a - (a_0 + a_1q + \dots + a_kq^k) = yq^{k+1}$, for some $y \in A$. Now $y - a_{k+1} \in \mathfrak{m}$, which means that $y - a_{k+1} = zq$, for some $z \in A$. Substituting, we get:

$$a - (a_0 + a_1q + \dots + a_{k+1}q^{k+1}) \in Aq^{k+2}.$$

Now, the n -th terms of this series converge to 0, and this is enough, for this metric, to make the series converge; (see the exercise which follows.) In view of Exercise **11.3.10**, the series clearly converges to a . ■

Exercise 11.3.14. Again, under the conditions of **11.3.12**, prove that a series $a_1 + a_2 + \cdots + a_n + \cdots$ converges if and only if $\lim_n a_n = 0$. (Note: As you know the necessity holds in any metric space; the novelty is the sufficiency.)

Definition & Remarks 11.3.15. *The ring J_p of p -adic integers.* We begin with some hot air, which is intended to plug the gap left by the decision to omit actual construction of the completion. It is not hard to check that if A^* is a complete discrete valuation ring, with maximal ideal \mathfrak{m}^* , and A is the completion of A with maximal ideal \mathfrak{m} , then \mathfrak{m}^* contracts to \mathfrak{m} , and the residue fields A/\mathfrak{m} and A^*/\mathfrak{m}^* are isomorphic. (Use the density and Exercise **11.3.10**.)

So let's consider (for a fixed prime number p) the discrete valuation ring $\mathbb{Q}_{(p)}$ with respect to the classical p -adic valuation v_p . Here $\mathbb{Q}_{(p)}/p\mathbb{Q}_{(p)}$ is isomorphic to the field \mathbb{Z}_p . Let's use the numbers $0, 1, \dots, p-1$ as the complete set of representatives modulo $p\mathbb{Q}_{(p)}$.

If J_p denotes the p -adic completion, that is to say, the valuation ring of the completion of v_p , then, according to Theorem **11.3.13**, each member a of J_p can be uniquely expressed as

$$a = \sum_{n=0}^{\infty} a_n p^n,$$

where the $a_n \in \{0, 1, \dots, p-1\}$. J_p is called *the ring of p -adic integers*.

Let's close with a word about the arithmetic of these power series. Suppose that $b = b_0 + b_1 p + \cdots + b_n p^n + \cdots$ is also a p -adic integer. What are the power series expansions for $a + b$ and ab ? Answer: add term-by-term, and *carry*, as in ordinary arithmetic:

$$a + b = \sum_{n=0}^{\infty} c_n p^n,$$

where the c_n are calculated as follows: $c_0 \equiv a_0 + b_0 \pmod{p}$; if $c_0 - (a_0 + b_0) = z_1 p$, then $c_1 = a_1 + b_1 + z_1 \pmod{p}$; and so on, by induction. The product is resolved in the same manner.

12. Commutative Algebra: The Classical Rings

We consider Noetherian rings and modules in this chapter, proving both Hilbert's Basis Theorem and the Nullstellensatz. The final section deals with Dedekind domains, which are the natural generalizations of UFD's, in terms of the arithmetic of ideals. Our main reference continues to be [AM69], mainly Chapters 7 and 9. As in the various previous chapters, all rings (if referred to without any additional designation) are assumed to be commutative and possess an identity. Categorical references are to **Crn1**.

12.1 Noetherian and Artinian Rings and Modules.

Definition 12.1.1. Suppose that A is a commutative ring with 1. An A -module M is said to be *Noetherian* (resp. *Artinian*) if the lattice of submodules of M satisfies the ascending (resp. descending) chain condition.

We say that the ring A is *Noetherian* (resp. *Artinian*) if it is a Noetherian (resp. Artinian) module over itself. These notions coincide with the ones introduced in Chapter 7. Eventually, in Proposition 12.1.18, we will show that, for commutative rings, any Artinian ring is also Noetherian. In Chapter 7 this had been established for not-necessarily commutative rings with identity, under the assumption of J -semisimplicity.

We proceed immediately to recall Exercise 1.3.7, adapted to the present context.

Proposition 12.1.2. *For an A -module M the following are equivalent.*

- (a) M is Noetherian.
- (b) Every A -submodule is finitely generated.
- (c) Every nonempty family of proper A -submodules has a maximal member.

In particular A is a Noetherian ring if and only if every ideal of A is finitely generated.

The "Artinian" counterpart of Proposition 12.1.2 follows, as an exercise:

Exercise 12.1.3. The A -module M is Artinian if and only if every nonempty family of nonzero A -submodules has a minimal member.

Much of the discussion in this section is patterned after Chapter 6 of [AM69].

Examples 12.1.4. (A) Over the ring \mathbb{Z} , the group of all complex p^n -th roots of unity (all $n \in \mathbb{N}$) is Artinian but not Noetherian.

(B) All PIDs are Noetherian. Is a UFD necessarily Noetherian?

(C) For any field F , the polynomial ring $F[T_1, T_2, \dots]$ in an infinite number of indeterminates is neither Noetherian nor Artinian.

(D) Recall that an integral domain which is Artinian is a field.

(E) Von Neumann regular rings need not be Noetherian. Example?

The following result is typical of theorems about chain conditions.

Proposition 12.1.5. *Suppose that $0 \rightarrow M \rightarrow N \rightarrow P \rightarrow 0$ is an exact sequence of A -modules. Then*

- (a) *N is Noetherian if and only if M and P are Noetherian.*
- (b) *N is Artinian if and only if M and P are Artinian.*

Proof. We prove (a) only, as the proof of (b) is quite similar.

Obviously, an infinite ascending sequence of submodules of M is an ascending sequence of submodules of N . Likewise, if

$$P_1 < P_2 < \dots$$

is an ascending sequence of submodules of P , then, by taking inverse images under the map $N \rightarrow P$, we obtain an infinite ascending sequence of submodules of N . Thus, if N is Noetherian, then so are the other two.

Suppose now that $N_1 < N_2 < \dots$ is an ascending sequence of submodules of N , yet both M and P are Noetherian. For a suitable index t ,

$$M \cap N_t = M \cap N_{t+1} = \dots,$$

and a suitable index s ,

$$M + N_s = M + N_{s+1} = \dots$$

For $u = \max(s, t)$ both sequences become stationary for $n \geq u$.

By one of the isomorphism theorems, $(M + N_i)/M \cong N_i/(M \cap N_i)$, for each i . Now, for $n \geq u$, the left side of these identities is constant, while the denominator of the right side is too. Thus $N_n = N_u$, for all $n \geq u$, a contradiction. ■

Corollary 12.1.6. *Suppose that M_i ($i = 1, 2, \dots, k$) are A -modules, and that M is their direct sum. Then M is Noetherian (resp. Artinian) if and only if each M_i is Noetherian (resp. Artinian).*

Proof. Apply Proposition 12.1.5 inductively to sequences

$$0 \rightarrow M_i \rightarrow M_1 \oplus M_2 \oplus \dots \oplus M_i \rightarrow M_1 \oplus M_2 \oplus \dots \oplus M_{i-1} \rightarrow 0.$$

Here's an example of a very natural commutative ring which is, in general, far from being Noetherian. ■

Exercise 12.1.7. Let $C(X)$ be the ring of all continuous real valued functions defined on the compact Hausdorff space X . Prove that $C(X)$ is Noetherian if and only if X is a finite space.

Next, we have another basic feature of Noetherian and Artinian modules.

Proposition 12.1.8. *If A is a Noetherian (resp. Artinian) ring, then every finitely generated A -module is Noetherian (resp. Artinian).*

Proof. If M is finitely generated, then it is a homomorphic image of a finitely generated free A -module. Thus, M is a homomorphic image of a finite direct sum of copies of A . Now apply Proposition 12.1.5. ■

More basic stuff on Noetherian and Artinian rings and modules in the next few exercises.

Exercise 12.1.9. If A is a Noetherian (resp. Artinian) ring, and \mathfrak{r} is an ideal of A , then A/\mathfrak{r} is a Noetherian ring. (However, a subring of a Noetherian ring need not be Noetherian; ditto for Artinian. Examples?)

The next exercise is sometimes referred to as Fitting's Lemma.

Exercise 12.1.10. Let M be a Noetherian A -module, and $f : M \rightarrow M$ be an $\mathbf{A}\mathbf{Mod}$ -morphism. If f is surjective it is necessarily an isomorphism. (Hint: Consider the sequence of submodules $\ker(f^n)$.) Do formulate and prove the corresponding property relative to Artinian modules. (Else, see Exercise 1, p. 78, in [AM69].)

Exercise 12.1.11. Suppose that M is a Noetherian A -module, and let $\mathfrak{r} = \text{Ann}(M)$. Prove that A/\mathfrak{r} is a Noetherian ring.

Is the corresponding result true for Artinian modules?

Exercise 12.1.12. Suppose that the ring A has the ascending chain condition for prime ideals. Does it follow that A is a Noetherian ring?

The rest of this section is devoted to the so-called "Jordan-Hölder" theory. Formally, by a *chain* of submodules of an A -module M we mean a sequence

$$M = M_0 > M_1 > M_2 > \cdots > M_n = 0.$$

n is called the *length* of the chain.

Definition & Remarks 12.1.13. Let M be an A -module. A chain $M = M_0 > \cdots > M_n = \{0\}$ is called a *composition series* if each M_{i-1}/M_i is a simple A -module. Observe that this condition is equivalent to stipulating that the chain is *irrefinable*; that is to say, if $M = N_0 > N_1 > \cdots > N_k = \{0\}$ is a chain, so that each M_i equals some N_j , then $n = k$, and $M_i = N_i$, for each $i = 0, 1, \dots, n$.

Suppose that M is an A -modules, and that a composition series exists for M . (\mathbb{Z} as a \mathbb{Z} -module has none!) Then M has a composition series of smallest length. This least number is called the *length* of M , denoted $\Lambda(M)$. In a moment we will establish that all composition series have the same length, but for now this concept is certainly well defined.

If M has no composition series, we say that $\Lambda(M) = \infty$. The notion of length ought to "feel" like the concept of dimension of a vector space.

The following proposition establishes the basic properties of the length of a module.

Proposition 12.1.14. *Suppose M is an A -module with a composition series. Then*

- (a) every submodule N has a composition series, and if N is proper, then $\Lambda(N) < \Lambda(M)$;
- (b) the length k of any chain in M satisfies $k \leq \Lambda(M)$; thus, it follows that any two composition series have the same length;
- (c) an A -module N has a composition series if and only if it is both Noetherian and Artinian.

Proof. (a) Suppose that N is a submodule of M , and let

$$M = M_0 > M_1 > \cdots > M_n = \{0\}, \quad (n = \Lambda(M))$$

be a shortest composition series. By intersecting with N , one obtains a chain for N . Put $N_i = N \cap M_i$; for each $i = 0, 1, \dots, n$. Notice that (by the appropriate isomorphism theorem)

$$\begin{aligned} N_{i-1}/N_i &= (N \cap M_{i-1})/((N \cap M_{i-1}) \cap M_i) \\ &\cong ((N \cap M_{i-1}) + M_i)/M_i. \end{aligned}$$

Since M_{i-1}/M_i is simple, it follows that $N_{i-1} = N_i$, or else $N_{i-1}/N_i \cong M_{i-1}/M_i$, hence simple. Thus, dropping repetitions in the N_i 's, one gets a composition series for N . Evidently $\Lambda(N) \leq \Lambda(M)$.

If $\Lambda(N) = \Lambda(M)$, then in the preceding paragraph, $N_{i-1}/N_i \cong M_{i-1}/M_i$, for each $i = 0, 1, \dots, n$. Proceeding inductively up the two chains, one obtains that $M_i = N_i$, for each i , whence $M = N$.

(b) For any chain $M = M_0 > M_1 > \dots > M_k = \{0\}$, we have from (a) that

$$\Lambda(M) > \Lambda(M_1) > \dots > \Lambda(M_k) = 0.$$

Then it is clear that $k \leq n$. The second claim in (b) should then be obvious.

(c) (Necessity) If $\Lambda(M) = n < \infty$, then for any ascending sequence of submodules $M_1 \leq M_2 \leq \dots$, we conclude that

$$\Lambda(M_1) \leq \Lambda(M_2) \leq \dots \leq n.$$

Then, clearly, the sequence of submodules cannot ascend without becoming stationary. Thus, M is Noetherian. By a similar argument, M is Artinian.

(Sufficiency) Suppose that M is both Noetherian and Artinian. Let $M_0 = M$. Pick any maximal (proper) submodule M_1 . (Which exists by Proposition 12.1.2(c). Next, let M_2 be any proper submodule of M_1 , which is maximal as a proper submodule of M_1 . Proceed by induction. Eventually, (after some k) $M_k = \{0\}$ (because M is Artinian), and by its very construction, the chain thus obtained is a composition series. ■

For vector spaces the length of a module is its dimension over the field.

Exercise 12.1.15. Suppose that V is a vector space over a field F . Then the following are equivalent.

- (a) V has finite dimension.
- (b) V has finite length (as an F -module).
- (c) V is Noetherian (as an F -module).
- (d) V is Artinian (as an F -module).

Moreover, if these are satisfied, then $\Lambda(V) = \dim(V)$.

Let's conclude this section by looking a bit more closely at Artinian rings.

Corollary 12.1.16. (Of the preceding exercise) *Suppose that A is a ring and that $\mathfrak{m}_1 \mathfrak{m}_2 \cdots \mathfrak{m}_k = 0$, for a suitable set of maximal ideals of A (which aren't necessarily distinct). Then A is Noetherian if and only if it is Artinian.*

Proof. Consider the chain of ideals (A -submodules)

$$A > \mathfrak{m}_1 \geq \mathfrak{m}_1 \mathfrak{m}_2 \geq \dots \geq \mathfrak{m}_1 \mathfrak{m}_2 \cdots \mathfrak{m}_k = \{0\}.$$

Each factor $(\mathfrak{m}_1 \cdots \mathfrak{m}_{i-1})/(\mathfrak{m}_1 \cdots \mathfrak{m}_i)$ can be regarded as an A/\mathfrak{m}_i -vector space. Now, by the exercise, each factor satisfies the a.c.c if and only if it does the d.c.c (on A/\mathfrak{m}_i -subspaces, and hence on A -submodules.) By repeated applications of Proposition 12.1.5, it follows that A has the a.c.c. on ideals if and only if it has the d.c.c. ■

We need a lemma before proceeding.

Lemma 12.1.17. *An Artinian ring A has a finite number of maximal ideals.*

Proof. Consider the family of all finite intersections of maximal ideals of A . By the Artinian property (Exercise 12.1.3), this set has a minimal element $\mathfrak{r} = \mathfrak{m}_1 \cap \mathfrak{m}_2 \cap \cdots \cap \mathfrak{m}_k$. Now, if \mathfrak{n} is any maximal ideal of A , then $\mathfrak{n} \cap \mathfrak{r} = \mathfrak{r}$, which is to say that $\mathfrak{r} \leq \mathfrak{n}$. We leave it to the reader to check that this implies that \mathfrak{n} actually contains one of the \mathfrak{m}_i , and therefore equals it. ■

Part (a) of the next proposition was Corollary 7.4.7, for semisimple rings. Part (b) is contained in 7.4.12(c).

Proposition 12.1.18. *Let A be a ring.*

- (a) *If A is Artinian it is also Noetherian.*
- (b) *If A is Artinian then $n(A) = \mathfrak{J}(A)$.*
- (c) *If A is Artinian then every prime ideal is maximal.*

Proof. (a) For some k , $n(A)^k = 0$ (by 7.4.12(b)). Now let $\mathfrak{m}_1, \mathfrak{m}_2, \dots, \mathfrak{m}_n$ be all the maximal ideals of A (apply Lemma 12.1.17). Then

$$(\mathfrak{m}_1 \mathfrak{m}_2 \cdots \mathfrak{m}_n)^k \leq (\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n)^k \leq n(A)^k = \{0\}.$$

Now apply Corollary 12.1.16, and it follows that A is Noetherian.

(c) $A/n(A)$ is Artinian and J -semisimple, and therefore also von Neumann regular, by Exercise 8.2.14(e). By Theorem 9.1.11(d). ■

Remark 12.1.19. Proposition 12.1.18, (a) and (c), tell us that an Artinian ring is Noetherian and has Krull dimension 0. We prove the converse of this in §3: a Noetherian ring which has 0 Krull dimension is Artinian.

12.2 Two Theorems of Hilbert. In this section we prove two classical theorems of David Hilbert about Noetherian rings.

So far nothing has been said about the effect on chain conditions of passing to a ring of fractions. Let's begin by correcting that omission.

Proposition 12.2.1. *Suppose that A is a Noetherian ring, and S any multiplicative system. Then $S^{-1}A$ is Noetherian as well.*

Proof. It is a consequence of Proposition 10.2.1(a) that the lattice of ideals of $S^{-1}A$ is isomorphic to the sublattice of contracted ideals of A (those of the form $\mathfrak{r} = \mathfrak{r}^{ec}$). Since the lattice of ideals of A satisfies the a.c.c., so does the lattice of contracted ones. ■

The following corollary is an immediate consequence. Being Noetherian is not a local property, however. If A is any von Neumann regular ring, then $A_{\mathfrak{p}}$ is a field, for each prime ideal \mathfrak{p} of A , and hence Noetherian. However, an infinite product of fields is not Noetherian, but it is locally Noetherian.

Corollary 12.2.2. *For any prime ideal \mathfrak{p} of a Noetherian ring A , the localization $A_{\mathfrak{p}}$ is Noetherian.*

Let us also summarize some properties from the previous section.

Proposition 12.2.3. (a) *If $f : A \rightarrow B$ is a surjective homomorphism of rings, and A is Noetherian, then so is B .*

(b) *Suppose that A is Noetherian, and a subring of B , so that B is a finitely generated A -module. Then B is a Noetherian ring.*

Proof. (a) By Exercise 12.1.9, plus the First Isomorphism Theorem.

(b) Apply Proposition 12.1.8: B is a Noetherian A -module, hence also Noetherian over itself. ■

Now the first of the theorems of Hilbert announced at the start of this section: polynomial rings over Noetherian rings are Noetherian.

Theorem 12.2.4. Hilbert's Basis Theorem. *If A is a Noetherian ring then so is the polynomial ring $A[T]$.*

Proof. We prove that every ideal of $A[T]$ is finitely generated. To that end, let \mathfrak{r} be an ideal of $A[T]$. The set of leading coefficients of elements in \mathfrak{r} , say \mathfrak{r}_0 , forms an ideal of A . Suppose that $\{a_1, \dots, a_n\}$ is a generating set for \mathfrak{r}_0 . Let $f_i(T)$ be a polynomial in \mathfrak{r} of degree r_i , with leading coefficient a_i . Then with

$$\mathfrak{r}' \equiv A[T]f_1 + \dots + A[T]f_n,$$

we have $\mathfrak{r}' \leq \mathfrak{r}$. Let r be the maximum of the r_i .

Now pick $f \in \mathfrak{r}$, with leading coefficient $a \in \mathfrak{r}_0$ and degree m . If $m \geq r$, write $a = u_1a_1 + \dots + u_na_n$. Then

$$f - \left(\sum_{i=1}^n u_i f_i T^{m-r_i} \right) \in \mathfrak{r},$$

and has degree less than m . Continuing in this manner, we may eventually write $f = g + h$, with $g \in \mathfrak{r}$, $\deg(g) < r$, and $h \in \mathfrak{r}'$.

Next, if M is the A -submodule generated by $\{1, T, \dots, T^{r-1}\}$, then the above paragraph shows that $\mathfrak{r} = (M \cap \mathfrak{r}) + \mathfrak{r}'$. Since M is finitely generated and A is a Noetherian ring, it follows that $M \cap \mathfrak{r}$ is finitely generated (as an A -module). Thus, \mathfrak{r} is finitely generated as an ideal of $A[T]$, by the A -generators of $M \cap \mathfrak{r}$ and the f_i . ■

We have two corollaries of Hilbert's Basis Theorem; one iterates to get the first, and very little else to get the second.

Corollary 12.2.5. *If A is Noetherian, then so is $A[T_1, T_2, \dots, T_n]$, for any finite number of indeterminates.*

Corollary 12.2.6. *Suppose that B is an A -algebra, and finitely generated as an A -algebra. Then, if A is Noetherian, so is B . In particular, any finitely generated ring (over \mathbb{Z}), and any finitely generated F -algebra, over any field F , is Noetherian.*

Proof. The concluding assertion is obvious from the first. As to the first claim, if B is a finitely generated A -algebra, then it is a homomorphic image of the free algebra $A[T_1, T_2, \dots, T_n]$, for suitable n . Now use Proposition 12.2.3(a). ■

Of course, in the context of fields, we already know more than is stated in the previous corollary. Recall Exercise 11.1.9: suppose that the field L is an extension of the field K , and a finitely generated K -algebra. Then L is a finite algebraic extension of K . Recall also the consequence of 11.1.9, namely Exercise 11.1.10. (Do it! It's a preliminary form of the Nullstellensatz ahead.)

