ALGEBRAS, LATTICES, VARIETIES

VOLUME I

Ralph N. McKenzie
University of California, Berkeley

George F. McNulty
University of South Carolina

Walter F. Taylor
University of Colorado

1987

Contents

C H A P T E R O N E

Basic Concepts

1.1 Algebras and Operations

An algebra is a set endowed with operations. algebras are the fundamental object with which we shall deal, so our first step is to make the preceding sentence precise. Let A be a set and n be a natural number. An **operation of rank** n **on** A is a function from A^n into A. Here we have used A^n to denote the n-fold direct power of A—the set of all n-tuples of elements of A. By a (**finitary**) **operation on** A we meanan operation of rank n on A for some natural number n. Because virtually every operation taken up in this book will be finitary, we will generally omit the word "finitary" and use "operation" to mean finitary operation. If A is nonempty, then each operation on A has a unique rank. Operations of rank 0 on a nonempty set are functions that have only one value; that is they are constant functions of a rather special kind. We call operations of rank 0 **constants** and identify them with their unique values. Similarly, we call operations of rank 1 on A **unary operations** and identify them with the functions from A into A. **Binary** and **ternary** operations are operations of rank 2 and 3, respectively. We use n**-ary operation, operation of rank** n and **operation of arity** n interchangeably. It is important to realize that the domain on an operation of rank n on A is the whole set A^n. Functions from a subset of A^n into A are called **partial** operations of rank n on A. Subtraction is a binary operation on the set $\mathbb{Z}$ of integers, but it is only a partial operation on the set ω of natural numbers.

The fundamental operations most frequently encountered in mathematics have very small ranks. A list of these important operations certainly includes addition, subtraction, multiplication, division, exponentiation, negation, conjugation, etc., on appropriate sets (usually sets of numbers, vectors, or matrices). This list should also include such operations as forming the greatest divisor of two natural numbers, the composition of two functions, and the union of two sets. Of course, one is almost immediately confronted with operations of higher rank that are compounded from these. Operations of higher finite rank whose mathematical significance does not depend on how they are built up from op-

erations of smaller rank seem, at first, to be uncommon. Such operations will emerge later in this work, especially in Chapter 4 and in later volumes. However, even then most of the operations will have ranks no larger than 5. While there is some evidence that operations of such small rank provide adequate scope for the development of a general theory of algebra, why this might be so remains a puzzle.

To form algebras, we plan to endow sets with operations. There are several ways to accomplish this. We have selected one that, for most of our purposes, leads to clear and elegant formations of concepts and theorems.

DEFINITION 1.1. An **algebra** is an ordered pair $\langle A, F\rangle$ such that A is a nonempty set and $F = \langle F_i : i \in I\rangle$ where F_i is a finitary operation on A for each $i \in I$. A is called the **universe** of $\langle A, F\rangle$, F_i is referred to as a **fundamental** or **basic operation** of $\langle A, F\rangle$ for each $i \in I$, and I is called the **index set** or the **set of operation symbols** of $\langle A, F\rangle$.

The reason we have endowed our algebras with *indexed* systems of operations rather than with mere sets of operations is so that we have a built-in means to keep the operations straight. From the customary viewpoint, rings have two basic binary operations labeled "addition" (or $+$) and "multiplication" (or $\cdot$). For the development of ring theory, it is essential to distinguish these operations from each other. In effect, most expositions do this by consistent use of the symbols $+$ and $\cdot$: The actual binary operations in any given ring are *indexed* by these symbols. This is why we have chosen to call the set of $\langle A, F\rangle$ its set of operation symbols. The distinction between operation symbols and operations is important, and we will have much to say regarding it in §4.11.

The notation implicit in the definition above is unwieldly in most situations. Quite often the set of operation symbols is small. For example, ring theory is accommodated by the operation symbols $+$, $\cdot$, and $-$ (this last symbol is intended to name the unary operation of negation). But surely

$$\langle \mathbb{Z}, \langle F_i : i \in \{+, \cdot, -\}\rangle\rangle$$

is an uncomfortable way to display the ring of integers. In this situation and others like it, we find

$$\langle \mathbb{Z}, +, \cdot, -\rangle$$

much more acceptable. Notice that in this last display $+$, $\cdot$, and $-$ are no longer operation symbols but operations; exactly which operations they are is clear from the context.