One more item should be mentioned in connection with Hilbert's Basis Theorem. It's the version for power series rings.

Exercise 12.2.7. If A is a Noetherian ring, prove that $A[[T]]$ is too. (Hint: Imitate the proof of the Basis Theorem, but instead of the leading coefficients – which may not be there – consider an ideal of lowest-term coefficients.)

Let us now proceed to the Nullstellensatz (Theorem 12.2.11), the classical theorem of Hilbert, promised in this section. This theorem can reasonably be regarded as an entry to the study of algebraic geometry.

Definition 12.2.8. Suppose that K is a field, and suppose that $A = K[T_1, \dots, T_n]$. For each ideal \mathfrak{r} of A , let

$$V(\mathfrak{r}) = \{ (x_1, \dots, x_n) \in K^n : f(x_1, \dots, x_n) = 0, \forall f \in \mathfrak{r} \}.$$

$V(\mathfrak{r})$ is referred to as the *variety (of points) in K^n defined by \mathfrak{r}* .

Conversely, suppose that V is a set of points in K^n . Let

$$\mathfrak{I}_V = \{ f \in A : f(x_1, \dots, x_n) = 0, \forall (x_1, \dots, x_n) \in V \}.$$

Thus, think of \mathfrak{I}_V as the set of polynomials which vanish on V . It is not hard to verify that \mathfrak{I}_V is an ideal.

The following are also easy to check, and are left to the reader. You should smell a topology coming on after reading it.

Proposition 12.2.9. Suppose that K is a field, $A = K[T_1, \dots, T_n]$, and V and V' are subsets of K^n , while \mathfrak{r} and \mathfrak{s} are ideals of A . Then

- (a) $V \subseteq V'$ implies that $\mathfrak{I}_{V'} \leq \mathfrak{I}_V$;
- (b) $\mathfrak{r} \leq \mathfrak{s}$ implies that $V(\mathfrak{s}) \subseteq V(\mathfrak{r})$.
- (c) $V(\mathfrak{r}\mathfrak{s}) = V(\mathfrak{r} \cap \mathfrak{s}) = V(\mathfrak{r}) \cup V(\mathfrak{s})$.
- (d) If $(\mathfrak{r}_i)_i$ is a family of ideals, then

$$V\left(\sum_i \mathfrak{r}_i\right) = \bigcap_i V(\mathfrak{r}_i).$$

- (e) $\mathfrak{r} \leq \mathfrak{I}_{V(\mathfrak{r})}$ and $V \subseteq V(\mathfrak{I}_V)$.
- (f) $V(\{0\}) = K^n$ and $V(A) = \emptyset$, while $\mathfrak{I}_\emptyset = A$.

Hot Air 12.2.10. There is an obvious omission in the list of Proposition 12.2.9. It is easy to construct examples (of finite fields, say) showing that \mathfrak{I}_{K^n} needn't be the trivial ideal.

Nonetheless, (c), (d) and (f), are enough to define a topology on K^n , for which the sets $V(\mathfrak{r})$ (the varieties of K^n) are precisely the closed sets of this topology.

For example, if $n = 2$, and $K = \mathbb{R}$, and $\mathfrak{r} = A(T_1^2 - T_2^2)$, then $V(\mathfrak{r})$ is the union of the principal axes in the plane. In general (for $n = 2$, and $K = \mathbb{R}$), $V(Af)$ will be a curve in the plane, or else a finite union of curves.

The topology referred to here is often called the *constructible* or *Zariski* topology on K^n . It is not the ordinary Euclidean topology, if $K = \mathbb{R}$ or \mathbb{C} .

This is as far as we shall take the peek into algebraic geometry. Hilbert's Nullstellensatz describes (referring to (d) in Proposition 12.2.9) $\mathfrak{I}_{V(\mathfrak{r})}$, when the underlying field is algebraically closed.

Theorem 12.2.11. Hilbert's Nullstellensatz. *Suppose that K is an algebraically closed field, $A = K[T_1, T_2, \dots, T_n]$, and \mathfrak{r} is any ideal of A . Then*

$$\mathfrak{I}_{V(\mathfrak{r})} = \sqrt{\mathfrak{r}} = \{f \in A : \exists k \in \mathbb{N}, f^k \in \mathfrak{r}\}.$$

Proof. If a power of $f \in A$ (say f^k) lies in \mathfrak{r} , then f^k vanishes at all the points of $V(\mathfrak{r})$, whence f does that. By definition, $f \in \mathfrak{I}_{V(\mathfrak{r})}$. So it is clear that $\sqrt{\mathfrak{r}} \subseteq \mathfrak{I}_{V(\mathfrak{r})}$.

Conversely, suppose that $f \notin \sqrt{\mathfrak{r}}$, and pick a prime ideal \mathfrak{p} of A , $\mathfrak{p} \supseteq \mathfrak{r}$, that does not contain f . Let $B = A/\mathfrak{p}$ and $C = B_f$, which – abusing the notation a bit – is the ring of fractions of B , the denominators of which are powers of $\mathfrak{p} + f$. Finally, let \mathfrak{m} be a maximal ideal of C . Since C/\mathfrak{m} is a finitely generated K -algebra, and K is algebraically closed, we have (owing to Exercise 11.1.10) that $C/\mathfrak{m} \cong K$. So let $x_i = \eta(\mathfrak{p} + T_i)$, where $\eta: C \rightarrow C/\mathfrak{m}$ is the canonical map. By construction, $f(x_1, \dots, x_n) \neq 0$, and as $\mathfrak{r} \subseteq \mathfrak{p}$, $(x_1, \dots, x_n) \in V(\mathfrak{r})$. ■

Over algebraically closed fields the missing item of Proposition 12.2.9 is recovered.

Corollary 12.2.12. *With K an algebraically closed field and $A = K[T_1, \dots, T_n]$, $\mathfrak{I}(K^n) = \sqrt{0} = 0$; thus, the only polynomial which vanishes at every point of K^n is the zero polynomial.*

To conclude this section, an item which is at once anticlimactic and intriguing.

Exercise 12.2.13. Suppose that K is a field, and $A = K[T_1, \dots, T_n]$. For each $x = (x_1, \dots, x_n) \in K^n$, let

$$\mathfrak{m}_x \equiv \{f \in A : f(x) = 0\};$$

that is, $\mathfrak{I}_{\{x\}}$.

(a) Prove that \mathfrak{m}_x is a maximal ideal of A .

(b) If K is algebraically closed, prove that each maximal ideal is of this form.

This means that, if K is algebraically closed, then the assignment $x \mapsto \mathfrak{m}_x$ is a bijection of K^n onto $\text{Max}(A)$. How are the topologies of K^n (the constructible one) and $\text{Max}(A)$ related?

12.3 Primary Decomposition. The goal of this section is to introduce primary ideals, and to show that, in a Noetherian ring, every ideal can be expressed as an intersection of primary ideals. This is, in a sense, one ideal-theoretic generalization of prime-power factorization of natural numbers. For the uniqueness conditions in this factorization we refer the reader to Chapter 4 of [AM69].

Definition 12.3.1. Let A be any commutative ring with identity, and \mathfrak{q} be an ideal of A . \mathfrak{q} is *primary* if \mathfrak{q} is proper, and $xy \in \mathfrak{q}$ implies that $x \in \mathfrak{q}$ or else $y^k \in \mathfrak{q}$, for a suitable k . It is readily seen that \mathfrak{q} is primary if and only if, in A/\mathfrak{q} , every zero-divisor is nilpotent. Thus, if \mathfrak{q} is primary, $\sqrt{\mathfrak{q}}$ is a prime ideal.

Before getting to some examples, let's make a basic observation about primary ideals. The proof is already done, more or less.

Proposition 12.3.2. *If \mathfrak{q} is a primary ideal of A then $\sqrt{\mathfrak{q}}$ is a prime ideal, and (obviously) the smallest prime ideal containing \mathfrak{q} .*

Definition 12.3.3. If \mathfrak{q} is a primary ideal of A and $\mathfrak{p} = \sqrt{\mathfrak{q}}$, then \mathfrak{q} is said to be \mathfrak{p} -primary.

Now some examples to disabuse the reader of possible misconceptions.

Examples 12.3.4. (a) In the ring of integers the nonzero primary ideals are precisely those generated by the power of a prime.

(b) Let K be any field, $A = K[T_1, T_2]$, with \mathfrak{q} the ideal generated by T_1 and T_2^2 . Now, $A/\mathfrak{q} = K[T]/(T^2)$; it is easy to verify that in A/\mathfrak{q} every zero-divisor is nilpotent. Thus, \mathfrak{q} is primary; its radical is $\mathfrak{p} = (T_1, T_2)$. Notice that $\mathfrak{p}^2 < \mathfrak{q} < \mathfrak{p}$, so that \mathfrak{q} itself is not the power of a prime ideal.

Therefore, a primary ideal need not be the power of a prime ideal.

(c) Conversely, the power of a prime ideal need not be primary.

Let K be a field, and $A = K[T_1, T_2, T_3]/(T_1T_2 - T_3^2)$. We shall denote the cosets of T_1, T_2 and T_3 by x, y and z , respectively. Thus, $A = K[x, y, z]$. Let $\mathfrak{p} = (x, z)$; this is a prime ideal, as $A/\mathfrak{p} = K[T]$, which is an integral domain. Next, $xy = z^2 \in \mathfrak{p}^2$, but $x \notin \mathfrak{p}^2$, and $y \notin \sqrt{\mathfrak{p}^2} = \mathfrak{p}$, which says that \mathfrak{p}^2 is not primary.

By contrast, we do have the following result.

Exercise 12.3.5. If $\sqrt{\mathfrak{r}}$ is a maximal ideal, then \mathfrak{r} is a primary ideal. In particular, all the powers of a maximal ideal are primary.

To establish the primary decomposition of ideals of a Noetherian ring, it's useful to introduce the following concept.

Definition 12.3.6. An ideal \mathfrak{r} of the ring A is said to be *irreducible* if $\mathfrak{r} = \mathfrak{s}_1 \cap \mathfrak{s}_2$ (for any ideals \mathfrak{s}_i of A) implies that $\mathfrak{r} = \mathfrak{s}_1$ or $\mathfrak{r} = \mathfrak{s}_2$.

Lemma 12.3.7. In a Noetherian ring A , every ideal is the intersection of a finite number of irreducible ideals.

Proof. By contradiction: if there is an ideal of A which is not a finite intersection of irreducible ones, then by the maximality principle, there is a maximal one \mathfrak{r} with this property. Now, \mathfrak{r} must be reducible, so write $\mathfrak{r} = \mathfrak{s}_1 \cap \mathfrak{s}_2$, so that both \mathfrak{s}_i 's properly contain \mathfrak{r} . But then each \mathfrak{s}_i can be written as a finite intersection of irreducible ideals, and then so can \mathfrak{r} , a contradiction. ■

Now we have the existence of primary decompositions. There are several uniqueness conditions that can be attached to this kind of decomposition. For details we refer the reader to Chapter 4 of [AM69]. See as well the concluding exercise in this section, and Proposition 12.4.2, in the next one.

Theorem 12.3.8. In a Noetherian ring A every ideal can be expressed as the intersection of a finite number of primary ideals.

Proof. By Lemma 12.3.7, all we need to do is show that every irreducible ideal in a Noetherian ring is primary.

Next, by passing to factor rings, it suffices to show that if $\{0\}$ is an irreducible ideal it is also primary. So suppose that $xy = 0$, but $x \neq 0$; consider the chain

$$\text{Ann}(y) \leq \text{Ann}(y^2) \leq \dots$$

From the a.c.c. we conclude that, for some positive integer n , $\text{Ann}(y^n) = \text{Ann}(y^{n+1}) = \dots$. This means that $Ax \cap Ay^n = \{0\}$; for if $a = bx = cy^n$, then $ay = 0$ and so $cy^{n+1} = 0$, whence $a = cy^n = 0$. Since $\{0\}$ is irreducible, it follows that $Ay^n = 0$ (because $Ax \neq 0$), and $y^n = 0$. ■

Hot Air 12.3.9. *Economy!* If \mathfrak{q} is a primary ideal and $\mathfrak{p} = \sqrt{\mathfrak{q}}$, then \mathfrak{p} is the least prime ideal containing \mathfrak{q} . Thus in the decomposition of Theorem 12.3.8, one can reduce the decomposition by dropping those primary ideals whose radicals are not minimal with respect to containing the ideal in question.

This is an informal account, and the reader ought to think this through.

With a little more work we get some nice corollaries. We leave the proof of the following as an exercise.

Proposition 12.3.10. *In a Noetherian ring A , every ideal \mathfrak{r} contains a power of its radical.*

Proof. Pick a (finite) generating set for $\sqrt{\mathfrak{r}}$. Then select a natural number m , large enough so that $x^m \in \mathfrak{r}$, for each $x \in \sqrt{\mathfrak{r}}$. The reader should then note that $(\sqrt{\mathfrak{r}})^m \leq \mathfrak{r}$. ■

Corollary 12.3.11. *In a Noetherian ring A , $n(A)^r = \{0\}$, for some natural number r . (That is, $n(A)$ is a nilpotent ideal.)*

Proof. Let $\mathfrak{r} = \{0\}$ in 12.3.10; $\sqrt{0} = n(A)$. ■

Corollary 12.3.12. *Suppose that \mathfrak{m} is a maximal ideal of the Noetherian ring A , and \mathfrak{q} is any ideal of A . The following are equivalent:*

- (a) \mathfrak{q} is \mathfrak{m} -primary.
- (b) $\sqrt{\mathfrak{q}} = \mathfrak{m}$.
- (c) $\mathfrak{m}^k \leq \mathfrak{q} \leq \mathfrak{m}$, for a suitable k .

Proof. By definition (a) implies (b), and the reverse follows from Exercise 12.3.5. Proposition 12.3.10 gives that (b) implies (c), and by computing radicals, (c) implies (b). ■

At long last we have the converse promised in 12.1.19.

Proposition 12.3.13. *Suppose that A is a ring. Then A is Artinian if and only if it is Noetherian and has zero Krull dimension.*

Proof. The necessity has been proved by application of Proposition 12.1.18. If the Krull dimension of A is zero and A is Noetherian, then any primary decomposition of the zero ideal of A (which exists by Theorem 12.3.8 can be rewritten as an intersection of a finite number of powers of maximal ideals, and therefore $\{0\}$ is a product of maximal ideals. Now invoke Corollary 12.1.16, and conclude that A is Artinian. ■

We conclude with an exercise which generalizes Lemma 10.2.2.

Exercise 12.3.14. Suppose that A is a ring, and S is a multiplicative system. Then the map $\mathfrak{q} \mapsto S^{-1}\mathfrak{q}$ defines a one-to-one, order-preserving correspondence between the primary ideals of A which do not meet S , and the set of all primary ideals of $S^{-1}A$. Moreover, if \mathfrak{q} is a \mathfrak{p} -primary ideal, then $\mathfrak{p} \cap S = \emptyset$ if and only if $\mathfrak{q} \cap S = \emptyset$, and then $S^{-1}\mathfrak{q}$ is $S^{-1}\mathfrak{p}$ -primary in $S^{-1}A$.

12.4 Dedekind Domains. In this section we investigate the rings which are prominent in the study of algebraic number theory, the Dedekind domains.

Definition 12.4.1. A Noetherian integral domain of Krull dimension 1, which is integrally closed is called a *Dedekind domain*. **All references to ring dimension in this section are to Krull dimension.**

It is immediate from the definition, and Exercise 10.3.8, that every principal ideal domain is a Dedekind domain. In some sense, as we shall see in the main theorem, Dedekind domains are a natural generalization of PID's. We already see this in the first result.

Proposition 12.4.2. *Suppose that A is a Noetherian (integral) domain of dimension one. Then every nonzero ideal of A can be uniquely expressed as a product of primary ideals, whose radicals are distinct.*

Proof. According to Theorem 12.3.8, if τ is a nonzero ideal of A , it is an intersection of a finite number of primary ideals $\mathfrak{q}_1, \dots, \mathfrak{q}_k$. Let $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$. Suppose, as indicated in 12.3.9, that the \mathfrak{p}_i are minimal with respect to containing τ . As the dimension of A is 1, each \mathfrak{p}_i is maximal, and therefore any two are comaximal (meaning that any two generate A as an ideal).

We leave it to the reader to verify that if the radicals of two ideals are comaximal, then so are the two ideals themselves. Also, that if a finite number of ideals are pairwise comaximal, then their intersection is, in fact their product. This shows existence of the decomposition mentioned in the proposition.

To establish uniqueness, suppose that

$$\tau = \mathfrak{q}_1 \cdots \mathfrak{q}_k = \mathfrak{t}_1 \cdots \mathfrak{t}_m,$$

so that each \mathfrak{q}_i and each \mathfrak{t}_j is primary, and their radicals are distinct. As was just observed, τ is also the intersection of the \mathfrak{q}_i and \mathfrak{t}_j , respectively. Let $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$ and $\mathfrak{n}_j = \sqrt{\mathfrak{t}_j}$. Observe that the \mathfrak{p}_i and \mathfrak{n}_j are all maximal ideals, and (unless some \mathfrak{p}_i equals some \mathfrak{n}_j) they are pairwise comaximal.

Suppose now that \mathfrak{p}_1 does not equal any of the \mathfrak{n}_j . Localize at \mathfrak{p}_1 ; i.e., let $S = A \setminus \mathfrak{p}_1$. Then

$$S^{-1}\tau = S^{-1}\mathfrak{q}_1 \cap \cdots \cap S^{-1}\mathfrak{q}_k = S^{-1}\mathfrak{t}_1 \cap \cdots \cap S^{-1}\mathfrak{t}_m,$$

and according to Exercise 12.3.14, all $S^{-1}\mathfrak{t}_j = S^{-1}A$, while $S^{-1}\mathfrak{q}_i = S^{-1}A$, for all $i \geq 2$. Hence, $S^{-1}\tau = S^{-1}\mathfrak{q}_1 = S^{-1}A$, which is absurd. Thus, each \mathfrak{p}_i equals some \mathfrak{n}_j , and it is quickly understood that $k = m$, and that (after some reindexing) $\mathfrak{n}_i = \mathfrak{p}_i$.

Now use induction, and once again appeal to Exercise 12.3.14, to conclude that the primaries themselves are unique. ■

The rings described in the following corollary are the Dedekind domains, as we shall see soon enough.

Corollary 12.4.3. *Suppose that A is a Noetherian domain of dimension one, in which every primary ideal is a power of its radical. Then every nonzero ideal of A can be uniquely expressed as a product of prime ideals.*

We now take a brief aside into a review of discrete valuation rings. Mainly, we need to recall Theorem 11.3.9; let us paraphrase: suppose that A is a valuation ring; then it is a discrete valuation ring if and only if it is a PID with (up to associates) one irreducible element. Indeed, from the proof of that theorem it becomes clear that, if A is a discrete valuation ring, then every nonzero ideal is a power of the maximal ideal.

We generalize slightly, and get many more points of view.

Lemma 12.4.4. *Suppose that A is a Noetherian local domain of dimension one; let \mathfrak{m} be its maximal ideal, and $A/\mathfrak{m} = K$ be its residue field. Then the following are equivalent.*

- (a) A is a discrete valuation ring.
- (b) A is integrally closed.
- (c) \mathfrak{m} is a principal ideal.
- (d) $\dim_K(\mathfrak{m}/\mathfrak{m}^2) = 1$ (as a K -vector space).
- (e) Every nonzero proper ideal of A is a power of \mathfrak{m} .
- (f) There exists a $q \in A$, so that every nonzero ideal of A is of the form Aq^k , for some integer $k \geq 0$.

Proof. Observe at the outset that if \mathfrak{q} is a nonzero ideal, then $\sqrt{\mathfrak{q}} = \mathfrak{m}$, because the Krull dimension is 1, and so (by Corollary 12.3.12), \mathfrak{q} is primary and contains some power of \mathfrak{m} . In addition, by Nakayama's Lemma (Exercise 9.1.14), $\mathfrak{m}^n \neq \mathfrak{m}^{n+1}$, for all $n \in \mathbb{N}$. Now to the equivalence of (a) through (f).

Theorem 11.3.9 gives that (a) implies (b), (c), (e) and (f), and, with a little thought, that (f) implies (a).

(b) \Rightarrow (c): pick $a \in \mathfrak{m}$, $a \neq 0$. Now, by Proposition 12.3.10, there is a positive integer n for which $\mathfrak{m}^{n+1} \leq Aa$, but $\mathfrak{m}^n \not\leq Aa$, so we choose $b \in \mathfrak{m}^n$ not divisible by a . Letting $x = a/b$, observe that $x^{-1} \notin A$, and so x^{-1} cannot be integral over A .

We claim that $x^{-1}\mathfrak{m} \not\leq \mathfrak{m}$. For suppose the contrary; then, since $(b/a)c = bc/a \in \mathfrak{m}$, for all $c \in \mathfrak{m}$, \mathfrak{m} is naturally an $A[x^{-1}]$ -module, which is then *faithful* as an $A[x^{-1}]$ -module. \mathfrak{m} is finitely generated as an A -module (because A is Noetherian), all of which contradicts Lemma 10.3.3(d). Thus $x^{-1}\mathfrak{m} = A$, since by the choice of b , $bc/a \in A$; which means that $\mathfrak{m} = Ax$, with $x \in A$, is the sought-after generator.

(c) \Rightarrow (d): As we have observed $\mathfrak{m}^2 \neq \mathfrak{m}$. Now $\mathfrak{m}/\mathfrak{m}^2$ is a K -vector space (how?), generated by $\mathfrak{m}^2 + x$, where $\mathfrak{m} = Ax$. That's sufficient.

(d) \Rightarrow (c) \Rightarrow (e): If \mathfrak{r} is a nonzero proper ideal of A , then $\mathfrak{m}^k \leq \mathfrak{r} \leq \mathfrak{m}$, for a suitable k . Since $\dim_K(\mathfrak{m}/\mathfrak{m}^2) = 1$, it is easily seen that, if $a \in \mathfrak{m} \setminus \mathfrak{m}^2$, then $Aa = \mathfrak{m}$. Assume that j is the largest number for which $\mathfrak{r} \leq \mathfrak{m}^j$. Then choose $y \in \mathfrak{r} \setminus \mathfrak{m}^{j+1}$; then $y = ra^j$ (as a^j generates \mathfrak{m}^j), and r is a unit of A . But this means that $\mathfrak{r} = \mathfrak{m}^j$.

(e) \Rightarrow (f): Any $y \in \mathfrak{m} \setminus \mathfrak{m}^2$ must generate \mathfrak{m} . This is sufficient to prove this implication.

The proof of the lemma is complete. ■

The lemma proves most of the following theorem:

Theorem 12.4.5. *Suppose that A is a Noetherian domain of dimension one. The following are then equivalent.*

- (a) A is a Dedekind domain.
- (b) Every nonzero primary ideal is a power of its radical.
- (c) Every nonzero ideal is uniquely expressible as a product of prime ideals.
- (d) Every localization $A_{\mathfrak{p}}$ is a discrete valuation ring.

Proof. By Theorem 10.3.15, A is integrally closed if and only if each localization $A_{\mathfrak{p}}$ of A is integrally closed. So the equivalence of (a) and (d) follows immediately from Lemma 12.4.4.

To see that (b) and (d) are equivalent, invoke the lemma, as well as Corollary 10.2.3(b) and Exercise 12.3.14.

From Corollary 12.4.3, (b) implies (c). As to the converse, if (c) holds then, in particular, every primary ideal \mathfrak{q} is a product of prime ideals $\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_k$. Since each \mathfrak{p}_i contains $\sqrt{\mathfrak{q}}$, and the dimension of A is 1, it follows that $\mathfrak{p}_i = \sqrt{\mathfrak{q}}$, for each i . ■

Hot Air 12.4.6. *Caution!* In some sense then, Dedekind domains are a natural extension of unique factorization to ideals (Corollary 12.4.3). However, a unique factorization domain need not be a Dedekind domain. In $\mathbb{Z}[T]$, which is Noetherian and integrally closed (by Exercise 10.3.8), one has the chain of prime ideals $\{0\} < (T) < (2, T)$, making the Krull dimension at least two.

Remark 12.4.7. To prove that in an algebraic number field (that is to say, a finite algebraic extension of \mathbb{Q}) the integral closure of \mathbb{Z} is a Dedekind domain, we are now able to prove all we need, except the Noetherian feature, which requires some elements of field theory. For this reason we postpone this result until the next chapter.

The final characterization of Dedekind domains is in terms of the so-called fractional ideals, which we will now define. This concept leads to the notions of “class groups” and “class numbers”, which we will merely brush against – and then brush aside.

Definition 12.4.8. Let A be an integral domain and $K = qA$. An A -submodule \mathfrak{m} of K , so that $x\mathfrak{m} \leq A$, for some nonzero $x \in A$, is called a *fractional ideal* of A . Clearly, each ideal of A is a fractional ideal. Note that for each $x = a/b \in K$ (with $a, b \in A$), the A -submodule Ax is a fractional ideal, as $bAx = Aa \leq A$. This is the *principal fractional ideal generated by x* . Since a fractional ideal of A is an A -submodule of K , it should be obvious that Ax is the smallest fractional ideal containing x .

Observe also that if \mathfrak{m} is a fractional ideal then, if $x\mathfrak{m} \leq A$, with $x \in A$, then $\mathfrak{r} = x\mathfrak{m}$ is an ideal of A , and that $\mathfrak{m} = x^{-1}\mathfrak{r}$.

An A -submodule \mathfrak{m} of K is *invertible*, if there exists an A -submodule \mathfrak{n} of K , for which $\mathfrak{m}\mathfrak{n} = A$ (where $\mathfrak{m}\mathfrak{n}$ is the A -submodule generated by all xy , with $x \in \mathfrak{m}$ and $y \in \mathfrak{n}$).

For any A -submodule \mathfrak{m} of K , let $[A : \mathfrak{m}]$ denote the set of all $x \in qA$ such that $x\mathfrak{m} \leq A$. It should be clear that $[A : \mathfrak{m}]$ is always an ideal of A ; it could be trivial. If \mathfrak{m} is invertible (with $\mathfrak{m}\mathfrak{n} = A$), then observe that

$$\mathfrak{n} \leq [A : \mathfrak{m}] = [A : \mathfrak{m}]\mathfrak{m}\mathfrak{n} \leq A\mathfrak{n} \leq \mathfrak{n},$$

whence $\mathfrak{n} = [A : \mathfrak{m}]$.

The following proposition describes the basic relationships between fractional and invertible submodules of K .

Proposition 12.4.9. *Suppose that A is an integral domain and K is its field of fractions. Then we have the following:*

- (a) *Every finitely generated A -submodule of K is a fractional ideal.*
- (b) *Every invertible submodule of K is finitely generated (and therefore a fractional ideal).*
- (c) *If A is Noetherian, then the converse of (a) holds.*

Proof. (a) If $\mathfrak{m} = Ax_1 + \cdots + Ax_k$, and we write $x_i = a_i/b_i$ (for each i , with $a_i, b_i \in A$), then by taking $b = b_1b_2 \cdots b_k$, we may write each $x_i = c_i/b$, with suitable $c_i \in A$. Then note that $b\mathfrak{m} = Ac_1 + \cdots + Ac_k \leq A$.

(b) If $\mathfrak{m}\mathfrak{n} = A$, write $1 = x_1y_1 + \cdots + x_my_m$, with the $x_i \in \mathfrak{m}$ and the $y_i \in \mathfrak{n}$. We showed in **12.4.8** that $\mathfrak{m} = [A : \mathfrak{n}]$. Now, if $a \in \mathfrak{m}$, then $a = x_1(y_1a) + \cdots + x_m(y_ma)$, and each $y_ia \in A$, proving that the x_i generate \mathfrak{m} .

(c) Each fractional ideal is (by definition) of the form $a^{-1}\mathfrak{r}$, and \mathfrak{r} is finitely generated, if A is Noetherian. ■

The rest of the section is concerned with the converse of (b) in the preceding proposition: with A Noetherian, when is every fractional ideal invertible? Obviously, an invertible module cannot be trivial.

First some further generalities wafting aloft.

Hot Air 12.4.10. *Keeping it all Straight.* Let A be an integral domain, $K = qA$. We use $\text{pr}(A)$, $\text{frac}(A)$, $\text{fg}(A)$ and $\text{inv}(A)$ to denote (respectively) the sets of principal fractional ideals, all fractional ideals, all finitely generated A -submodules of K , and all invertible A -submodules of K . Since it is clear that every principal fractional ideal Ax is invertible (its inverse being Ax^{-1}), we have that

$$\text{pr}(A) \subseteq \text{inv}(A) \subseteq \text{fg}(A) \subseteq \text{frac}(A).$$

The latter two inclusions come from Proposition **12.4.9**; equality in the last one occurs if A is Noetherian.

Moreover, with respect to the product $\mathfrak{m}\mathfrak{n}$ of A -submodules of K , observe that the product of two fractional ideals is a fractional ideal. So, if we view the sets introduced here as semigroups (with the type of a single binary operation), then the above inclusions are as subsemigroups. $\text{inv}(A)$ and $\text{pr}(A)$ are, in fact, subgroups of $\text{frac}(A)$, where the identity is A .

Now we show that “being invertible” is a local property, for finitely generated submodules of the field of fractions.

Proposition 12.4.11. *Let \mathfrak{m} be an A -submodule of the field qA of fractions of the domain A . The following are equivalent.*

- (a) \mathfrak{m} is invertible.
- (b) \mathfrak{m} is finitely generated and $\mathfrak{m}_{\mathfrak{p}}$ is invertible, for each prime ideal \mathfrak{p} of A .
- (c) \mathfrak{m} is finitely generated and $\mathfrak{m}_{\mathfrak{q}}$ is invertible, for each maximal ideal \mathfrak{q} of A .

Proof. (a) implies (b): We have already seen that an invertible ideal is finitely generated. If $\mathfrak{m}\mathfrak{n} = A$, then also $A_{\mathfrak{p}} = (\mathfrak{m}\mathfrak{n})_{\mathfrak{p}} = \mathfrak{m}_{\mathfrak{p}}\mathfrak{n}_{\mathfrak{p}}$.

Since it is obvious that (b) implies (c), we move on to show that (c) implies (a). Note that $\mathfrak{r} \equiv \mathfrak{m}[A : \mathfrak{m}]$ is an ideal of A , and if it is proper, then it is contained in a maximal ideal \mathfrak{q} of A . Now observe that $\mathfrak{r}_{\mathfrak{q}}$ is a proper ideal, while $\mathfrak{r}_{\mathfrak{q}} = \mathfrak{m}_{\mathfrak{q}}[A : \mathfrak{m}]_{\mathfrak{q}}$; the reader should also note that $[A : \mathfrak{m}]_{\mathfrak{q}} = [A_{\mathfrak{q}} : \mathfrak{m}_{\mathfrak{q}}]$; it is here that the assumption that \mathfrak{m} is finitely generated is required. By assumption,

$$\mathfrak{r}_{\mathfrak{q}} = \mathfrak{m}_{\mathfrak{q}}[A_{\mathfrak{q}} : \mathfrak{m}_{\mathfrak{q}}] = A_{\mathfrak{q}},$$

and this is a contradiction. ■

All of which brings us to the local version of the characterization we have been pursuing.

Lemma 12.4.12. *Suppose that A is a local domain. Then A is a discrete valuation ring if and only if every nonzero fractional ideal of A is invertible.*

Proof. (Necessity) Let x be a generator of the maximal ideal \mathfrak{q} of A ; (recall Lemma 12.4.4). Suppose that \mathfrak{m} is a nonzero fractional ideal of A . As has already been noted, $\mathfrak{m} = y^{-1}\mathfrak{r}$, for a suitable ideal \mathfrak{r} of A (also nontrivial) and $y \in A$. However, by Lemma 12.4.4, \mathfrak{r} is of the form $\mathfrak{r} = Ax^r$, that is to say, it is principal, which makes \mathfrak{m} principal. Thus, \mathfrak{m} is invertible.

(Sufficiency) First, since every nonzero ideal is invertible, and therefore finitely generated, it follows that A is Noetherian. It suffices to show that every nonzero ideal of A is a power of the maximal ideal \mathfrak{q} of A . (Because, if this is so, then the Krull dimension is automatically 1, and we can apply Lemma 12.4.4.)

Suppose that this is false. If there is a nonzero ideal which is not a power of \mathfrak{q} , then we may select a maximal one \mathfrak{r} ; clearly, $\mathfrak{r} < \mathfrak{q}$. Let \mathfrak{n} be a fractional ideal of A , so that $\mathfrak{q}\mathfrak{n} = A$. Then $\mathfrak{r}\mathfrak{n} < \mathfrak{q}\mathfrak{n} = A$ (properly, otherwise $\mathfrak{r} = \mathfrak{q}$, by cancellation), and also $\mathfrak{r} \leq \mathfrak{r}\mathfrak{n}$. Mark well that $\mathfrak{r}\mathfrak{n}$ is an (ordinary) ideal of A ! But \mathfrak{r} is also invertible, so that the identity $\mathfrak{r}A = \mathfrak{r} = \mathfrak{r}\mathfrak{n}$ and cancellation would imply that $\mathfrak{n} = A$, which is impossible. Thus, $\mathfrak{r} < \mathfrak{r}\mathfrak{n}$; in view of the maximality of \mathfrak{r} among nonpowers of \mathfrak{q} , we conclude that $\mathfrak{r}\mathfrak{n} = \mathfrak{q}^k$, for some positive integer k . From this it follows that

$$\mathfrak{r} = \mathfrak{r}\mathfrak{n}\mathfrak{q} = \mathfrak{q}^{k+1},$$

a contradiction. ■

And now finally, the global counterpart of Lemma 12.4.12, which is the sought-after characterization of Dedekind domains.

Theorem 12.4.13. *Let A be an integral domain. Then A is a Dedekind domain if and only if every nonzero fractional ideal of A is invertible.*

Proof. (Necessity) Suppose that \mathfrak{m} is a nonzero fractional ideal of A . As A is Noetherian, \mathfrak{m} is finitely generated. For each prime ideal \mathfrak{p} of A , $\mathfrak{m}_{\mathfrak{p}}$ is a nonzero fractional ideal of the discrete valuation ring $A_{\mathfrak{p}}$, and therefore invertible, according to Lemma 12.4.12. Now, by Proposition 12.4.11, it follows that \mathfrak{m} is invertible.

(Sufficiency) If every nonzero fractional ideal is invertible, then, in particular, so is every ideal, making A Noetherian. To complete the proof, it suffices to show that each localization $A_{\mathfrak{p}}$ is a discrete valuation ring (because that shows that the dimension is 1, at which point Theorem 12.4.5 can be applied).

To use Lemma 12.4.12, it suffices to show that every nonzero ideal of $A_{\mathfrak{p}}$ is invertible. Now, if \mathfrak{s} is an ideal of $A_{\mathfrak{p}}$, then its contraction \mathfrak{s}^c to A is nonzero, hence invertible. As $\mathfrak{s} = (\mathfrak{s}^c)_{\mathfrak{p}}$, we may conclude, from Proposition 12.4.11, that \mathfrak{s} is invertible. ■

Hot Air 12.4.14. *More to Keep Straight.* Suppose that A is a Dedekind domain. In terms of the discussion in 12.4.10, what we have is that

$$\text{inv}(A) = \text{fg}(A)^{\#} = \text{frac}(A)^{\#}.$$

This is a group under multiplication of fractional ideals, called the *group of ideals* of A .

If $K^{\#}$ denotes the group of nonzero elements of $K = qA$, and U the group of units of A , then the map $\theta : K^{\#} \rightarrow \text{frac}(A)$, defined by $\theta(x) = Ax$, is a homomorphism, for which $\ker(\theta) = U$ and $\theta(K^{\#}) = \text{pr}(A)^{\#}$. Now form

$$\text{ic}(A) = \text{inv}(A)/\text{pr}(A),$$

and let $\mu_A : \text{inv}(A) \rightarrow \text{ic}(A)$ be the canonical homomorphism. If α denotes the inclusion of U in $K^\#$, then we obtain the following exact sequence:

$$0 \rightarrow U \xrightarrow{\alpha} K^\# \xrightarrow{\theta} \text{inv}(A) \xrightarrow{\mu_A} \text{ic}(A) \rightarrow 0.$$

The group $\text{ic}(A)$ is called the *ideal class group* of A . For instance, to say that $\text{ic}(A)$ is trivial, is to say that A is a PID. If K is an algebraic number field, and A is the integral closure of \mathbb{Z} in K , it will turn out (as has already been telegraphed) that A is a Dedekind domain. Now, it can be shown that under these circumstances, $\text{ic}(A)$ is a finite group; its order is referred to as the *class number of the field K* .

It also turns out that U is finitely generated (as an abelian group). In fact, it can be shown that the elements of U of finite order are precisely the roots of unity which lie in K , and that these elements form a cyclic group Z . Note that it then follows that U/Z is a finitely generated free abelian group.

A great deal is known about this situation, and about the number of free generators for U/Z . We refer the reader to algebraic number theory for further information on this subject.

We now conclude the section and chapter with a number of exercises.

Exercise 12.4.15. Suppose that A is a Dedekind domain. If $f(T)$ is a polynomial over A , let $c(f)$ be the ideal generated in A by the coefficients. Prove “Gauss’ Lemma”: $c(fg) = c(f)c(g)$, for any two polynomials over A .

(Hint: localize at each maximal ideal.)

Exercise 12.4.16. Suppose that A is a local domain, in which the maximal ideal \mathfrak{m} is principal, and that $\bigcap_{n \in \mathbb{N}} \mathfrak{m}^n = \{0\}$. Prove that A is a discrete valuation ring.

The following three exercises have to do with torsion freeness over an integral domain, and the relationship to flatness. Compare with Exercise 8.2.12. Recall that a module M over the integral domain A is *torsion free* if $ax = 0$ (with $a \in A$ and $x \in M$) implies that $a = 0$ or $x = 0$. Note that the subset

$$T(M) = \{x \in M : ax = 0, \text{ for some } a \neq 0\}$$

is an A -submodule of M , if A is any integral domain.

In the next exercise, A stands for an integral domain.

Exercise 12.4.17. Suppose that M is an A -module and that S is a multiplicative system of A . Prove that $S^{-1}(T(M)) = T(S^{-1}M)$. Now show the following are equivalent.

- (a) M is torsion free.
- (b) $M_{\mathfrak{p}}$ is torsion free, for each prime ideal \mathfrak{p} .
- (c) $M_{\mathfrak{q}}$ is torsion free, for each maximal ideal \mathfrak{q} .

Exercise 12.4.18. Suppose that A is a Noetherian ring, and M is a finitely generated A -module.

- (a) If A is local, prove that M is free if and only if it is flat.
- (b) Now prove that the following are equivalent.
 - (i) M is a flat A -module;

- (ii) $M_{\mathfrak{p}}$ is a free $A_{\mathfrak{p}}$ -module, for each prime ideal \mathfrak{p} .
- (iii) $M_{\mathfrak{q}}$ is a free $A_{\mathfrak{q}}$ -module, for each maximal ideal \mathfrak{q} .

Put the previous two exercises together to show:

Exercise 12.4.19. Suppose that A is a Dedekind domain, and that M is a finitely generated A -module. Prove that M is flat if and only if M is torsion free.

Exercise 12.4.20. If A is a Dedekind domain, then prove that the lattice of all ideals of A is distributive.

(Hint: localize!)

Exercise 12.4.21. *A Chinese Remainder Theorem.* Suppose that A is a Dedekind domain. Suppose that $\mathfrak{r}_1, \dots, \mathfrak{r}_m$ are ideals of A , and $a_1, \dots, a_m \in A$. The congruences $x \equiv a_i \pmod{\mathfrak{r}_i}$ ($1 \leq i \leq m$) have a solution if and only if $a_i \equiv a_j \pmod{\mathfrak{r}_i + \mathfrak{r}_j}$, whenever $i \neq j$.

Exercise 12.4.22. Suppose that A is a Dedekind domain, having only a finite number of prime ideals. Prove that A is a PID.

(Hint: It suffices to show that each prime ideal is principal. To do that, apply the Chinese Remainder Theorem as follows: let $\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_n$ be the distinct prime ideals of A ; apply 12.4.21 to \mathfrak{p}_i^2 and the other \mathfrak{p}_j , choosing an element of $\mathfrak{p}_i \setminus \mathfrak{p}_i^2$.)

Exercise 12.4.23. In a Dedekind domain every ideal can be generated by one or two elements.

(Hint: Use the preceding exercise to show that, if A is a Dedekind domain, then A/\mathfrak{r} is a principal ideal ring (though not a domain, unless it is also a field). If \mathfrak{r} is principal, fine; otherwise, let $x \in \mathfrak{r}$, and then apply the previous observation about A modulo a nonzero ideal to Ax .)