As a general convention, we use uppercase boldface letters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\cdots$ to denote algebras and the corresponding uppercase letters A, B, C, $\cdots$ to denote their universes, attaching subscripts as needed. Thus, in most uses, A is the universe of $\mathbf{A}$, B_3 is the universe of $\mathbf{B}_3$, and so on. If Q is an operation symbol of $\mathbf{A}$, then we use $Q^{\mathbf{A}}$ to stand for the fundamental operation of $\mathbf{A}$ indexed by Q; we say that Q **denotes** $Q^{\mathbf{A}}$ or that $Q^{\mathbf{A}}$ is the **interpretation** of Q in $\mathbf{A}$. Whenever the cause of clarity or the momentum of customary usage dictates, we will depart from these conventions.

Given an algebra $\mathbf{A}$ with index set I, there is a function ρ called the **rank function** from I into the set ω of natural numbers defined by:

$$\rho(Q) \text{ is the rank of } Q^{\mathbf{A}} \text{ for all } Q \in I.$$

The rank function of an algebra is also referred to as its **similarity type** or, more briefly, its **type**. Algebras $\mathbf{A}$ and $\mathbf{B}$ are said to be **similar** if and only if they have the same rank function. The similarity relation between algebras is an equivalence relation whose equivalence classes will be called **similarity classes**. Most of the time (with some important exceptions), only algebras of the same similarity type will be under consideration. In fact, this hypothesis that all algebras at hand are similar is so prevalent that we have left it unsaid, even in the statement of some theorems.

The rank functions are partially ordered by set inclusion (that is, by extension of functions). This ordering can be imposed on the similarity class as well. For individual algebras, we say that $\mathbf{A}$ is a **reduct** of $\mathbf{B}$ (and that $\mathbf{B}$ is an **expansion** of $\mathbf{A}$) if and only if $\mathbf{A}$ and $\mathbf{B}$ have the same universe, the rank function of $\mathbf{A}$ is a subset of the rank function of $algB$, and $Q^{\mathbf{A}}$ and $Q^{\mathbf{B}}$ for all operation symbols Q of $\mathbf{A}$. In essence, this means that $\mathbf{B}$ can be obtained by adjoining more basic operations to $\mathbf{A}$. For example, each ring is an expansion of some Abelian group.

We close this section with a series of examples of algebras. Besides illustrating the notions just introduced, these examples specify how we formalize various familiar kinds of algebras and serve as resources for later reference. In formulating the examples, we use the following operation symbols:

Constant symbols: e, 1, 0, and $1'$

Unary operation symbols: $-$, $^{-1}$, $^{-}$, $^{\cup}$, and f_r for each $r \in R$

Binary operation symbols: $+$, $\cdot$, $\wedge$, $\vee$

Semigroups

A **semigroup** is an algebra $\mathbf{A} = \langle A, \cdot^{\mathbf{A}} \rangle$ such that:

$$\left(a \cdot^{\mathbf{A}} b\right) \cdot^{\mathbf{A}} c = a \cdot^{\mathbf{A}} \left(b \cdot^{\mathbf{A}} c\right) \text{ for all } a, b, c \in A.$$

Thus a semigroup is a nonempty set endowed with an associative binary operation. A typical example of a semigroup is the collection of all functions from X into X, where X is any set, with the operation being composition of functions. A more sophisticated example is the collection of all $n \times n$ matrices of integers endowed with matrix multiplication.

Monoids

A **monoid** is an algebra $\mathbf{A} = \langle \cdot^{\mathbf{A}}, e^{\mathbf{A}} \rangle$ such that $\langle A, \cdot^{\mathbf{A}} \rangle$ is a semigroup and

$$a \cdot^{\mathbf{A}} e^{\mathbf{A}} = e^{\mathbf{A}} \cdot^{\mathbf{A}} a = a \text{ for all } a \in A.$$

To obtain some concrete examples, we can let e denote the identity function in our first example of a semigroup and the identity matrix in the second example. Although every monoid is an expansion of a semigroup, not every semigroup can be expanded to a monoid.