Back to homological ideas for a finish:

Exercise 12.4.24. Suppose that A is a commutative ring with identity, and that \mathfrak{r} is an ideal of A containing a regular element. Prove that \mathfrak{r} is projective (as an A -module) if and only if \mathfrak{r} is invertible (as a fractional ideal of its classical ring of fractions). Deduce from this that, if A is an integral domain, then it is a Dedekind domain if and only if every ideal of A is projective.

13. Field Theory: Galois Theory of Equations

In this chapter we give an account of the Galois theory of polynomial equations, with applications to radical extensions and to finite fields. The main reference for this chapter is [Ka69].

13.1 Galois Connexion of Fields. Throughout this section F will be a fixed base field, over which we shall consider polynomials. The immediate aim is Galois' Fundamental Theorem linking certain extension fields with groups of automorphisms that fix the members of those fields.

Definition 13.1.1. Suppose that E is an extension field of F (which simply means that F is a subfield of E). For completeness we shall recall a few basic definitions. $z \in E$ is *algebraic over F* if there is a polynomial $f(T)$ with coefficients in F , so that $f(z) = 0$. If no such polynomial exists we say that E is *transcendental over F* .

E may be viewed as a vector space over F ; as such, we write $[E : F]$ for the dimension of E over F . If $[E : F]$ is finite we speak of E being a *finite* extension of F . Recall that every finite extension of F is necessarily algebraic. Recall also that, if $F \leq L$ and $L \leq E$, then $[E : F] = [E : L][L : F]$.

The first result should be familiar to the reader, we recall it, once again, for the sake of completeness.

Proposition 13.1.2. Suppose that E is an extension of F , and that z in E is algebraic over F . Then $I_z = \{f(T) \in F[T] : f(z) = 0\}$ is an ideal of $F[T]$, which has a unique monic generator, say $p(T)$. The subfield generated by F and z is $F[z]$, and $[F[z] : F] = n$, where n is the degree of $p(T)$. Moreover, $F[z] \cong F[T]/(p(T))$.

Proof. Let $e_z : F[T] \rightarrow E$ be the evaluation map at z ; $e_z(f(T)) = f(z)$. Then $\ker(e_z) = I_z$ and $e_z(F[T]) = F[z]$. ■

Definition 13.1.3. Under the terms and notation of Proposition 13.1.2, the unique monic generator $p(T)$ of I_z is called the *minimum polynomial* of z . It is an irreducible polynomial over F . Its degree is also called the *degree of z over F* .

Definition 13.1.4. Suppose that E is a field extension of F . The *Galois group* of E over F , denoted $\text{Gal}(E/F)$, is the group (under composition of maps) of all automorphisms of E which fix all the elements of F .

Now, observe right away that $\text{Gal}(E/F)$ could be trivial, even though E is a proper extension of F . Let $E = \mathbb{Q}[\sqrt[3]{2}]$, with $F = \mathbb{Q}$. Since every member of $\text{Gal}(E/F)$, in this instance, must send a root of $T^3 - 2$ to another root of this polynomial, and E contains no other root of $T^3 - 2$, it follows that $\text{Gal}(E/F) = \{1\}$.

Although, for the most part, our applications of Galois Theory will be to finite extensions, there is no reason, in principle, to limit consideration to such cases. Indeed the Galois group of the algebraic closure of \mathbb{Q} over \mathbb{Q} is an important infinite group.

Example 13.1.5. Let n be a positive integer. The polynomial $\Phi_n(T)$ one gets as the product of all terms $T - z$, where z ranges over all primitive n -th complex roots of unity, is called the n -th

cyclotomic polynomial. In general, $\Phi_n(T)$ is not irreducible over the rationals; $\Phi_8(T)$, for example, is not. $\Phi_n(T)$ came up briefly in Exercise 7.4.21. If p is a prime integer then

$$\Phi_p(T) = T^{p-1} + \dots + T + 1.$$

By a clever application of Eisenstein's Criterion, it can be shown that $\Phi_p(T)$ is irreducible over \mathbb{Q} .

Consider now a primitive p -th root of 1; that is to say, since p is prime, any root $z \neq 1$. The roots of $\Phi_p(T)$ are $\{1, z, \dots, z^{p-1}\}$, and they form a cyclic group under multiplication. The extension $\mathbb{Q}[z] \cong \mathbb{Q}[T]/(\Phi_p(T))$ is of dimension $p - 1$, and the roots $\{1, z, \dots, z^{p-2}\}$ form a basis for the extension over \mathbb{Q} .

Recall that the automorphism group of a cyclic group of order p is isomorphic to the cyclic group of order $p - 1$. The point is that we have an automorphism of the roots of $\Phi_p(T)$ of order $p - 1$, and this automorphism α is completely determined on the roots by $\alpha(z) = z^i$. One has to check that the extension by linearity of α (via the basis mentioned in the preceding paragraph) to $\mathbb{Q}[z]$, is a field automorphism. Once this is done it becomes clear that $\text{Gal}(\mathbb{Q}[z]/\mathbb{Q})$ is cyclic of order $p - 1$.

Definition & Remarks 13.1.6. *The Galois Connection.* Let E be a field extension of F , and $G = \text{Gal}(E/F)$. For each intermediate field K , let

$$K' = \{ g \in G : g(z) = z, \forall z \in K \};$$

for each subgroup H of G , let

$$H' = \{ z \in E : g(z) = z, \forall g \in H \}.$$

Then the reader will easily verify the following facts:

- (a) For any two intermediate fields $F_1 \leq F_2$, we have $F_2' \leq F_1'$.
- (b) For any two subgroups of G , $H_1 \leq H_2$ implies that $H_2' \leq H_1'$.
- (c) For any subgroup H of G , $H \leq H''$.
- (d) For any intermediate field K , $K \leq K''$.
- (e) $\{1\}' = E$; $F' = G$ and $E' = \{1\}$. (The reader will observe the conspicuous absence of the fourth item: $G' = F$. Just recall the example of $\mathbb{Q}[\sqrt[3]{2}]$! The fact that $G' = F$ fails, in general, serves as the intellectual push for the theory we are about to develop.)

The above also imply the following:

- (f) $H''' = H'$, for every subgroup H of G .
- (g) $K''' = K'$, for any intermediate field K .

The failure discussed in (e) gives rise to the next definition.

Definition 13.1.7. Suppose that E is a field extension of F . We say that E is a *Galois extension of F* if $G' = F$, where $G = \text{Gal}(E/F)$. To put it in a more expansive manner: E is a Galois extension of F if for each $z \in E \setminus F$ there exists a $g \in G$, so that $g(z) \neq z$.

We will also refer to the passage $H \mapsto H''$, for subgroups H of G (resp. $K \mapsto K''$, for subfields of E that contain F) as the *closure* of the subgroup H (resp. subfield K). A subgroup H (resp. subfield K , intermediate to E and F) is *closed* if $H = H''$ (resp. $K = K''$). There is a topology on G , under which the subgroups H which satisfy $H'' = H$ are (topologically) closed, and vice versa. We shall discuss this in the next chapter.

Here is a preliminary form of the fundamental Galois connection:

Theorem 13.1.8. *The maps $H \mapsto H'$ and $K \mapsto K'$, from the set of closed subgroups to the set of closed subfields and back, are mutually inverse bijections of these two sets, inverting the inclusion.*

Proof. That the maps invert inclusion follow from (a) and (b) of 13.1.6. Observe that a subgroup H of G (resp. intermediate field K) is closed if and only if $H = L'$, for some intermediate field L (resp. $K = A'$, for some subgroup A of G). Then it is clear that, on the two sets, as specified in the theorem, $H \mapsto H'$ and $K \mapsto K'$ are mutually inverse bijections. ■

The two lemmas which follow are fundamental to the eventual theorem on finite Galois extensions. For subgroups, $[G : H]$ denotes the index of H in G .

Lemma 13.1.9. *Suppose that $F \leq L \leq M \leq E$, all subfields, with $[M : L] = n$. Then $[L' : M'] \leq n$.*

Proof. By induction on n . The case $n = 1$ is, of course, trivial.

If there is a field L_o , properly situated between L and M , then by the inductive hypothesis $[L_o : L] \geq [L' : L'_o]$ and $[M : L_o] \geq [L'_o : M']$, whence it follows that

$$[M : L] = [M : L_o][L_o : L] \geq [L' : M'].$$

So we may as well assume that there are no subfields strictly between L and M , and that $M = L[u]$, for a suitable element u , which must be algebraic over L . Its minimum polynomial $p(T)$ over L has degree n .

Now, since the members of M' leave each element of M fixed, any two elements of L' which lie in the same coset modulo M' act the same way on u . So, for each coset C of L' , modulo M' , let Cu represent the result of applying any member of C to u ; this is unambiguously defined. For distinct cosets C and D , if $Cu = Du$, then each $\alpha \in C$ and each $\beta \in D$ act the same way on u , which means that $\alpha^{-1}\beta$ fixes u , and therefore M , contradicting the assumption that C and D are distinct. Conclusion: the number of cosets, that is to say, $[L' : M']$, is equal to the number of distinct images Cu . But each Cu must be a root of $p(T)$. It follows that $[L' : M'] \leq n$. ■

Lemma 13.1.10. *Suppose that $G = \text{Gal}(E/F)$, and $S \leq T$ are subgroups, with $[T : S] = n$. Then $[S' : T'] \leq n$.*

Proof. Suppose that $C = \alpha S$ is a coset, and that $x \in S'$. Since x is left fixed by any automorphism of S , it follows (as in the previous proof) that any two members of C act the same way on x . So, by setting $Cx = \alpha x$, we have a well defined action of the cosets of T modulo S , on S' .

By way of contradiction, suppose $[S' : T'] > n$. Then we may find $u_1, u_2, \dots, u_{n+1} \in S'$, linearly independent over T' . Let C_1, C_2, \dots, C_n denote the n distinct cosets of T modulo S . Consider the system of equations, in the unknowns x_i , given below:

$$\begin{aligned} (C_1 u_1)x_1 + (C_1 u_2)x_2 + \cdots + (C_1 u_{n+1})x_{n+1} &= 0 \\ (C_2 u_1)x_1 + (C_2 u_2)x_2 + \cdots + (C_2 u_{n+1})x_{n+1} &= 0 \\ &\dots\dots\dots \\ (C_n u_1)x_1 + (C_n u_2)x_2 + \cdots + (C_n u_{n+1})x_{n+1} &= 0. \end{aligned} \tag{!}$$

The coefficients of (!) all lie in $S' \leq E$, and so there must be a nontrivial solution in E . Pick one with the largest possible number of zeroes, and, after rearrangement, if necessary, assume that the solution is

$$(x_1, \dots, x_{n+1}) = (a_1, a_2, \dots, a_r, 0, \dots, 0),$$

where each $a_i \neq 0$. Without loss of generality, one can also assume that $a_1 = 1$.

Observe that not all the solutions can lie in T' ; for if this were so then, since one of the C_i is the identity coset, the equation corresponding to that coset would give a linear combination with coefficients in T' , contradicting the independence of the u_i over T' . So at least one of the a_i is not in T' , and without loss of generality, we may assume it is a_2 .

Now, there is an automorphism $\beta \in T$ which moves a_2 ; apply β to the equations in (!). The result is a system, the i -th equation of which is

$$\beta(C_i u_1) \beta x_1 + \cdots + \beta(C_i u_{n+1}) \beta x_{n+1} = 0.$$

However, by acting on the cosets C_1, C_2, \dots, C_n , β simply permutes them. Thus, the action of β on the system (!) reproduces the system, with the solution $(\beta a_1, \beta a_2, \dots, \beta a_r, 0, \dots, 0)$, and $\beta a_1 = 1$. Subtracting this from the first solution, one gets a nontrivial solution to (!), with a larger number of zeroes, a contradiction. ■

The two preceding lemmas establish the following theorem.

Theorem 13.1.11. *Suppose that $G = \text{Gal}(E/F)$.*

- (a) *If $F \leq L \leq M \leq E$ are intermediate fields, with L closed, and so that M is a finite extension of L , then M too is closed. Moreover, $[L' : M'] = [M : L]$.*
- (b) *If $S \leq T$ are subgroups of G , with S closed, and such that the index of S in T is finite, then T is also closed. Moreover, $[S' : T'] = [T : S]$.*

Proof. By the lemmas $[M'' : L''] \leq [L' : M'] \leq [M : L]$, and $L = L''$, by assumption. Since $M \leq M''$ (always), this implies that equalities hold throughout, proving (a). The proof of (b) is similar. ■

Corollary 13.1.12. *Suppose that E is an extension of the field F and $G = \text{Gal}(E/F)$. Then*

- (a) *any finite subgroup of G is closed;*
- (b) *if E is a Galois extension of F , and L is an intermediate field, so that L is finite over F , then E is Galois over L .*

Proof. (a) The subgroup $\{1\}$ is closed; apply Theorem 13.1.11(b).

(b) F is closed, by assumption. Apply 13.1.11(a). ■

Although quite anticlimactically, here is the traditional main theorem:

Theorem 13.1.13. (Fundamental Theorem of Galois Theory) *Suppose that E is a finite Galois extension of F , and $G = \text{Gal}(E/F)$. Then the priming operations set up a one-to-one correspondence between the intermediate subfields of E , and the subgroups of G , which inverts inclusion.*

For each pair of intermediate fields $L \leq M$, $[M : L] = [L' : M']$, and every intermediate subfield is closed. Likewise, for each pair of subgroups $S \leq T$ of G , $[T : S] = [S' : T']$, and each subgroup of G is closed.

If H is any subgroup of G , then $\text{Gal}(E/H') = H$. In particular, $|G| = [E : F]$.

Example 13.1.14. Suppose $E = k(T_1, T_2, \dots, T_n)$, the field of fractions of the polynomial ring $k[T_1, \dots, T_n]$, over the field k . Let G be the group of automorphisms of E obtained by all permutations of the indeterminates; this is, of course, a copy of S_n . Let F be the fixed field of G . Then, by arrangement, E is a Galois extension of F , and $\text{Gal}(E/F) = G$. So S_n is the Galois group of some Galois extension.

Since (by Cayley's Theorem) every finite group H is isomorphic to a subgroup of some S_n , it follows from Theorem 13.1.13, that H too is the Galois group of some Galois extension. This is cute, but not especially informative.

We conclude this section with a selection of exercises. [Ka69] contains most of these, and many other interesting ones.

Exercise 13.1.15. Suppose that E is a field extension of F , and that L and M are two intermediate fields. Let $L * M$ denote the set of sums of the form $x_1y_1 + \cdots + x_ky_k$, with the $x_i \in L$ and $y_i \in M$. Prove the following:

- (a) $L * M$ is a subring of E .
- (b) $q(L * M)$ is the subfield of E generated by L and M .
- (c) If L and M are algebraic over F , then $q(L * M) = L * M$, and the latter is algebraic over F .

Exercise 13.1.16. With the notation of the preceding exercise, suppose that both L and M are finite extensions of F . If

$$[L * M : F] = [L : F][M : F],$$

show that $L \cap M = F$. Prove that the converse is valid, provided that either $[L : F]$ or $[M : F]$ is 2. However, by considering cube roots of 2, show that the converse is false even if the dimensions $[L : F]$ and $[M : F]$ equal 3.

Exercise 13.1.17. Suppose that E is an extension of the field F , and that L is an intermediate field. If L is Galois over F , E is Galois over L , and each automorphism of L over F can be extended to E , then prove that E is Galois over F .

Exercise 13.1.18. Suppose that F is an infinite field, and $E = F(T)$, where T is an indeterminate. Prove that E is Galois over F .

(Hint: For each $a \in F$, the translation $T \mapsto T + a$, induces an element of $G = \text{Gal}(E/F)$. Now, if $f(T)/g(T)$ is in the fixed field of G , show that

$$h(T_1, T_2) = f(T_1)g(T_1 + T_2) - g(T_1)f(T_1 + T_2)$$

vanishes for all substitutions by elements of F . Since F is infinite, this means that the polynomial $h = 0$.)

Exercise 13.1.19. As in the previous exercise, F is a field, and $E = F(T)$. Let $G = \text{Gal}(E/F)$. Prove that

- (a) if $L > F$ is an intermediate field, then $[E : L]$ is finite. (Note that if $t = f(T)/g(T) \in L \setminus F$, then $tg(T) - f(T) = 0$, which means that the indeterminate T is algebraic over L , doesn't it?)
- (b) If F is infinite, show that the only closed subgroups of G are the finite subgroups and G itself.

Exercise 13.1.20. Prove that every automorphism of \mathbb{R} , the field of real numbers, is trivial.

(Hint: every rational number must be fixed. On the other hand, every automorphism of \mathbb{R} preserves order, and is, for that reason, continuous.)

The last exercise highlights the role of normal subgroups in the Galois correspondence.

Exercise 13.1.21. Suppose that E is a field extension of F , and that $G = \text{Gal}(E/F)$. An intermediate subfield L is said to be *stable* if, for each $\alpha \in G$, $\alpha L \leq L$ (which implies that $\alpha L = L$). (Thus, L is stable if and only if each automorphism of E over F induces, via restriction, an element of $\text{Gal}(L/F)$.) Prove that

- (a) if L is a stable subfield, then L' is a normal subgroup of G ;
- (b) if H is a normal subgroup of G then H' is a stable subfield of E .
- (c) If E is Galois over F and L is a stable subfield, then L is Galois over F .
- (d) If every automorphism of L over F extends to an automorphism of E , and L is stable, then $\text{Gal}(L/F) = G/L'$.

To summarize, in the Galois correspondence, of closed intermediate subfields vs. closed subgroups of G , stable subfields correspond to normal subgroups.

13.2 Splitting Fields. There are two results from elementary field theory, which are needed throughout. We state them here without proof. The first concerns the existence of roots in suitable extensions.

Proposition 13.2.1. *Suppose that F is a field and that $p(T) \in F[T]$ is an irreducible polynomial. Then there is a field extension E , so that $[E : F] = \deg(p)$, and $p(T)$ has a root in E .*

The second result is a crucial fact regarding extension of isomorphisms between fields.

Proposition 13.2.2. *Suppose that $f : F_1 \rightarrow F_2$ is an isomorphism of fields. Let $p_1(T) \in F_1[T]$ be an irreducible polynomial, and $p_2(T)$ be the polynomial over F_2 , obtained by applying f to each of the coefficients of $p_1(T)$. Let $E_i = F_i[u_i]$ ($i = 1, 2$), where u_i is a root of $p_i(T)$. Then there is an isomorphism $\tilde{f} : E_1 \rightarrow E_2$ which maps u_1 to u_2 , and extends f .*

Now to the main definition of this section.

Definition 13.2.3. Suppose that $f(T)$ is a polynomial over the field F , and that E is a field extension of F , containing all the roots of $f(T)$, and which is also the subfield generated by the roots of $f(T)$. Then we say that E is a *splitting field of $f(T)$ over F* . Another way to put all this is as follows: E is a splitting field of $f(T)$ over F if

- (i) $f(T)$ factors completely over E and,
- (ii) assuming that u_1, u_2, \dots, u_k are all the roots of $f(T)$ in E , then $E = F[u_1, \dots, u_k]$.

In tandem with Propositions 13.2.1 and 13.2.2 we have existence, followed by uniqueness of splitting fields.

Proposition 13.2.4. *Suppose that $f(T)$ is a polynomial over the field F . Then $f(T)$ has a splitting field over F .*

Proof. By induction on the degree of $f(T)$. If $\deg(f) = 1$, there is nothing to do; F itself is a splitting field. Now suppose that $g(T)$ is an irreducible factor of $f(T)$ over F , of degree > 1 . (Note: if all the irreducible factors are linear, again F is a splitting field, and we're finished.) If $\deg(g) < \deg(f)$, then $f(T) = g(T)h(T)$ over F , both of smaller degree, and (by induction) we may find, first, a splitting field L of $g(T)$, and then, considering $h(T)$ over L , a splitting field E of $h(T)$ over L . It is easy to check that E is then a splitting field of $f(T)$ over F .

So it suffices to take $f(T)$ itself to be irreducible over F . Apply Proposition 13.2.1 to obtain an extension $L = F[u]$ and a factorization $f(T) = (T - u)q(T)$ over L . Once more, by induction, there is a splitting field E of $q(T)$ over L , which is a splitting field of $f(T)$ over F . ■

The companion uniqueness theorem for splitting fields is next; we leave the proof to the reader. It should also be clear that, if E is a splitting field of $f(T)$ over F , then the members of $G = \text{Gal}(E/F)$ act transitively on the roots of $f(T)$.

Proposition 13.2.5. *Suppose that $f : F_1 \rightarrow F_2$ is an isomorphism of fields. Let $p_1(T)$ be a polynomial over F_1 , and $p_2(T)$ be the polynomial over F_2 obtained by applying f to all the coefficients of $p_1(T)$. Suppose that E_i is a splitting field of $p_i(T)$ over F_i , with $i = 1, 2$. Then there is an isomorphism $g : E_1 \rightarrow E_2$ extending f .*

We aim, in this section, for a characterization of finite Galois extensions, in terms of splitting fields of polynomials. There is a subtlety, for prime characteristics, which should be pointed out as a separate issue. First, the formal notion of “derivative”.

Definition 13.2.6. If $f(T) = a_n T^n + a_{n-1} T^{n-1} + \cdots + a_1 T + a_0$ is a polynomial over the field F , then the *derivative* is

$$f'(T) = n a_n T^{n-1} + (n-1) a_{n-1} T^{n-2} + \cdots + a_1.$$

It is easy to verify the usual rules for derivatives:

- (i) $(f + g)' = f' + g'$;
- (ii) $(fg)' = f'g + fg'$;
- (iii) $(cf)' = cf'$, for each $c \in F$.

We want to use the derivative to detect repetition of roots of a polynomial. Here's the first step.

Lemma 13.2.7. *Suppose that F is a field, $f(T) \in F[T]$, and $a \in F$. Then $(T - a)^2$ divides $f(T)$ if and only if $T - a$ divides both f and its derivative.*

Proof. If $f(T) = (T - a)^2 g(T)$, over F , then

$$f'(T) = 2(T - a)g(T) + (T - a)^2 g'(T),$$

whence $T - a$ divides $f'(T)$. Conversely, suppose that $f(T) = (T - a)h(T)$ and $f'(T) = (T - a)k(T)$. Note that $f'(T) = h'(T) + (T - a)h'(T)$, from which we may conclude that $T - a$ divides $h(T)$. ■

Lemma 13.2.7 has the following immediate consequence.

Lemma 13.2.8. *Suppose that $p(T)$ is an irreducible polynomial over the field F . Then the following are equivalent.*

- (a) *In every splitting field of $p(T)$ over F , $p(T)$ factors into distinct linear factors.*

(b) In some splitting field of $p(T)$ over F , $p(T)$ factors into distinct linear factors.

(c) $p'(T) \neq 0$.

Proof. As (b) is a special case of (a), we proceed to show that (b) implies (c). If $p'(T)$ is identically zero, then in particular (over the splitting field E guaranteed by (b)) $p(T)$ and $p'(T)$ share a root. Lemma 13.2.7 then tells us that $p(T)$ has a repeated root in E .

Next, assume (c). If, in some splitting field L , $p(T)$ has a repeated root a , then $T - a$ is a common divisor of $p(T)$ and $p'(T)$, which has smaller degree than p . Now, since $p(T)$ is irreducible we have (for suitable polynomials $r(T)$ and $s(T)$ over F) that $r(T)p(T) + s(T)p'(T) = 1$. Substituting a in this equation produces a contradiction. Thus, (c) implies (a). ■

Lemma 13.2.8 produces the following corollary.

Corollary 13.2.9. *Suppose that $p(T)$ is irreducible over the field F . Then $p(T)$ has repeated roots in some splitting field if and only if the characteristic of F is prime (say p), and $p(T) = g(T^p)$, for a suitable polynomial $g(T) \in F[T]$.*

Proof. Exercise. ■

We have come to the notion of separability; the contrasting feature, pure inseparability, will come up in §14.3.

Definition 13.2.10. Suppose that $p(T)$ is an irreducible polynomial over the field F . If $p(T)$ factors into distinct linear factors in some splitting field over F (and hence in all of them) we say that $p(T)$ is *separable*. An element u in a field extension of F , which is algebraic over F , is *separable* if its minimum polynomial over F is separable. An algebraic extension is called *separable* over F if every element is separable over F .

We should emphasize that – in view of Corollary 13.2.9 – over characteristic zero, separability is automatic.

Here is the main theorem of this section. It implies that if E is a Galois extension of F , then E is Galois over every intermediate subfield of E , and therefore, that every intermediate subfield is closed under the priming Galois correspondence.

Theorem 13.2.11. *Suppose that E is a finite field extension of the field F . Then the following are equivalent.*

- (a) E is Galois over F .
- (b) E is separable over F , and a splitting field (of some polynomial) over F .
- (c) E is the splitting field of some polynomial over F , of which all the irreducible factors are separable.

Proof. (a) implies (b): Suppose that $u \in E$, and that $p(T)$ is its minimum polynomial over F . We prove that $p(T)$ is separable. To that end, let $u = u_1, u_2, \dots, u_k$ be the distinct images of u under the elements of $G = \text{Gal}(E/F)$. Note that $k \leq \deg(p)$. Let $q(T) = (T - u_1)(T - u_2) \cdots (T - u_k)$; it is easy to see that each $g \in G$ fixes every coefficient of $q(T)$, which says that $q(T) \in F[T]$. This must mean that $p(T)$ divides $q(T)$, whence $p(T) = q(T)$.

Now let v_1, v_2, \dots, v_r be a basis of E over F . Let $p_i(T)$ be the minimum polynomial of v_i over F , and $f(T) = p_1(T) \cdots p_r(T)$. $f(T)$ is a separable polynomial, and it should be clear that $E = F[v_1, v_2, \dots, v_r]$ is a splitting field of $f(T)$ over F .

As it is clear that (b) implies (c), we turn to the remaining implication: that (a) is a consequence of (c).

Suppose that E is a splitting field of $f(T)$ over F , so that each irreducible factor of $f(T)$ is separable. Let $G = \text{Gal}(E/F)$. To prove that E is Galois over F it suffices to show that the order of G is $[E : F]$. If $f(T)$ factors completely over F , then $F = E$, and there is nothing to prove. Now let $q(T)$ be an irreducible factor of $f(T)$ of degree $r > 1$. Let u be a root of $q(T)$, and set $L = F[u]$, with $H = L'$. As in the proof of Lemma 13.1.9, $[G : H]$ is the number of distinct images of u under the automorphisms of G . However, each of the distinct roots of $q(T)$ is such an image; this is seen by applying Propositions 13.2.2 and 13.2.5. Thus, $[G : H] = r = [L : F]$. Since $H = \text{Gal}(E/L)$, we have that $|H| = [E : L]$ (by induction) and so $|G| = r|H| = [L : F][E : L] = [E : F]$. This proves that E is Galois over F . ■

In the following exercises the reader is cautioned that no assumption is made about the extensions being Galois. The first is a characterization of splitting fields that does not name a specific polynomial.

Exercise 13.2.12. Suppose that E is a finite extension of the field F . Then E is a splitting field over F (of some polynomial) if and only if each irreducible polynomial over F , which has a root in E , factors completely over E .

(Hint: For the necessity, suppose the contrary, namely that $g(T)$ is an irreducible polynomial over F , having a root $u \in E$, but also an irreducible factor $h(T)$ over E , of degree greater than one. Let v be a root of $h(T)$ in $E[v]$, which exists on account of Proposition 13.2.1. There is an isomorphism of $F[u]$ onto $F[v]$ (why?), which in turn can be extended to an isomorphism of E onto $E[v]$ (why?). This is nonsense, by comparing dimensions.)

Exercise 13.2.13. Suppose that E is a splitting field over F (of a suitable polynomial), and that L is an intermediate field. Prove that L is stable if and only if L is a splitting field over F . If so, then $\text{Gal}(L/F) = \text{Gal}(E/F)/L'$.

(Hint: Restrict members of $\text{Gal}(E/F)$ to L , if L is stable; use the preceding exercise to show that the restriction map is surjective.)

Exercise 13.2.14. Suppose that L is a finite extension of F . Then there is a field extension E of L , which is a splitting field over F , and such that no proper subfield of E , containing L is a splitting field over F .

(Hint: Pick a basis $\{v_1, \dots, v_r\}$ of L over F . Let $p_i(T)$ be the minimum polynomial of v_i over F , and $f(T)$ be the product of the $p_i(T)$. Let E be the splitting field of $f(T)$ over L .)

The preceding exercise motivates the following concept.

Definition 13.2.15. If L is a finite extension of the field F , and E is the least splitting field over F containing L , then E is called the *split closure* of L over F . This closure is unique, owing to Proposition 13.2.5.

Next, a huge anticlimax, that the separable elements of a splitting field form a subfield.

Proposition 13.2.16. Suppose that K is a splitting field over the field F . The subset E of all separable elements over F is a subfield of K .

Proof. Suppose that a and b in K are both separable over F . Let L be the splitting field of the minimum polynomial (over F) of a . By Exercise 13.2.12, $L \leq K$, and L is Galois over F , according to Theorem 13.2.11. Of course, b is separable over L , and K is still a splitting field over L . Let M

be the splitting field of the minimum polynomial of b over L . Then (for the same reasons) $M \leq K$, and M is Galois over L . Observing that every automorphism of L over F can be extended to one of M over F (since M is a splitting field, and owing to Proposition 13.2.5), we apply Exercise 13.1.17, and conclude that M is Galois over F , and therefore that M is separable over F . Since M contains $a \pm b$, ab and a^{-1} (if $a \neq 0$), this proves the proposition. ■

To conclude this section, an exercise which gives a criterion for when the Galois group of a splitting field consists of even permutations. We remind the reader that if a splitting field E of the polynomial $p(T)$ over F is Galois, then $\text{Gal}(E/F)$ can be viewed as a permutation group of the roots of $p(T)$.

Exercise 13.2.17. Suppose that F is a field of characteristic $\neq 2$. Suppose $p(T)$ is a polynomial over F , with distinct roots in its splitting field E over F . Let $G = \text{Gal}(E/F)$. Note that, by Theorem 13.2.11, E is Galois over F .

Now define

$$d = \prod_{i < j} (u_i - u_j),$$

where the u_i are the roots of $p(T)$. Let $\alpha = d^2$.

Prove that $\alpha \in F$, and that $d \in F$ if and only if G consists of even permutations. (d is called the *discriminant* of $p(T)$.)

13.3 Solutions by Radicals. In this section we solve the classical problem of when a polynomial equation (over \mathbb{Q}) can be solved by radicals; that is to say, by rational arithmetical operations plus the taking of radicals. It is the problem which gave rise to Galois theory. To begin, let us formally define the notion of a “radical extension”.

Definition 13.3.1. The field extension E of F is called a *radical extension* of F , if E is algebraic over F , and it is of the form $E = F[u_1, u_2, \dots, u_m]$, where (for each $i = 1, 2, \dots, m$) some power of u_i lies in $F[u_1, u_2, \dots, u_{i-1}]$. (In the case $i = 1$, we are saying that some power of u_1 lies in F .)

Note that, by factoring the exponents involved in the above, and introducing more elements, one can in fact assume that the powers of the u_i are prime powers. Whenever convenient we shall assume this has been done.

We say that a polynomial $f(T)$ over F is *solvable by radicals* if its splitting field is a radical extension.

We shall reach our goal, of proving that the Galois group of a radical extension is solvable, via a series of lemmas.

Lemma 13.3.2. Suppose that E is an algebraic extension of F , and $E = E_1 * E_2 * \dots * E_k$, where each E_i is a radical extension of F . Then E is a radical extension of F .

Proof. Exercise. ■

Lemma 13.3.3. If L is a radical extension of F , and E is the split closure of L over F , then E too is a radical extension.

Proof. $E = L_1 * \dots * L_k$, where each $L_i \cong L$, over F . (We leave the justification of this as an exercise.) Now apply Lemma 13.3.2. ■

The next lemma identifies how the abelian quotients of the Galois group of a radical extension will arise.

Lemma 13.3.4. *Suppose that p is a prime number, and that E is a splitting field of $T^p - 1$ over F . Then $\text{Gal}(E/F)$ is abelian.*

Proof. If the characteristic is p , then $T^p - 1 = (T - 1)^p$, and so $E = F$, and the conclusion is trivially true. If the characteristic is not p then $T^p - 1$ has distinct roots in E (by Lemma 13.2.8). Let y be a root of $T^p - 1$, different from 1. Then (in $F^\#$) y has order p , and $\{1, y, \dots, y^{p-1}\}$ are the distinct roots of $T^p - 1$. This means that $E = F[y]$.

Since each automorphism of E over F is determined by its action on y , and, for each $\alpha \in \text{Gal}(E/F)$, $\alpha(y) = y^i$, for some $i = 1, 2, \dots, p - 1$, it should be clear that $\text{Gal}(E/F)$ is abelian. ■

The final piece of the puzzle, in advance of the main theorem of this section is this:

Lemma 13.3.5. *Suppose that F is a field containing all the n -th roots of unity. Suppose that $a \in F$, and E is a splitting field of $T^n - a$ over F . Then $\text{Gal}(E/F)$ is abelian.*

Proof. Suppose that $u \in E$ is a root of $T^n - a$. In general, the roots of $T^n - a$ are of the form eu , where e is an n -th root of unity. Thus, $E = F[u]$, and so the members of $\text{Gal}(E/F)$ are completely determined by their action on u . So if $g, h \in \text{Gal}(E/F)$, $g(u) = eu$, $h(u) = e'u$, where e and e' are n -th roots of unity. Note that

$$g(h(u)) = g(e'u) = e'eu = h(g(u)),$$

which is enough to prove that $gh = hg$. Thus, $\text{Gal}(E/F)$ is abelian. ■

Remark 13.3.6. In fact, under the situations of the preceding two lemmas, the Galois groups are cyclic. We leave the verification of this to the reader.

Before stating the main theorem, the reader is advised to review the basic properties of solvable groups. Chiefly, recall that any subgroup and any homomorphic image of a solvable group is solvable. Also, if G is a group, having a normal subgroup N , so that both N and G/N are solvable, then G itself is solvable.

Now, here is the main theorem:

Theorem 13.3.7. *Suppose that E is a radical extension of F . For any intermediate subfield L of E , $\text{Gal}(L/F)$ is solvable.*

Proof. By passing to the fixed field F_o of $\text{Gal}(L/F)$, the Galois group is unchanged, and E remains a radical extension of F_o . So without loss of generality we may assume that L is Galois over F .

Next, Lemma 13.3.3 insures that the split closure of E over F is also a radical extension, so we may assume that E is a splitting field over F . By Exercise 13.2.13 and Theorem 13.2.11, it follows that L is a stable subfield. Thus, $\text{Gal}(L/F)$ is a homomorphic image of $\text{Gal}(E/F)$, and if we can show that the latter is solvable, we are done.

We therefore prove: if E is a splitting field over F and a radical extension of F , then $\text{Gal}(E/F)$ is solvable.

Suppose that $E = F[u_1, u_2, \dots, u_k]$ is the “presentation” which witnesses the fact that E is a radical extension of F . We will induct on k . As mentioned in the definition of “radical extension”, we may suppose that $u_1^p \in F$, for a suitable prime number p . Let \hat{E} be the splitting field of $T^p - 1$ over E . For any root v of this polynomial ($\neq 1$), we have that $\hat{E} = E[v]$. Let $\hat{F} = F[v]$.

Consider now the situation inside \hat{E} ; it is a splitting field over F , and since E too is a splitting field, E is stable under $\text{Gal}(\hat{E}/F)$, by Exercise 13.2.13. So if $\text{Gal}(\hat{E}/F)$ were shown to be solvable, then $\text{Gal}(E/F)$, being a homomorphic image of it, would also be solvable.

Now \hat{F} too is stable (thanks to Exercise 13.2.13, once more), and $\text{Gal}(\hat{F}/F)$ is abelian, according to Lemma 13.3.4. If we can show that $\text{Gal}(\hat{E}/\hat{F})$ is solvable, then so is $\text{Gal}(\hat{E}/F)$, as $\text{Gal}(\hat{E}/\hat{F}) = (\hat{F})'$ (under the Galois connection) is a normal subgroup of it, with $\text{Gal}(\hat{F}/F)$ as quotient.

The proof therefore reduces to showing that $\text{Gal}(\hat{E}/\hat{F})$ is solvable. Now, $\hat{E} = \hat{F}[u_1, u_2, \dots, u_k]$, and since \hat{F} contains all the p -th roots of unity $\hat{F}[u_1]$ is a splitting field over F , with abelian Galois group, by Lemma 13.3.5. Thus, $\hat{F}[u_1]$ is stable. Furthermore, $\text{Gal}(\hat{E}/\hat{F})$ has a normal subgroup isomorphic to $\text{Gal}(\hat{E}/\hat{F}[u_1])$, which is solvable by induction, and a quotient group, $\text{Gal}(\hat{F}[u_1]/F)$ which is abelian. We finally are able to conclude that $\text{Gal}(\hat{E}/\hat{F})$ is solvable, and as explained in the preceding paragraphs, this shows that $\text{Gal}(E/F)$ is solvable. ■

The following “application” of Theorem 13.3.7 permits us to produce reasonable splitting extensions which are not radical extensions.

Proposition 13.3.8. *Let p be a prime number, and suppose that $p(T)$ is an irreducible polynomial of degree p over \mathbb{Q} . Assume that $p(T)$ has exactly two nonreal roots. Then the Galois group of the splitting field of $p(T)$ is S_p .*

Proof. In building the splitting field E of $p(T)$, we adjoin a root of $p(T)$, which means that E contains a subfield of dimension p over \mathbb{Q} . Thus, p divides the order of $\text{Gal}(E/\mathbb{Q})$. Viewing this Galois group as a permutation group on the p distinct roots, then there is an element in it of order p , which is necessarily a p -cycle. As there are only two nonreal roots, complex conjugation induces a transposition.

We now leave it to the reader to check that a subgroup of S_p which contains a p -cycle and a transposition is all of S_p . ■

It is reasonably straightforward to produce a quintic polynomial, over \mathbb{Q} , which satisfies the hypotheses of Proposition 13.3.8, and therefore has a splitting field E for which $\text{Gal}(E/\mathbb{Q}) = S_5$, which is not solvable. Such a quintic is not solvable by radicals.

Remark 13.3.9. Theorem 13.3.7 has a converse. One has to be careful with characteristics, although in characteristic zero, there are no problems. Developing this converse takes some effort, which we prefer to bestow on different matters. The reader is referred to [Ka69], §7.

We now turn to the theory of finite fields.

13.4 Finite Fields. Although only tangentially connected to the subject of finite fields, the following result is overdue.

Proposition 13.4.1. *If G is any finite multiplicative subgroup of (the nonzero elements of) a field F , then G is cyclic.*

Proof. If G is finite, of order r , then $x^r = 1$, for each $x \in G$. On the other hand, as G is an abelian group, by the Fundamental Theorem on finitely generated abelian groups, G can be expressed as a direct sum of cyclic subgroups C_1, C_2, \dots, C_k , so that the order of C_i divides that of C_{i+1} . If $k \geq 2$, then there is a positive integer $m < r$, so that $x^m = 1$.

Since the polynomial $T^m - 1$ has no more than m roots, this amounts to a contradiction, unless in the above decomposition $k = 1$, and G is cyclic. ■

Corollary 13.4.2. *The group of nonzero elements of a finite field is cyclic.*

Hot Air 13.4.3. *Last gasps.* Suppose that F is a finite field. Then it has prime characteristic, say p , and contains a copy of \mathbb{F}_p , the field of integers modulo p . If $n = [F : \mathbb{F}_p]$, then it is easily seen that $|F| = p^n$. Moreover, each nonzero element x of F satisfies the polynomial $T^{p^n} - 1$.

Conversely, consider the polynomial $T^{p^n} - T$ over \mathbb{F}_p , and let E be the splitting field of $T^{p^n} - T$ over \mathbb{F}_p . Verify that the subset F of E , consisting of the roots of this polynomial, is a subfield of E , which means that $F = E$. (The reason for this is the so-called “freshman’s dream”, over characteristic p ; namely, that $(x + y)^{p^n} = x^{p^n} + y^{p^n}$, which implies that the map $x \mapsto x^{p^n}$ is a field isomorphism. Thus, $|E| = p^n$, and it follows that $[E : \mathbb{F}_p] = n$. All of this proves the following theorem.

Theorem 13.4.4. *Each finite field F , of characteristic p , has order p^n , for suitable n . Moreover, for each positive integer n , there is (up to isomorphism) exactly one field of order p^n , namely the splitting field of $T^{p^n} - T$ over \mathbb{F}_p .*

(Note: the unique field of order p^n is denoted by \mathbb{F}_{p^n} , but also, frequently, by $\text{GF}(p^n)$.)

Theorem 13.4.4 has the following consequence.