Groups

A **group** is an algebra $\mathbf{A} = \langle A, \cdot, ^{-1}, e\rangle$ such that $\langle A, \cdot, e\rangle$ is a monoid and

$$a \cdot a^{-1} = a^{-1} \cdot a = e \text{ for all } a \in A.$$

(The reader will have observed that the superscript $^{\mathbf{A}}$ has grown tiresome and been dropped.) A typical example of a group is the collection of all one-to-one functions from X onto X (such functions are called **permutations** of X), where X is an arbitrary set and the operations are composition of functions, inversion of functions, and the identity function. A more intricate example is the collection of isometries (also called distance-preserving functions) of the surface of the unit ball in ordinary Euclidean three-dimensional space endowed with the same sorts of basic operations. Groups have been construed as algebras of several different similarity types. Most of the popular renditions of group theory define groups as certain special kinds of semigroups. Our choice of fundamental operations was motivated by the desire to have the class of groups turn out to be a variety. Still, there are a number of quite satisfactory ways to present the intuitive notion of a group. For instance, there is no real need to devote a basic operation to the unit element. It is even possible to make do with a single binary operation, though this cannot be $\cdot$. The sense in which such formulations are equivalent, not only for groups but for algebras in general, is made precise in §4.12. Some interesting aspects of semigroups and groups are explored in Chapter 3.

Rings

A **ring** is an algebra $\langle A, +, \cdot, -, 0\rangle$ such that $\langle A, +, -, 0\rangle$ is an Abelian group, $\langle A, \cdot\rangle$ is a semigroup, and

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \text{ for all } a, b, c \in A$$

and

$$(b + c) \cdot a = (b \cdot a) + (c \cdot a) \text{ for all } a, b, c \in A.$$

A **ring with unit** is an algebra $\langle A, +, \cdot, -, 0, 1\rangle$ such that $\langle A, +, \cdot, -, 0\rangle$ is a ring and $\langle A, \cdot, 1\rangle$ is a monoid. A familiar example of a ring (with unit) is the integers endowed with the familiar operations. Another example is the set of $n \times n$ matrices with real entries endowed with the obvious operations. We regard fields as special kinds of rings.

Vector Spaces and Modules

In the familiar treatments, vector spaces and modules are equipped with a binary operation called addition and a scalar multiplication subject to certain conditions.

As ordinarily conceived, the scalar multiplication is not what we have called an operation. An easy way around this trouble is to regard scalar multiplication as a schema of unary operations, one for each scalar. Actually, this is in accord with geometric intuition by which these operations amount to stretching or shrinkingly. Let

$$\mathbf{R} = \langle R, +, \cdot, -, 0, 1 \rangle$$

be a ring with unit. An **R-module** (sometimes called a left unitary **R**-module) is an algebra $\langle M, +, -, 0, f_r \rangle_{r \in R}$ such that $\langle M, +, -, 0 \rangle$ is an Abelian group and for all $a, b \in M$ and for all r, s, and $t \in R$ the following equalities hold:

$$\begin{aligned} f_r(f_s(a)) &= f_t(a) \qquad \text{where } r \cdot s = t \text{ in } \mathbf{R} \\ f_r(a+b) &= f_r(a) + f_r(b) \\ f_r(a) + f_s(a) &= f_t(a) \qquad \text{where } r + s = t \text{ in } \mathbf{R} \\ f_1(a) &= a. \end{aligned}$$

In essence, what these conditions say is that f_r is an endomorphism of the Abelian group $\langle M, +, -, 0 \rangle$ for each $r \in R$, that this collection of endomorphisms is itself a ring with unit, and that mapping $r \mapsto f_r$ is a homomorphism from **R** onto this ring. Although we will soon formulate such notions as homomorphism in our general setting, this last sentence is meaningful as it stands in the special context of ring theory. part of the importance of modules lies in the fact that very ring is, up to isomorphism, a ring of endomorphisms of some Abelian group. This fact is analogous to the more familiar theorem of Cayley to the effect that every group is isomorphic to a group of permutations of some set. In the event that **R** is a field, we call the **R**-modules **vector spaces over R**.