Proposition 13.4.5. *Suppose that E is a finite field (of order p^n , for a suitable prime number p). Then $\text{Gal}(E/\mathbb{F}_p)$ is cyclic of order n . In particular, if m divides n , then $\text{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_{p^m})$ is cyclic of order n/m .*

Proof. We prove the first claim, and leave the second to the reader.

The map $\alpha(x) = x^p$ is an automorphism of E , because E is finite. Clearly $\alpha^n = 1$. If $\alpha^k = 1$, for a smaller exponent k , then it is easy to see that $x^{p^k} - x = 0$, for each $x \in E$. This contradicts the fact that E has order p^n . Since E is the splitting field of $T^{p^n} - T$, which is separable, we have $\text{Gal}(E/\mathbb{F}_p) = [E : \mathbb{F}_p] = n$. Now α is an element of order n in a group of order n , whence $\text{Gal}(E/\mathbb{F}_p)$ is cyclic of order n . ■

We close with a few exercises. The first is the so-called Theorem of the Primitive Element. We say that E is a *simple extension* of F , if, for some $u \in E$, E is the smallest subfield containing F and u . (If E is algebraic over F , this is equivalent to saying that $E = F[u]$.)

Exercise 13.4.6. Suppose that E is any finite extension of the field F , (which is not necessarily finite). Then E is a simple extension of F if and only if E has only a finite number of intermediate subfields. Conclude from this that if L is any finite, separable extension of F , then L is a simple extension.

(Hint: Take the case where F is finite first; then E is also finite, and so Proposition 13.4.1 can be applied. The number of intermediate subfields is finite, vacuously.)

Now suppose that F is infinite. For the sufficiency, pick u in E , so that $[F[u] : F]$ is as large as possible. Suppose that $F[u]$ is not E , $v \in E \setminus F[u]$, and consider the list of intermediate subfields $\{F[u + av] : a \in F\}$. This must be a finite list. As to the necessity, suppose that $E = F[u]$. Prove that if L is any intermediate subfield of E , then $L = F[a_1, \dots, a_r]$, where the a_i are the coefficients of the minimum polynomial of u over L . On the other hand, every such polynomial must be a factor of the minimum polynomial of u over F .

The final statement in the exercise is a consequence of the first and the Fundamental Theorem of Galois Theory.)

By contrast, observe the following.

Exercise 13.4.7. Let F be an infinite field of prime characteristic p , and $E = F[u, v]$, such that u^p and v^p lie in F , with $[E : F] = p^2$. Show that E is not a simple extension of F , by exhibiting an infinite number of intermediate subfields.

Exercise 13.4.8. Suppose that F is any finite field, and n is any positive integer. Show that there is an irreducible polynomial over F , of degree n .

Exercise 13.4.9. Let $n \geq 3$. Prove that $T^{2^n} + T + 1$ is reducible over \mathbb{F}_2 .
 (Hint: if u is a root of this polynomial, raise the equation $u^{2^n} = u + 1$ to the 2^n -th power.)

13.5 Integral Closures in Algebraic Number Fields This section addresses the matter which was left unresolved during the discussion on Dedekind domains, in the preceding chapter. We prove that the ring of integers of a finite, separable extension of \mathbb{Q} is a Dedekind domain.

The key to the result is the independence of automorphisms.

Definition 13.5.1. Let A be any commutative ring with identity, and S any set of automorphisms of A . We say that the members of S are *linearly independent* over A if, for any $\alpha_1, \alpha_2, \dots, \alpha_n \in S$ and $r_1, r_2, \dots, r_n \in A$,

$$r_1\alpha_1(x) + r_2\alpha_2(x) + \dots + r_n\alpha_n(x) = 0,$$

for all $x \in A$, implies that each $r_i = 0$.

Over fields we have the following independence result.

Lemma 13.5.2. Any distinct automorphisms of a field F are linearly independent over F .

Proof. By way of contradiction, suppose that S is a set of distinct automorphisms of F are linearly dependent. Among the relations of dependence, pick one which is minimal, in the sense of involving the fewest number of elements of S . We write down such a dependence:

$$u_1\theta_1(x) + u_2\theta_2(x) + \dots + u_k\theta_k(x) = 0, \quad \forall x \in F.$$

Of course, $k > 1$, and there exists a $y \in F$ such that $\theta_1(y) \neq \theta_2(y)$. Now, replace x in the equation above by yx :

$$\begin{aligned} 0 &= u_1\theta_1(yx) + u_2\theta_2(yx) + \dots + u_k\theta_k(yx) \\ &= u_1\theta_1(y)\theta_1(x) + u_2\theta_2(y)\theta_2(x) + \dots + u_k\theta_k(y)\theta_k(x). \end{aligned}$$

Multiply the original equation by $\theta_1(y)$ and subtract:

$$u_2(\theta_2(y) - \theta_1(y))\theta_2(x) + \dots + u_k(\theta_k(y) - \theta_1(y))\theta_k(x) = 0,$$

which is a shorter equation of dependence. This is a contradiction. ■

The concept of “trace”, along with the companion “norm”, are classical tools for the study of field extensions, the so-called Kummer Theory. We shall not explore any of that here. But the notion of trace is handy in establishing the result we want for algebraic number fields.

Definition 13.5.3. Suppose that E is a finite field extension of the field F . Suppose that L is a finite Galois extension of F , containing E . Let $G = \text{Gal}(L/F)$ and $H = E'$. Let $\tau_1, \tau_2, \dots, \tau_k$ be a complete set of (right) coset representatives for H , and define the *trace* of $x \in E$ to be

$$T_E(x) = \tau_1(x) + \tau_2(x) + \cdots + \tau_k(x).$$

Observe that, since the quotient of two members of the same coset, modulo H , leave each element of E fixed, the definition of $T_E(x)$ is independent of the set of representatives used.

The following is easy to prove; it is left to the reader.

Proposition 13.5.4. *Suppose that L is a finite Galois extension of F and that E is an intermediate field. The trace map T_E satisfies*

- (a) $T_E(x) \in F$, for each $x \in E$;
- (b) $T_E(x + y) = T_E(x) + T_E(y)$, for all $x, y \in E$;
- (c) $T_E(ax) = aT_E(x)$, for all $a \in F$, $x \in E$;
- (d) $T_E(x) = kx$, for each $x \in F$, where $k = [G : H]$, and $G = \text{Gal}(L/F)$ and $H = E'$.

Thus, T_E is an F -linear functional defined on E .

We are now ready for the theorem on algebraic number fields.

Theorem 13.5.5. *Suppose that E is a finite extension of \mathbb{Q} , and that A is the integral closure of \mathbb{Z} in E . Then A is a Dedekind domain.*

Proof. According to the material presented in §12.4, it must be shown that A is a Noetherian ring of Krull dimension 1, since it is integrally closed, by definition. That A has Krull dimension 1 is a consequence of the Going-up Theorem. It remains to show that A is Noetherian. We prove, in fact, that A is a finitely generated abelian group.

If $x \in E$, consider the minimum polynomial of x over \mathbb{Q} , $a_n T^n + \cdots + a_1 T + a_0$. We may assume that the $a_i \in \mathbb{Z}$. Multiplying by a_n^{n-1} , we get the equation

$$u^n + a_{n-1}u^{n-1} + \cdots + a_1 a_n^{n-2}u + a_0 a_n^{n-1} = 0,$$

where $u = a_n x$, signifying that u is integral over A . Hence $u \in A$. The point of this observation is that if x_1, x_2, \dots, x_k is any basis of E over \mathbb{Q} , we may multiply each x_i by a suitable element of A and get a new basis, consisting of elements of A . Suppose that this has been done, and that x_1, x_2, \dots, x_k is such a basis.

Moreover, a basis for E over \mathbb{Q} may be chosen, so that the trace of each element is nonzero; (think about this; given any linear functional f on a finite dimensional vector space, one can pick a basis such that $f(v) = 1$, for each member v of the basis.) The procedure in the preceding paragraph, which transfers the basis to A preserves the condition that the trace of each basis member is nonzero. We assume that this too has been done.

Let L be the split closure of E over \mathbb{Q} ; it is a Galois extension of \mathbb{Q} , as we are in characteristic zero. Let $T = T_E$. For each $i = 1, 2, \dots, k$, let $T_i(u) = T(x_i u) = T(x_i)T(u)$; this is a linear functional, for each i . If $r_1 T_1 + \cdots + r_k T_k = 0$, with rational numbers r_i ($1 \leq i \leq k$), then for each $u \in E$ we have,

$$r_1 T(x_1)T(u) + \cdots + r_k T(x_k)T(u) = 0,$$

which, by applying the definition of the trace map, can be converted into an equation of linear dependence of automorphisms over \mathbb{Q} . This yields, for each $i = 1, 2, \dots, k$, that $r_i T(x_i) = 0$. As none of the $T(x_i)$ is zero, we conclude that each $r_i = 0$.

We've shown that the set $\{T_1, T_2, \dots, T_k\}$ of linear functionals is linearly independent, and, therefore, a basis. From linear algebra, there is a dual basis for E over \mathbb{Q} ; that is, a basis y_1, y_2, \dots, y_k over \mathbb{Q} so that

$$T(x_i y_j) = T_i(y_j) = \delta_{ij}.$$

Now, suppose that $a \in A$, and $a = s_1 y_1 + s_2 y_2 + \dots + s_k y_k$, where each $s_i \in \mathbb{Q}$. Observe that $T(ax_i) = s_i$, by Proposition 13.5.4. On the other hand, each $ax_i \in A$, thus integral over \mathbb{Z} ; for each $\tau \in \text{Gal}(L/\mathbb{Q})$, $\tau(ax_i)$ is also integral over \mathbb{Z} , whence it follows that each $T(ax_i)$ is integral over \mathbb{Z} ; that is to say, $T(ax_i) \in A$. But since \mathbb{Z} is integrally closed in \mathbb{Q} , we conclude that each $s_i \in \mathbb{Z}$.

Conclusion: $A \leq \mathbb{Z}y_1 + \dots + \mathbb{Z}y_k$. This says that, as an abelian group, A is a subgroup of a finitely generated free abelian group. Invoking Theorem 6.3.7 (and its proof), A is also a finitely generated free abelian group. In particular, A is a Noetherian module over itself, and, thus, a Noetherian ring, and the proof of the theorem is, at last, complete. ■

The preceding proof generalizes, almost word for word, to yield the following:

Theorem 13.5.6. *Suppose that A is a Dedekind domain, and $K = qA$. Suppose that L is a finite separable extension of K , and that B is the integral closure of A in L . then B is a Dedekind domain.*

We close with the following curious exercise on trace maps.

Exercise 13.5.7. Suppose that L is a finite, Galois field extension of the field F , and E is an intermediate extension. For each $u \in E$, let $p(T) = T^n + a_{n-1}T^{n-1} + \dots + a_1T + a_0$ be its minimum polynomial over F . Then $T_E(u) = -ma_{n-1}$, where $m = [E : F]/n$.

(Hint: a_{n-1} is the negative of the sum of the roots of $p(T)$. Now, in the definition of $T_E(u)$, count the number of times each root is repeated.)

14. Field Theory: Infinite Field Extensions

This chapter is dedicated to extensions of fields, in general. In the first section the existence and uniqueness of the algebraic closure is shown. We then survey various topics, including transcendence bases, Lüroth's Theorem and pure inseparability, closing with an account of the theory of real closed fields, due to Artin and Schreier.

14.1 Algebraic Closure.

Definition 14.1.1. For the record, a field F is *algebraically closed* if every nonconstant polynomial over F factors completely over F . An *algebraic closure* of F is an algebraic extension of F which is also algebraically closed.

Since “algebraic over algebraic is algebraic”, an algebraic closure of F is precisely an algebraic extension which is not properly contained in any algebraic extension of F .

First, we establish existence.

Theorem 14.1.2. *Every field F possesses an algebraic closure.*

Proof. To begin, observe the following. If E is any algebraic extension of the field F , then every element of E is the root of some polynomial with coefficients in F . Since each polynomial involves only a finite number of elements of F as coefficients, and the number of roots of any polynomial (in any extension) cannot exceed the degree of that polynomial, a simple cardinality argument shows that if E is algebraic, then $|F| = |E|$, if F is infinite, and $|E|$ is countable if F is finite.

The importance of this is that every algebraic extension of F has an underlying set whose cardinality is bounded by $|F|^\omega$. This will permit the use of Zorn's Lemma. To put it precisely, let \tilde{Q} be a set, the cardinality of which exceeds that of the field F . Define a set \mathcal{S} as follows: $L \in \mathcal{S}$ if it is a subset of \tilde{Q} , which admits operations making it a field, so that F is a subfield of L , and L is an algebraic extension of F . Order \mathcal{S} by declaring $L_1 \leq L_2$ if L_1 is a subfield of L_2 . It is easy to see that \mathcal{S} fulfills the conditions one needs to apply Zorn's Lemma.

Now, let E be a maximal member of \mathcal{S} . If L is an algebraic extension of E (but not necessarily a subset of \tilde{Q}), and $u \in L \setminus E$, then consider $E[u] > E$. Let $p(T)$ be the minimum polynomial of u over E . Pick an arbitrary element v in $\tilde{Q} \setminus E$ (which we can do, since the cardinality of \tilde{Q} is larger than that of E), and via any bijection α between $E[u]$ and a subset \hat{E} of \tilde{Q} , which is the identity on E and sends u to v , define the field operations on \hat{E} , so as to make \hat{E} an algebraic field extension of E , and hence of F , violating the maximality of E .

Thus, E is algebraically closed. ■

Next, we generalize the notion of a splitting field.

Definition 14.1.3. Suppose that E is a field extension of F . If there is a family $\{p_i(T) : i \in I\}$ of polynomials over F , which factor completely over E , and such that E is generated by F and the roots of all the $p_i(T)$, then we say that E is a *splitting field* over F (of the polynomials in question). We observe that an algebraic closure of F is a splitting field over F ; namely, of all the nonconstant polynomials over F .

We also leave to the reader the following generalization of Exercise 13.2.12.

Exercise 14.1.4. Suppose that E is an algebraic extension of F . Then E is a splitting field over F if and only if each irreducible polynomial over F , having a root in E , factors completely over E .

To establish the uniqueness of algebraic closures, we have to, once again, do a set-theoretic song and dance.

Theorem 14.1.5. *Suppose that F_1 and F_2 are fields, and that $f : F_1 \rightarrow F_2$ is an isomorphism. Let $\{p_{i,1}(T) : i \in I\}$ be a family of polynomials over F_1 , and let $\{p_{i,2}(T) : i \in I\}$ be the corresponding family of polynomials over F_2 (via f). Suppose that E_1 (resp. E_2) is a splitting field of the $(p_{i,1}(T))_{i \in I}$ (resp. $(p_{i,2}(T))_{i \in I}$) over F_1 (resp. F_2). Then there is an isomorphism $g : E_1 \rightarrow E_2$ extending f .*

Proof. We consider \mathcal{E} , the family of all triples (L_1, L_2, h) , such that $h : L_1 \rightarrow L_2$ is an isomorphism which extends f , and L_i is an intermediate subfield of E_i ($i = 1, 2$). Order \mathcal{E} as follows: $(L_1, L_2, h) \leq (M_1, M_2, h')$ if $L_i \leq M_i$ ($i = 1, 2$), and h' extends h .

Now, it is straightforward that Zorn's Lemma can be applied to \mathcal{E} . We leave that to the reader, as well as the verification that, if (N_1, N_2, k) is a maximal triple in \mathcal{E} , then $N_i = E_i$ ($i = 1, 2$). ■

As a consequence of Theorem 14.1.5, we have the following feature of algebraically closed fields.

Proposition 14.1.6. *Suppose that E is an algebraic field extension of F . Let L be any algebraically closed field, containing F as a subfield. Then there is an isomorphism α of E onto an intermediate subfield of L , which fixes the elements of F .*

Proof. Embed E in a splitting field over F , say \hat{E} . (Take a basis S of E over F , and form the subfield of the algebraic closure of F generated by the roots of the minimum polynomial $p_s(T)$, over F , for each $s \in S$.) By Theorem 14.1.5, there is a subfield L_o of L , containing F , and an isomorphism of \hat{E} onto L_o , which extends the identity map on F . ■

We have two more generalizations of the finite case. The details are left to the reader. First, let's generalize Theorem 13.2.11.

Proposition 14.1.7. *Suppose that E is an algebraic extension of F . Then E is Galois over F if and only if it is separable and a splitting field over F .*

Then, as a consequence of Proposition 14.1.7, we have, generalizing a comment made following Theorem 13.2.11:

Proposition 14.1.8. *If E is a Galois algebraic extension of F , then it is Galois over any intermediate subfield, and hence every intermediate subfield of E is closed under the Galois correspondence.*

Definition & Remarks 14.1.9. *The Krull Topology.* To close these proceedings, we observe that if E is an algebraic Galois extension of F , then $G = \text{Gal}(E/F)$ can be topologized, so that the group operations become continuous, and G is compact and Hausdorff in the topology. Under this topology, called the *Krull topology* (after Wolfgang Krull, who first noted these facts), the closed subgroups are precisely the ones which are closed under the Galois correspondence.

None of this should be mysterious; we give a brief account of the Krull topology. For each positive integer n , and each pair of n -tuples (a_1, a_2, \dots, a_n) and (b_1, b_2, \dots, b_n) of elements of E , let

$$V(a_1, \dots, a_n; b_1, \dots, b_n) \equiv \{g \in G : g(a_i) = b_i, \forall i = 1, \dots, n\}.$$

It is not hard to verify that the family of all finite unions of sets of the form $V(a_1, \dots, a_n; b_1, \dots, b_n)$, constitute a base for the closed sets on a topology on G . This is the Krull topology.

Now, here are some basic features, which follow rapidly from the definition of the basic closed sets.

- (a) For each finite extension L of F , which is a subfield of E , L' (under the Galois correspondence) is (topologically) closed.

Proof. Let a_1, a_2, \dots, a_m be a basis for L over F , and then observe that

$$L' = V(a_1, \dots, a_m; a_1, \dots, a_m).$$

■

- (b) L' is closed in the Krull topology, for each intermediate subfield L of E .

Proof. L is a directed union of finite extensions of F , making L' an intersection of subgroups of the form M' , where M is a finite extension of F . Then apply (a). ■

- (c) The Krull topology is Hausdorff.

Proof. It suffices to show that any nontrivial automorphism can be separated from 1. But, if $g \in G$ is nontrivial, then $gu \neq u$, for some $u \in E$. Now, since E is algebraic over F , the set of images of u under G is finite. Let $\{u, gu = g_1u, \dots, g_ru\}$ be the distinct images of u under action by G . Then the open sets $W_1 = G \setminus V(u; u)$ and $W_2 = G \setminus V(u; g_1u) \cap \dots \cap G \setminus V(u; g_ru)$ are disjoint. Furthermore, $1 \in W_2$ and $g \in W_1$. ■

- (d) The Krull topology is compact. (The proof is left as an exercise.)

- (e) Composition of automorphisms in G is continuous in the Krull topology. (Exercise.)

- (f) Inversion in G is continuous.

Proof. Note that $g \in V(a_1, \dots, a_n; b_1, \dots, b_n)$ precisely when $g^{-1} \in V(b_1, \dots, b_n; a_1, \dots, a_n)$. ■

- (g) For each $X \subseteq G$, $(X)'$ (under the Galois correspondence) is the topological closure of X . In particular, any closed subgroup of G is of the form L' , for some intermediate field L . (Proof left as an exercise.)

- (h) If E is a finite Galois extension then the Krull topology is discrete.

Proof. E is a separable extension, so that, by Exercise 13.4.6, there is an element $u \in E$, so that $E = F[u]$. Let $u_1 = u, u_2, \dots, u_k$ be the distinct roots of the minimum polynomial of u over F . Since the action of an automorphism in G is completely determined by its action on u , it follows that $G \setminus V(u; u_2) \cap \dots \cap G \setminus V(u; u_k) = 1$. Since translation in G is a homeomorphism, by (e), this is enough to show that the Krull topology is discrete. ■

14.2 Transcendence Bases. The main objective of this section is Lüroth's Theorem on purely transcendental extensions of degree one.

Definition 14.2.1. Suppose that K is a field extension of the field F , and $a_1, a_2, \dots, a_m \in K$. These are said to be *algebraically independent over F* , if there is no nonzero polynomial $f(T_1, \dots, T_m) \in F[T_1, T_2, \dots, T_m]$ such that $f(a_1, \dots, a_m) = 0$. A single element of K is algebraically independent if and only if it is transcendental over F .

An arbitrary subset $S \subseteq K$ is *algebraically independent over F* if every finite subset of S is algebraically independent over F . A maximal algebraically independent subset of K , over F , is called a *transcendence basis of K over F* .

The following result gives the basic facts about transcendence bases. Existence is proved by a typical Zorn's Lemma argument. The fact that any two finite transcendence bases have the same number of elements is shown by a replacement argument, akin to the one used to show that any two bases of a finitely generated vector space have the same number of elements. Part (c) of Proposition 14.2.2 is a direct consequence of maximality.

Proposition 14.2.2. *Let K be a field extension of F .*

- (a) K has a transcendence basis over F .
- (b) If $\{a_1, a_2, \dots, a_k\}$ and $\{b_1, b_2, \dots, b_m\}$ are transcendence bases of K over F , then $k = m$.
- (c) If S is a transcendence basis of K over F , then K is algebraic over $F(S) \equiv q(F[S])$.

Definition & Remarks 14.2.3. We caution that, if S is a transcendence basis of K over F , then $F(S)$ need not equal K . For example, if T is a single indeterminate, then $\{T\}$ and $\{T^2\}$ are both transcendence bases of $F(T)$ over F , but $F(T^2)$ is a proper subfield of $F(T)$.

The *transcendence degree* of an extension is infinite, if there is no finite transcendence basis, and, otherwise, the cardinality n of any finite transcendence basis. If K has a transcendence basis S over F , such that $F(S) = K$, then we say that K is a *purely transcendental extension* of F .

Exercise 14.2.4. Show that in $\mathbb{Q}(T, \sqrt{T^3 - T})$ the only algebraic elements over \mathbb{Q} are the rationals themselves, but $\mathbb{Q}(T, \sqrt{T^3 - T})$ is not a purely transcendental extension of \mathbb{Q} .

Hot Air 14.2.5. *Is a subfield of a purely transcendental extension also purely transcendental?* The answer is no. And NO as well, to the following: if K is purely transcendental over F , and L is an intermediate subfield of K , of the same transcendence degree n , is L purely transcendental? There are examples with $n \geq 3$.

The answer is YES with $n = 2$, a theorem due to Castelnuovo. The goal of this section is to prove this for $n = 1$.

Theorem 14.2.6. (Lüroth's Theorem) *Suppose that $K = F(T)$, and L is a subfield of K , properly containing F . Then there is an $r \in K$, so that $L = F(r)$.*

A key ingredient in the proof of Lüroth's Theorem is the following result, which is of independent interest anyway.

Proposition 14.2.7. *Suppose that $K = F(T)$. Let $g = g(T)$ and $h = h(T)$ be relatively prime polynomials (in $F[T]$), at least one of which is not constant. Then*

$$[K : F(g/h)] = \max \{ \deg(g), \deg(h) \}.$$

Proof. Let $s = g/h$; consider the polynomial (in the variable X) $f(X) = g(X) - sh(X)$. First, one shows that $f(X)$ is irreducible over $F(s)$: first prove it is irreducible over $F[s]$, which is enough, by Gauss' Lemma. Then it is enough to show that the degree of $f(X)$ is $\max \{ \deg(g), \deg(h) \}$. Since T is a root of $f(X)$, the result then follows. ■

Here's a timely look at a classical group.

Example 14.2.8. *The projective linear group $\text{PGL}(2, F)$.* Let us now apply Proposition 14.2.7 to compute the Galois group of $F(T)$ over F . For each $s \in F(T) \setminus F$, the assignment $T \mapsto s$ extends to an isomorphism of $F(T)$ onto a subfield of $F(T)$. By Proposition 14.2.7, $F(s) = F(T)$ if and only if $s = (aT + b)/(cT + d)$, for suitable $a, b, c, d \in F$, with $ad - bc \neq 0$.

Each $T \mapsto (aT + b)/(cT + d)$, with $ad - bc \neq 0$, is called a *fractional linear transformation*. Moreover, the map $\Theta : \text{GL}(2, F) \rightarrow \text{Gal}(F(T)/F)$, defined by

$$\Theta \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \left\{ T \mapsto \frac{aT + b}{cT + d} \right\},$$

is a surjective homomorphism, with kernel $Z(\text{GL}(2, F))$, which consists of the scalar matrices. Thus,

$$\text{PGL}(2, F) \equiv \text{GL}(2, F)/Z(\text{GL}(2, F)) = \text{Gal}(F(T)/F).$$

Note, for example, in the case where F is a finite field, say of order q , that $\text{PGL}(2, F)$ is finite, of order $q(q-1)(q+1)$. Thus, F cannot be the fixed field of its Galois group, as $F(T)$ is an infinite extension of F .

Proof. of Lüroth's Theorem. Suppose that E is a subfield of $F(T)$ (T an indeterminate), properly containing F . Let $y \in E \setminus F$. As $F(T)$ is algebraic over $F(y)$, it is also algebraic over E . So suppose that $q(X)$ is the minimum polynomial of T over E ; say,

$$q(X) = c_0 + c_1X + \cdots + X^n.$$

Each $c_i = u_i(T)/v_i(T)$, for suitable polynomials in T . Multiplying by the correct polynomial in T gives

$$q(T, X) = d_n(T)X^n + \cdots + d_1(T)X + d_0(T),$$

a polynomial in $F[T, X]$, which is a primitive polynomial in X . Note too that $c_i = d_i(T)/d_n(T) \in E$, and not all the c_i are in F , because T is transcendental over F . This means that at least one of the c_i is of the form $u(T)/v(T)$, with u and v relatively prime, and such that $m = \max\{\deg(u), \deg(v)\} > 0$.

From Proposition 14.2.7, and its proof, one obtains that $u(X) - c_i v(X)$ is the minimum polynomial of T over $F(c_i)$, and that $[F(T) : F(c_i)] = m$. Also, $c_i \in E$ implies that $n \leq m$, proving that $m = n$ shows that $E = F(c_i)$, and proves Lüroth's Theorem.

Now T is a root of $u(X) - c_i v(X)$, which has coefficients in E . Thus, $u(X) - c_i v(X) = q(X)t(X)$, in $E[X]$. Multiplying by the correct denominator, a polynomial in T , one gets an expression:

$$(*) \quad w(T)[u(X)v(T) - u(T)v(X)] = q(T, X)t(T, X).$$

Recalling that $q(T, X)$ is primitive as an element of $F(T)[X]$, we are able to conclude that $w(T)$ divides $t(T, X)$, and then, dividing $w(T)$ out, the expression $(*)$ simplifies to

$$(**) \quad u(X)v(T) - u(T)v(X) = q(T, X)t'(T, X).$$

Observe now that the degree of the expression on the left (in $(**)$) is at most m . On the other hand, the degree in X of the right is at least n . Hence, $m = n$, unless the left side is identically zero. The latter would imply that $u(0)v(T) = u(T)v(0)$, and so that $u(T)/v(T) \in F$, contrary to our stipulation. ■

We conclude this brief excursion into the subject of transcendental extensions with the following exercise.

Exercise 14.2.9. Prove that a purely transcendental extension of a field is never algebraically closed.

14.3 Separable vs Inseparable. The subject under examination in this section is at the opposite extreme of separability. Over a field of prime characteristic, we look at the situation involving an irreducible polynomial with a single root in its splitting field.

Definition & Remarks 14.3.1. Suppose that E is a field extension of the field F ; $x \in E$ is *purely inseparable over F* if the minimum polynomial of x over F has a single root, namely x . According to Corollary 13.2.9, if this happens, then the characteristic of F is a prime p , and the minimum polynomial $g(T)$ of x is of the form $g(T) = h(T^p)$, for a suitable $h(T) \in F[T]$.

Let us analyze this definition more closely. Corollary 13.2.9 leads to the following lemma.

Lemma 14.3.2. Suppose that F is a field of prime characteristic p , and E is a field extension of F , with $x \in E \setminus F$ algebraic over F . Let $g(T)$ be the minimum polynomial of x over F . Then there is a separable polynomial $g_s(T) \in F[T]$, irreducible over F , so that $g(T) = g_s(T^{p^n})$, for some nonnegative integer n .

Proof. If $g(T)$ is separable, then let $g_s = g$, and $n = 0$. Proceeding by induction, if $g(T)$ is not separable, then by Corollary 13.2.9, $g(T) = h(T^p)$, for some $h(T) \in F[T]$. Obviously, $h(T)$ is irreducible, and the minimum polynomial of x^p . Thus, $h(T) = k(T^{p^n})$, with $k(T)$ irreducible and separable over F . Let $k = g_s$, and we're done. ■

Next, we have an elementary characterization of purely inseparable elements.

Proposition 14.3.3. Suppose that $x \in K$, a field extension of the field F , and that $g(T)$ is the minimum polynomial of x over F . The following are equivalent.

- (a) x is purely inseparable over F ;
- (b) $g_s(T) = T - a$, for some $a \in F$;
- (c) $g(T) = T^{p^n} - a$, for some $a \in F$.

Proof. If x is purely inseparable over F , then as $g_s(T)$ is separable over F , it must be linear, otherwise $g(T)$ has at least two distinct roots in its splitting field. Thus, (a) implies (b). Also, (c) follows from (b) and Lemma 14.3.2.

If (c) holds, then $g(T) = (T - x)^{p^n}$, where $x^{p^n} = a$, from which it is clear that x is purely inseparable. ■

Globally we have the following conclusion:

Proposition 14.3.4. Suppose that K is a field extension of F . The following are equivalent.

- (a) Each element of K is purely inseparable over F .
- (b) The only separable elements of K over F are the elements of F .
- (c) K has characteristic p , and for each $x \in K$ there is a positive integer n so that $x^{p^n} \in F$.

Definition 14.3.5. Suppose that K is an algebraic extension of the field F . K is *purely inseparable over F* if every element of K is purely inseparable over F .

Now the main result of this section; an account of how a splitting field “splits” into a separable and a purely inseparable part.

Theorem 14.3.6. *Suppose that K is a splitting field over the field F (and not necessarily of finite dimension). Denote by K^{ins} the subset of K of all purely inseparable elements over F , and K^{sep} the subset of all separable elements over F . Then*

- (a) K^{ins} and K^{sep} are subfields of K , and $K^{\text{ins}} \cap K^{\text{sep}} = K$;
- (b) K is purely inseparable over K^{sep} ;
- (c) K is separable over K^{ins} , and therefore also Galois over K^{ins} ;
- (d) the subfield of K generated by K^{ins} and K^{sep} is K itself;
- (e) K is Galois over K^{ins} , and

$$\text{Gal}(K/K^{\text{ins}}) = \text{Gal}(K^{\text{sep}}/F).$$

Proof. If the characteristic of F is zero, then $K = K^{\text{sep}}$, $F = K^{\text{ins}}$, and there is nothing to prove. So we assume that the characteristic of F is the prime p .

It is clear from (c) of Proposition 14.3.3 that K^{ins} is a subfield of K . As to K^{sep} , the reader is referred to Proposition 13.2.16, with the remark that the proof of that proposition did not require K to be finite dimensional over F . That $K^{\text{ins}} \cap K^{\text{sep}} = F$ should be obvious. Thus, (a) is proved. Lemma 14.3.2 proves (b).

As to (c), if $x \in K$, then according to Lemma 14.3.2, there is a separable irreducible polynomial $g(T)$ over F , and an n , so that $g(T^{p^n})$ is the minimum polynomial of x over F . Let $h(T)$ be the polynomial whose i -th coefficient is the p^n -th root of the i -th coefficient of $g(T)$ (in the algebraic closure of F). Observe that $g(T^{p^n}) = h(T)^{p^n}$. Now, since K is a splitting field over F , Exercise 14.1.4 tells us that $g(T)$ factors completely over K , as does $g(T^{p^n})$. Note that the roots of $g(T^{p^n})$ are precisely the p^n -th roots of the (say k) distinct roots of $g(T)$, and that $h(T)$ has precisely those k p^n -th roots for its zeroes. Since the coefficients of a polynomial are, in turn, polynomials in its roots, it follows that $h(T) \in K[T]$. In fact, the coefficients of $h(T)$ are in K^{ins} . From here one easily concludes that x is separable over K^{ins} . That K is Galois over K^{ins} follows from Proposition 14.1.7. Thus, (c) is proved.

Since K is both separable and purely inseparable over the subfield generated by K^{ins} and K^{sep} , it should be clear that these do generate K itself, proving (d).

Finally (e): since K is separable over K^{ins} , Proposition 14.1.7 once more, shows that K is Galois over K^{ins} . And any automorphism of K over K^{ins} restricts to an automorphism of K^{sep} which fixes F . In view of (d), this restriction map is one-to-one. To argue that it is surjective, apply the appropriate extension theorem for splitting fields to an automorphism of K^{sep} over F , and then observe that any automorphism of K which fixes F must fix K^{ins} , since p^n -th roots are unique. The restriction of maps in $\text{Gal}(K/K^{\text{ins}})$ to $\text{Gal}(K^{\text{sep}}/F)$ is the desired isomorphism of groups.

This completes the proof of this theorem. ■

As a corollary we observe what, surely, must be obvious by now. We leave the proof to the reader.

Corollary 14.3.7. *Suppose that E is a separable algebraic extension of the field F and K a separable algebraic extension of E . Then K is separable over F .*

To conclude this section, a definition and two exercises.

Definition 14.3.8. A field F is said to be *perfect* if every algebraic extension of F is separable. From the above it is easy to see that F is perfect if and only if K^{ins} (in the algebraic closure K of

F) is F itself. Or: F is perfect if and only if F has characteristic zero, or else p (prime), and then every element of F has a p -th root in F .

Note that each finite field is perfect: the field \mathbb{F}_{p^n} of order p^n consists of all the roots of the polynomial $T^{p^n} - T$ (Theorem 13.4.4). However, the field of fractions of $\mathbb{F}_p[T]$ is not perfect.

Exercise 14.3.9. Prove that if E is a finite purely inseparable extension of the field F , then $[E : F]$ is a power of the characteristic p of F .

Exercise 14.3.10. Prove that a finite extension of a perfect field is perfect.

14.4 Artin–Schreier Theory of Real Closed Fields. This material is a classical and beautiful piece of mathematics. We may not have time to develop it in class. I cannot resist including it, however. This development follows [J64], Chapter VI.

Definition 14.4.1. Suppose F is a field endowed with a total ordering \leq , so that

- (i) $a \leq b$ implies that $a + c \leq b + c$, for each $c \in F$;
- (ii) $a \leq b$ implies that $ac \leq bc$, for each $c \geq 0$ in F .

We then call F is an *ordered field*. The set

$$F^+ = \{x \in F : x \geq 0\}$$

is referred to as the *positive cone* of F . If $x \geq 0$ we say that x is *positive*; *negative* elements are defined in the obvious way. If $x > 0$ (resp. $x < 0$) we say that x is *strictly positive* (resp. *negative*).

If a field F admits a total ordering, so that F is an ordered field with respect to it, we say that F is *orderable*.

The first observation about ordered fields is routine, and the proof is left to the reader.

Proposition 14.4.2. *Let F be an ordered field. Then we have the following.*

- (a) *The positive cone F^+ of F is closed under addition and multiplication.*
- (b) *Every square in F is positive.*
- (c) *-1 is not a sum of squares in F .*

Definition & Remarks 14.4.3. A field F which satisfies (c) of Proposition 14.4.2 will be called a *formally real field*. (The adverb “formally” is often suppressed; we will do so.) Thus, every ordered field is formally real; the first goal of this section is to show the converse; namely that on every formally real field, an ordering may be defined, making F an ordered field.

Observe that the field F is formally real if and only if $x_1^2 + \cdots + x_n^2 = 0$ implies that each $x_i = 0$. Note that the field \mathbb{C} of complex numbers cannot be ordered, on account of (c) in Proposition 14.4.2. More generally, no algebraically closed field can be ordered.

A field F is *partially ordered* if there is a partial order on F for which (i) and (ii) in 14.4.1 are satisfied. As we did there, we refer to the set $F^+ = \{x \in F : x \geq 0\}$ as the positive cone of F . The reader should observe that Proposition 14.4.2(a) also holds in partially ordered fields (but not necessarily (b) and (c)). Conversely, suppose that P is a subset of F , subject to

- (po-i) $P + P \leq P$,
 (po-ii) $PP \leq P$, and
 (po-iii) $P \cap -P = \{0\}$.

Define $x \leq y$ if $y - x \in P$; then it is easy to verify that \leq defines a partial ordering on F , with which F becomes a partially ordered field and $F^+ = P$.

Observe now that if F is formally real, then the subset Σ of all sum of squares of F satisfies (po-i), (po-ii) and (po-iii). Note as well that Σ defines the smallest partial ordering, in which every square is positive. Also, $1 > 0$ implies that the characteristic of F is zero.

We summarize the above remarks:

Proposition 14.4.4. *Suppose that F is a formally real field. Then F has characteristic zero, and Σ , the subset of F consisting of all sums of squares, is the positive cone of the least partial ordering on F which every square is positive.*

To prove that a formally real field is orderable we need a Zorn's Lemma argument to extend the cones from partial ones to total ones.

Proposition 14.4.5. *Every formally real field is orderable. Moreover, if P is the positive cone of a partial ordering on F , containing every square, then there is a total ordering on F which extends the given one, in the sense that $P \subseteq F^+$.*

Proof. The second claim obviously implies the first, by letting $P = \Sigma$.

So suppose that P is the positive cone for a partial ordering on F . Let $\Gamma(P)$ be the set of all cones containing P ; that is, $Q \subseteq F$ belongs to $\Gamma(P)$ if $P \subseteq Q$ and Q satisfies (po-i), (po-ii) and (po-iii). We let the reader verify that Zorn's Lemma can be applied to $\Gamma(P)$, with respect to inclusion.

We therefore assume that Q is maximal in $\Gamma(P)$; the aim is to prove that $x \leq y$, defined by $y - x \in Q$ is a total ordering. That is, we must show that if $x \in F$, then either $x \in Q$ or else $-x \in Q$. Note first that if $a \in Q$ and $a \neq 0$, then $a^{-1} \in Q$. For otherwise, the set

$$Q' = \left\{ \frac{a}{b} : a, b \in Q, b \neq 0 \right\}$$

belongs to $\Gamma(P)$, and Q is properly contained in Q' , contradicting the maximality of Q .

Suppose, by way of contradiction, that $x \in F$, and that neither $x \in Q$ nor $-x \in Q$. Let

$$\hat{Q} = \{ a + bx : a, b \in Q \}.$$

We prove that $\hat{Q} \in \Gamma(P)$. It should be clear that (po-i) holds. If a, b, c and d are in Q , then

$$(a + bx)(c + xd) = (ac + bdx^2) + (ad + bc)x;$$

since every square belongs to Q , this means that (po-ii) holds. Finally, let a, b, c, d once again be elements of Q , and suppose that $a + bx = -(c + dx)$. Then $a + c = -(b + d)x$; since $x \neq 0$, it follows that, if $a + c = 0$ then $b + d = 0$, and $a = b = c = d = 0$. If $a + c \neq 0$, then $-x = (a + c)/(b + d) \in Q$, by the comment at the end of the preceding paragraph. This is a contradiction, proving that (po-iii) also holds, and that $\hat{Q} \in \Gamma(P)$. This, in turn, contradicts the maximality of Q , because $x \in \hat{Q} \setminus Q$.

We are forced to conclude that Q is a total ordering on F , and that F is orderable. ■

We come now to the fundamental definition of this section.

Definition 14.4.6. A formally real field F is *real closed* if it has no proper algebraic extensions which are formally real.

The goal of this section is the following theorem, due to Artin and Schreier.

Theorem 14.4.7. *The following are equivalent for a formally real field F .*

- (a) F is real closed.
- (b) F can be ordered so that
 - (i) every positive element of F is a square, and
 - (ii) every polynomial over F of odd degree has a root in F .
- (c) $\sqrt{-1} \notin F$, and $F[\sqrt{-1}]$ is algebraically closed.
- (d) F can be ordered so that, for every polynomial $f(T) \in F[T]$,

$$(IVT) \quad f(a)f(b) < 0, a < b, \Rightarrow \exists c \in F, a < c < b, f(c) = 0.$$

Hot Air 14.4.8. *Inevitable Ruminations.* (a) Theorem 14.4.7 makes mention of a particular ordering; in spite of the fact (guaranteed by Proposition 14.4.5) that formally real fields admit an ordering, in principle, many such orderings are possible. First, and as an initial step in the proof of Theorem 14.4.7, we show – Lemma 14.4.9 – that a real closed field bears only one ordering; it is the set of squares.

(b) The condition in (d) of Theorem 14.4.7 is labelled (IVT), for “Intermediate Value Theorem”. It is the Intermediate Value Theorem for polynomials, at least.