Bilinear Algebras over a Field

Let $\mathbf{F} = \langle F, +, \cdot, -, 0, 1 \rangle$ be a field. An algebra $\mathbf{A} = \langle A, +, \cdot, -, 0, f_r \rangle_{r \in F}$ is a **bilinear algebra over F** if $\langle A, +, -, 0, f_r \rangle_{r \in F}$ is a vector space over **F** and for all $a, b, c \in A$ and all $r \in F$:

$$\begin{aligned} (a+b) \cdot c &= (a \cdot c) + (b \cdot c) \\ c \cdot (a+b) &= (c \cdot a) + (c \cdot b) \\ f_r(a \cdot b) &= (f_r(a)) \cdot b = a \cdot f_r(b). \end{aligned}$$

If, in addition, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in A$, then **A** is called an **associative algebra over F**. Thus an associative algebra over a field has both a vector space reduct and a ring reduct. An example of an associative algebra can be constructed from the linear transformations of any vector space into itself. A concrete example of this kind is obtained by letting A be the set of all 2×2 matrices over the field of real numbers and taking the natural matrix operations. Lie algebras, Jordan algebras, and alternative algebras provide important examples of non-associative bilinear algebras that have arisen in connection with physics and analysis. A **Lie algebra** is a bilinear algebra that satisfies two further equal-

ities:

$$a \cdot a = 0 \text{ for all } a \in A$$
$$((a \cdot b) \cdot c) + ((b \cdot c) \cdot a) + ((c \cdot a) \cdot b) = 0 \text{ for all } a, b, c \in A.$$

Suppose $\langle A, +, \cdot, -, 0, f_r \rangle_{r \in F}$ is an associative algebra over $\mathbf{F}$. Define

$$a * b = (a \cdot b) + (-(b \cdot a)) \text{ for all } a, b \in A.$$

It is not difficult to verify that $\langle A, +, *, -, 0, f_r \rangle_{r \in F}$ is a Lie algebra. A good but brief introduction to bilinear algebras is available in [] [Jacobson 1985]. [] [Pierce 1982] offers an excellent account of associative algebras over fields. We remark that a common usage of the word "algebra" in the mathematical literature is to refer to those mathematical structures we have called "bilinear algebras over fields." The objects that we have called "algebras" are then referred to as "universal algebras" (although there is nothing especially universal about, say, the three-element group, which is one of these objects) or as "Ω-algebras" (perhaps because Ω is a kind of Greek abbreviation for "operation").

The establishment of the theories of groups, rings, fields, vector spaces, modules, and various kinds of bilinear algebras over fields is a sterling accomplishment of nineteenth-century mathematics. This line of mathematical research can be said to have reached its maturity at the hands of such mathematicians as Hilbert, Burnside, Frobenius, Wedderburn, Noether, van der Waerden, and E. Artin by the 1930s. It has continued to grow in depth and beauty, being today one of the most vigorous mathematical enterprises.

There is another important series of examples of algebras, different in character from those described above.

Semilattices

A **semilattice** is a semigroup $\langle A, \wedge \rangle$ with the properties

$$a \wedge b = b \wedge a \text{ for all } a, b \in A$$

and

$$a \wedge a = a \text{ for all } a \in A.$$

A typical example of a semilattice is formed by taking A to be the collection of all subsets of an arbitrary set with the operation being intersection. Another example is formed by taking A to be the compact convex sets on the Euclidean plane and the operation to be the formation of the closed convex hull of the union of two compact convex sets.

Lattices

A **lattice** is an algebra $\langle A, \wedge, \vee \rangle$ such that both $\langle A, \wedge \rangle$ and $\langle A, \vee \rangle$ are semilattices and the following two equalities hold:

$$a \vee (a \wedge b) = a \text{ for all } a, b \in A$$

and

$$a \wedge (a \vee b) = a \text{ for all } a, b \in A.$$

A typical example of a lattice is formed by taking A to be the collection of all equivalence relations on an arbitrary set, $\wedge$ to be intersection, and $\vee$ to be the transitive closure of the union of two given equivalence relations. Another example is formed by taking A to be the set of natural numbers, $\vee$ to be the formation of the least common multiples, and $\wedge$ to be the formation of greatest common divisors. Lattices have a fundamental role to play in our work. Chapter 2 is devoted to the elements of lattice theory. The operation $\wedge$ is referred to as **meet**, and the operation $\vee$ is called **join**.

Boolean Algebras

A **Boolean algebra** is an algebra $\langle A, \wedge, \vee, ^- \rangle$ such that $\langle A, \wedge, \vee \rangle$ is a lattice and for all $a, b, c \in A$ the following equalities hold:

$$\begin{aligned} a \wedge (b \vee c) &= (a \wedge b) \vee (a \wedge c) \\ a \vee (b \wedge c) &= (a \vee b) \wedge (a \vee c) \\ (a \vee b)^- &= a^- \wedge b^- \\ a^{--} &= a \\ (a^- \wedge a) \vee b &= b \\ (a^- \vee a) \wedge b &= b. \end{aligned}$$

Thus a Boolean algebra is a distributive lattice with a unary operation of complementation (denoted here by $^-$) adjoined. As an example, take A to be the collection of all subsets of an arbitrary set X, let the join $\vee$ be union, the meet $\wedge$ be intersection, and the complementation $^-$ be set complementation relative to X. Another example is afforded by the clopen (simultaneously open and closed) subsets of a topological space under the same operations as above.