Lemma 14.4.9. *In any real closed field an element is either a square or the negative of a square.*

Proof. Suppose that F is a real closed field, and $x \in F$ is not a square. Consider the proper quadratic extension $F[\sqrt{x}]$. Since $F[\sqrt{x}]$ is a proper extension, it cannot be formally real, and we must therefore have

$$(a_1 + b_1\sqrt{x})^2 + \cdots + (a_k + b_k\sqrt{x})^2 = 0,$$

with not all the a_i and b_i equal to zero. Expanding squares, we see that

$$\sum_{i=1}^k a_i^2 + b_i^2 x = 0 \quad \text{and} \quad \sum_{i=1}^k a_i b_i = 0,$$

since the characteristic of F is zero. Since F is formally real, it follows that $\sum_{i=1}^k b_i^2 \neq 0$, and dividing by this sum of squares we obtain that

$$-x = \frac{\sum_{i=1}^k a_i^2}{\sum_{i=1}^k b_i^2}.$$

Since $y^{-1} = y(y^{-2})$, it is clear that if y is a sum of squares then so is y^{-1} . Thus, $-x$ is itself a sum of squares. Since -1 is not a sum of squares, then neither is x .

We have proved that if x is not a square, then it is not a sum of squares either. Contrapositively: any element which is a sum of squares is itself a square. This means that $-x$ is a square if x is not.

■

Corollary 14.4.10. *If F is a real closed field then it admits only ordering, that which has the set of squares for positive cone. Moreover, any automorphism of F is an order automorphism.*

Proof. Let P be the set of squares of F . We showed in the proof of Lemma 14.4.9 that a sum of squares is a square, which means that (po-i) is satisfied. That (po-ii) holds is clear. Finally, $x \in P \cap -P$ implies that $x = 0$, because F is formally real. Lemma 14.4.9 then insures that the partial ordering induced by P is in fact total.

The final claim is left to the reader. ■

Proof. that (a) \Rightarrow (b) in Theorem 14.4.7. We've already proved (i). So it remains to be shown that every polynomial over F of odd degree has a root in F .

Let $f(T)$ be a polynomial of odd degree. Proceed by induction on the degree n . If $n = 1$, there is nothing to do. If $f(T)$ is reducible then one of the factors must have odd degree, so by induction, it should have a root. So we may assume that $f(T)$ is irreducible.

Let $E = F[u]$, where u is a root of f . If $u \notin F$, then E is not formally real, and so we have polynomials $g_i(T)$ ($1 \leq i \leq m$), all of degree $n - 1$, or less, so that $g_1(u)^2 + \cdots + g_m(u)^2 = -1$, which in turn implies that

$$\sum_{i=1}^m g_i(T)^2 = -1 + f(T)v(T),$$

for suitable $v(T) \in F[T]$. Since the degree on the left is even and at most $2(n - 1)$, that of $v(T)$ is at most $2n - 2 - n = n - 2$, and odd. (Note: we need the assumption that F is formally real, to insure that the sum of the leading coefficients of the $g_i(T)$ is not zero.) Therefore, $v(T)$ has a root c in F , and substituting into the above identity, we contradict the fact that F is formally real. ■

The proof that (b) in Theorem 14.4.7 implies (c) is essentially one of standard proofs that the field of complex numbers is algebraically closed.

Proof. that (b) \Rightarrow (c) in Theorem 14.4.7. Clearly $\sqrt{-1} \notin F$, because it is formally real. We consider $L = F[\sqrt{-1}]$. Let $i = \sqrt{-1}$, and denote the automorphism $a + bi \mapsto a - bi$, as with complex numbers, by $z \mapsto \bar{z}$. Suppose that $f(T) \in L[T]$. If all polynomials with coefficients in F have a root in L , then we're done, because $f(T)\bar{f}(T) \in F[T]$, and a root $u \in L$ of $f(T)\bar{f}(T)$ is either a root of $f(T)$, or else \bar{u} is a root of $f(T)$. So assume that $f(T) \in F[T]$. If the degree of $f(T)$ is odd, there is a root in F , by (ii) in (b).

The next task is to show that every element of L has a square root in L . This is taken care of if $a \in F$, by (b)(i), considering the two possible cases, $a \geq 0$ and $a < 0$. More generally, observe that to solve the equation $a + bi = (c + di)^2$ one must solve $a = c^2 - d^2$ and $b = 2cd$, simultaneously. It suffices to do this with $b = 2$, say; that is, to solve $a = c^2 - d^2$ and $cd = 1$. Or: to solve $a = t - t^{-1}$, where $t = c^2$. Consider now the polynomial $T^2 - aT - 1 \in F[T]$; it has the solution $(a + \sqrt{a^2 + 4})/2 \in F$, since $a^2 + 4 > 0$. In addition, since $4 > 0$, this gives $a + \sqrt{a^2 + 4} > 0$. Thus,

$$c = \sqrt{a + \sqrt{a^2 + 4}}$$

exists in F , and as indicated already, this yields $a = c^2 - d^2$ with $cd = 1$. All of which proves that, in L , every element has a square root. The important consequence of this is that every quadratic polynomial over L splits in L .

Now let E be the splitting field of $(T^2 + 1)f(T)$ over F . We may assume that $L \leq E$. Since F has characteristic zero, E is Galois over F . Let $G = \text{Gal}(E/F)$; note: $|G| = 2^k m$, with $k \geq 1$ and m odd. Let H be a Sylow 2-subgroup of G . Let K be the fixed field of H . Then $[E : K] = 2^k$ and $[K : F] = m$. By (b)(ii) in Theorem 14.4.7, F has no proper odd extension, and so $m = 1$ and $K = F$. So we're down to G being a 2-group, which is nilpotent, and therefore solvable.

Finally, regarding E over L , if $E \neq L$, we put together the Fundamental Theorem of Galois Theory, the facts that L is a splitting field and any 2-group has an element of order 2, to obtain that L has an extension of degree 2. This contradicts the fact that in L every quadratic polynomial splits. Thus, $E = L$, and $k = 1$, proving that every polynomial over F splits in L .

We've therefore shown that L is algebraically closed. ■

To prove that (c) implies (a) in Theorem 14.4.7, we need an observation, the proof of which is left as an exercise to the reader.

Exercise 14.4.11. Suppose that F is a field satisfying (c) in Theorem 14.4.7. Prove that every irreducible polynomial over F has degree 1 or 2.

Proof. **proof that (c) \Rightarrow (a) in Theorem 14.4.7.** Consider the polynomial

$$g(T) = (T^2 - a)^2 + b^2,$$

where $a, b \in F$, and assume that $b \neq 0$. The linear factors of $g(T)$ over $F[\sqrt{-1}]$ are $T \pm a \pm bi$. By Exercise 14.4.11, $g(T)$ is a product of two irreducible quadratic factors (over F). We leave it to the reader to check that the one which has $\sqrt{a+bi}$ as a root must be $(T - \sqrt{a+bi})(T - \sqrt{a-bi})$ or $(T - \sqrt{a+bi})(T + \sqrt{a-bi})$. Either way, $a^2 + b^2$ has a square root in F . This shows that the sum of any two squares of F is a square. Since i is not a square in F (and invoking induction), we conclude that F is formally real.

Finally, any proper algebraic extension of F is isomorphic to $F[i]$, which (being algebraically closed) cannot be formally real. ■

We still have to link (d) in Theorem 14.4.7 with the other conditions. That (d) \Rightarrow (b) is not hard; we leave it as an exercise. It might require a little ingenuity, but nothing deep.

We finish the proof of Theorem 14.4.7 then, by showing that (d) holds in any real closed field.

Proof. **(a), (b) & (c) in Theorem 14.4.7 \Rightarrow (d).** Owing to Exercise 14.4.11, every irreducible polynomial over F (real closed) has degree one or two. From the usual quadratic formula, it is easily verified that a quadratic $T^2 + uT + v$ is irreducible if and only if $u^2 - 4v < 0$. If this is the case, then $4v - u^2 = 4w^2$, for some $w \in F$, and

$$T^2 + uT + v = \left(T + \frac{u}{2}\right)^2 + w^2.$$

The point being that, if $g(T)$ is an irreducible quadratic polynomial, then $g(x) > 0$, for all $x \in F$, or $g(x) < 0$, for all $x \in F$. Now, suppose that $f(T) \in F[T]$ with $f(a)f(b) < 0$ and a and b in F . Factor $f(T)$ into irreducible polynomials over F :

$$f(T) = c(T - a_1) \cdots (T - a_k)g_1(T) \cdots g_n(T),$$

where the $g_i(T)$ are monic quadratic irreducible polynomials. If none of the a_i lie between a and b , then all $T - a_i$ have the same sign at a and b . Since $g_i(x) > 0$ for each $x \in F$, this means that $f(a)$ and $f(b)$ have the same sign, a contradiction. ■

This completes the proof of Theorem 14.4.7.

We close the proceedings with a comment about existence and uniqueness of the so-called real closure. Existence is established in the predictable manner. But, first, let's be clear about the term "real closure".

Definition 14.4.12. Suppose that F is a formally real field. R is a *real closure* of F if R is a real closed field and algebraic over F .

Proposition 14.4.13. *Every formally real field F has a real closure. More precisely, if F^* is the algebraic closure of F , then there is a real closed subfield R of F^* containing F . Necessarily $F^* = R[i]$, where $i = \sqrt{-1}$.*

Proof. The existence is easily settled by applying Zorn's Lemma to formally real subfields of the algebraic closure of F . From Theorem 14.4.7, it is clear that, if R is any real closed subfield of F^* , containing F , then $F^* = R[i]$. ■

Definition & Remarks 14.4.14. *Disappointment?* Uniqueness is a different kettle of fish, however. We shall be content with outlining the arguments. Key to the development is Sturm's Theorem, a classical result, which counts the number of roots, in a given interval, of a polynomial with coefficients in a real closed field. We state it presently, referring the reader to Section 3 of [J64], Chapter VI, for the proof.

Let R be a real closed field, and $f(T)$ be a polynomial over R , of positive degree. We construct the so-called *standard Sturm sequence* for $f(T)$. Let $f_0(T) = f(T)$ and $f_1(T) = f'(T)$, the formal derivative of $f(T)$. Let $f_2(T)$ be the remainder obtained as follows:

$$f_0(T) = f_1(T)q_1(T) - f_2(T), \quad \text{with } f_2 = 0 \quad \text{or} \quad \deg(f_2) < \deg(f_1);$$

and inductively

$$f_{i-1}(T) = f_i(T)q_i(T) - f_{i+1}(T), \quad \text{with } f_{i+1} = 0 \quad \text{or} \quad \deg(f_{i+1}) < \deg(f_i),$$

continuing this iteration of the Euclidean Algorithm until the remainder is zero. The only difference with the usual application of the algorithm, is that the signs of the remainders oscillate. The last entry in the standard Sturm sequence $f_s(T)$ is the greatest common divisor in $R[T]$ of $f(T)$ and its derivative.

Let V_c ($c \in R$) denote the number of sign changes in the sequence $\{f_0(c), f_1(c), \dots, f_s(c)\}$.

Now here is Sturm's Theorem.

Theorem 14.4.15. (Sturm's Theorem) *Suppose that R is a real closed field and that $f(T)$ is any nonconstant polynomial in $R[T]$. Let $\{f_0(T), f_1(T), \dots, f_s(T)\}$ be the standard Sturm sequence for $f(T)$. Finally, suppose that $a < b$ in R , and that $f(a)f(b) \neq 0$. Then the number of distinct roots of $f(T)$ in the interval (a, b) is $V_a - V_b$.*

Now we recast the questions of existence and uniqueness of real closures, taking a narrower perspective.

Remark 14.4.16. Let us be a little clearer, as well as more ambitious, in the formulation of existence and uniqueness of real closures. Suppose that F is an ordered field; we say that the field extension R is a *real closure of F* if R is real closed, algebraic over F , and the ordering of R extends that of F . So the new ingredient in this definition, over 14.4.12, is attention to the ordering with which F is already endowed.

To prove the existence of real closures in this setting we first need the following lemma, the proof of which is sketched.

Lemma 14.4.17. *Suppose that F is an ordered field, and F^* is an algebraic closure of F . Let E be the splitting field in F^* over F of all the polynomials $T^2 - a$, with $a > 0$ in F . Then E is formally real.*

Proof. Prove the ostensibly stronger result that, for any $c_1, c_2, \dots, c_k \in F$ which are positive,

$$\sum_{i=1}^k a_i x_i^2 = 0$$

implies that each $x_i = 0$, whenever each $a_i > 0$ in F , and the $x_i \in F[\sqrt{c_1}, \dots, \sqrt{c_k}]$. Do this by induction on the dimension of the extension $F[\sqrt{c_1}, \dots, \sqrt{c_k}]$. ■

The preceding lemma, plus Sturm's Theorem, give us the grand finale. Before proceeding it is useful to record the following observation:

Exercise 14.4.18. If $f(T) = T^n + a_1 T^{n-1} + \dots + a_n$ is a polynomial over a real closed field R , then all the roots of f in R lie in the interval $(-c, c)$, where $c = n + 1 + a_1^2 + \dots + a_n^2$.

Theorem 14.4.19. Every ordered field F has a real closure. Moreover, if F_1 and F_2 are ordered fields, and $\phi : F_1 \rightarrow F_2$ is an order isomorphism of these fields, and R_1 and R_2 are real closures of F_1 and F_2 , respectively, then ϕ has a unique extension to an isomorphism ϕ' of R_1 onto R_2 ; ϕ' preserves order, necessarily.

Proof. Suppose that F is an ordered field. In any algebraic closure F^* of F , let E be the field generated by all the roots of polynomials of the form $T^2 - a$, with $a > 0$ in F . According to the preceding lemma, E is a formally real field. By Proposition 14.4.13, there is a real closed field R containing E , so that $F^* = R[\sqrt{-1}]$. Now, R is a real closure of F , for if $a > 0$ in F , then $a = b^2$, with $b \in R$ (since the ordering on R is unique), and hence $a > 0$ in R .

To prove the uniqueness claimed here, one first observes the following: if $f_1(T) \in F_1[T]$ is monic, and $f_2(T)$ is the polynomial (over F_2) obtained by mapping the coefficients of f_1 to F_2 , then $f_1(T)$ has the same number of roots in R_1 as f_2 in R_2 . This is where Sturm's Theorem comes in: to begin there is a $c > 0$ in R_1 such that all the roots of f_1 in R_1 lie in the interval $(-c, c)$ (Exercise 14.4.18), and this bound is a polynomial in the coefficients of f_1 . It should now be clear from Sturm's Theorem that the number of roots of f_1 in R_1 , which is also the number of roots of f_1 in the interval $(-c, c)$, is the same as the number of roots of f_2 in $(-\phi(c), \phi(c))$, and that is the total number of roots of f_2 in R_2 .

The next thing to realize is that, for each subset $x_1 < x_2 < \dots < x_k$ of R_1 , there is a subfield E_1 containing F_1 and the x_i , as well as an extension of ϕ to an isomorphism θ of E_1 into R_2 , so that $\theta(x_1) < \dots < \theta(x_k)$. To this end, let $g(T)$ be a polynomial over F_1 , which has among its roots the x_i and also the elements $y_i = \sqrt{x_{j+1} - x_j}$ with $1 \leq j \leq k-1$. Note that the latter belong to R_1 . Let E_1 be the field generated by the roots of $g(T)$ in R_1 . All these fields have characteristic zero, and so, by Exercise 13.4.6, $E_1 = F_1[t_1]$, for a suitable $t_1 \in E_1$. Lift the map ϕ to $\theta : E_1 \rightarrow F_2[t_2]$ as follows: use the minimum polynomial $h_1(T)$ of t_1 over F_1 , and its counterpart $h_2(T)$ over F_2 , and pick any root t_2 of $h_2(T)$. Since any isomorphism takes squares to squares, it follows that $\theta(x_1) < \dots < \theta(x_k)$.

Now we are ready to define the extension of ϕ to R_1 . Pick $x \in R_1$, and let $u_1(T)$ be its minimum polynomial over F_1 . Suppose that $x_1 < \dots < x_k$ are the distinct roots of $u_1(T)$ in R_1 and $x = x_m$. The corresponding polynomial over F_2 , $u_2(T)$, has exactly k roots in R_2 , labelled $y_1 < \dots < y_k$. We define $\phi'(x) = y_m$. It should be easy to see that $\phi : R_1 \rightarrow R_2$ is a bijection, and that ϕ' extends ϕ .

What remains is to show that ϕ' preserves the operations. We have already had occasion to observe that any isomorphism between real closed fields necessarily preserves order.

We sketch the rest: let S be a finite set containing all the roots in R_1 of the minimum polynomials of $x, y, x+y$ and xy . As we have seen, there is an extension θ to a subfield of R_1 , which contains S , and preserves the ordering in S . We leave it to the reader to verify that θ agrees with ϕ' on $x, y, x+y$ and xy , and that the latter is therefore an isomorphism. ■

Hot Air 14.4.20. *Ending on a Note of Caution.* A formally real field may have nonisomorphic real closures, if Definition 14.4.12 is used. For example, let $\mathbb{Q}(\pi)$ be the subfield of \mathbb{R} generated by π . $\mathbb{Q}(\pi)$ bears the natural ordering it inherits from \mathbb{R} . On the other hand, this field can also formally be regarded as the field of fractions of $\mathbb{Q}[T]$. The polynomial ring can be ordered by declaring $f(T) \in \mathbb{Q}[T]$ positive if the nonzero coefficient with the lowest index (corresponding to the lowest power of T) is positive. This is a so-called lexicographic ordering, which admits *infinitesimals*, that is, elements $f > 0$ and $g > 0$, so that $nf < g$, for each $n \in \mathbb{N}$. This ordering extends to $\mathbb{Q}(T) = \mathbb{Q}(\pi)$. Obviously, the real closures of $\mathbb{Q}(\pi)$ with these two different orderings are not isomorphic.

References

- [AM69] M. F. Atiyah & I. G. MacDonald, *Introduction to Commutative Algebra*. Addison–Wesley (1969), Reading, Mass.
- [BS80] S. Burris & V. Sankappanavar, *A Course in Universal Algebra*. Grad. Texts in Math. 78, Springer Verlag (1980), Berlin–Heidelberg–New York.
- [C81] P. M. Cohn, *Universal Algebra*. Math. and its Appl. **6**, Reidel Publ. Co. (1981), Dordrecht, the Netherlands.
- [D95] M. R. Darnel, *The Theory of Lattice–Ordered Groups*. Pure & Appl. Math. **187**, Marcel Dekker (1995), Basel–New York–Hong Kong.
- [Fu63] L. Fuchs, *Partially Ordered Algebraic Structures*. Pergamon Press (1963).
- [Gr79] G. Grätzer, *Universal Algebra*. (2nd Ed.) Springer Verlag (1979), Berlin–Heidelberg–New York.
- [Ha63] P. R. Halmos, *Lectures on Boolean Algebras*. Van Nostrand Math. Studies (1963), Princeton.
- [HS79] H. Herrlich & G. Strecker, *Category Theory*. Sigma Series in Pure Math. **1**, Heldermann Verlag (1979), Berlin.
- [Hu74] T. W. Hungerford, *Algebra*. Holt, Rinehart & Winston (1974), New York.
- [J64] N. Jacobson, *Lectures in Abstract Algebra, III*. Van Nostrand (1964).
- [Ja64] J. P. Jans, *Rings and Homology*. Holt, Rinehart & Winston (1964), New York.
- [Ka68] I. Kaplansky, *Infinite Abelian Groups*. (Revised Ed.) U. of Mich. Press (1968), Ann Arbor.
- [Ka69] I. Kaplansky, *Fields and Rings*. Chicago Lect. Math., Univ. of Chicago Press (1969), Chicago–London.
- [L86] J. Lambek, *Lectures on Rings and Modules*. (3rd Ed.) Chelsea Publ. Co. (1986), New York.
- [LM71] M. Larsen & P. McCarthy, *Multiplicative Theory of Ideals*. Pure & Applied Math. **43**, Academic Press (1971), New York & London.
- [LZ71] W. A. J. Luxemburg & A. C. Zaanen, *Riesz Spaces, I*. North Holland (1971), Amsterdam.
- [Nm67] H. Neumann, *Varieties of Groups*. Ergebnisse der Math., **37** Springer Verlag, (1967) Berlin–Heidelberg–New York.
- [P68] R. S. Pierce, *Introduction to the Theory of Abstract Algebras*. Holt, Rinehart & Winston (1968), New York.
- [Sh77] I. R. Shafarevich, *Basic Algebraic Geometry*. Springer Study Ed., Springer Verlag (1977), Berlin–Heidelberg–New York.
- [Si60] R. Sikorski, *Boolean Algebras*. Ergebnisse der Math. **25** Springer Verlag (1960), Berlin–Heidelberg–New York.
- [Wi70] S. Willard, *General Topology*. Addison–Wesley (1970), Reading, Mass.