Relation Algebras

A **relation algebra** is an algebra $\langle A, \wedge, \vee, \cdot, ^-, ^\cup, 1' \rangle$ such that $\langle A, \wedge, \vee, ^- \rangle$ is a Boolean algebra, $\langle A, \cdot, 1' \rangle$ is a monoid, and the following equalities hold for all $a, b, c \in A$:

$$\begin{aligned} a \cdot (b \vee c) &= (a \cdot b) \vee (a \cdot c) \\ (a \cdot b)^\cup &= b^\cup \cdot a^\cup \\ (a^\cup)^\cup &= a \\ (1')^\cup &= 1' \\ (a \vee b)^\cup &= a^\cup \vee b^\cup \\ \left(a^\cup \cdot (a \cdot b)^-\right) \wedge b &= a^- \wedge a. \end{aligned}$$

An example of a relation algebra can be formed by taking A to be the collection of all binary relations on an arbitrary set X, giving A the Boolean operations by

regarding A as the power set of X^2, and defining the remaining operations so that

$$\begin{aligned} R \cdot S &= \{\langle x,y\rangle : \langle x,z\rangle \in R \text{ and } \langle z,y\rangle \in S \text{ for some } z \in X\} \\ R^{\cup} &= \{\langle x,y\rangle : \langle y,x\rangle \in R\} \\ 1' &= \{\langle x,x\rangle : x \in X\}. \end{aligned}$$

The relation algebra obtained in this way from the set X is sometimes denoted by $\mathrm{Rel}X$. [] [Jónsson 1982] provides a thorough and very readable overview. Relation algebras have a rich theory, indeed rich enough to offer a reasonable algebraic context for the investigation of set theory. The essential idea for such an investigation is to regard set theory as the theory of membership. Membership is a binary relation. Adjoining an additional constant (nullary operation) e to the type of relation algebras to stand for membership opens the possibility of developing set theory by distinguishing e from other binary relations by means of the algebraic apparatus of relation algebras. It turns out to be possible, for example, to render the content of Zermelo-Fraenkel axioms for set theory entirely as equations in this setting. The deep connections these algebras have with the foundations of mathematics emerges in the monograph of [][Tarski and Givant 1987].

Our list of examples of algebras ends here, having merely touched on a small section. Further kinds of algebras will be introduced from time to time to serve as examples and counterexamples.

Exercises 1.2

1. Let A be a nonempty set and Q be a finitary operation on A. Prove that the rank of Q is unique.

2. Let A be a nonempty set. Describe the operations on A of rank 0 and in set-theoretic terms.

3. Construct a semigroup that cannot be expanded to a monoid.

4. Construct a semigroup that is not the multiplicative semigroup of any ring.

5. Prove that every ring can be embedded in a ring that can be expanded to a ring with unit.

6. Let $\langle A,+,\cdot,-,0,f_r\rangle_{r\in R}$ be an associative algebra over the field $\mathbf{F}$ and define

$$a * b = (a \cdot b) + (-(b \cdot a)) \text{ for all } a,b \in A.$$

Prove that $\langle A,+,*,-,0,f_r\rangle_{r\in R}$ is a Lie algebra.

7. Let A be a set and denote by Eqv A the set of all equivalence relations on A. For $R, S \in \text{Eqv } A$ define

$$R \wedge S = R \cap S$$
$$R \vee S = R \cup R \circ S \cup R \circ S \circ R \cup R \circ S \circ R \circ S \cup \cdots$$

where $\circ$ stands for relational product (that is, $a(R \circ S)b$ means that there is some c such that both aRc and cSb). Prove that $\langle \text{Eqv } A, \wedge, \vee \rangle$ is a lattice.

1.2 Subalgebras, Homomorphisms, and Direct Products

One of the hallmarks of algebraic practice is the prominent role played by relationships holding among algebras. Some of the subtleties of complicated algebras can be more readily understood if some tractable way can be found to regard them as having been assembled from less complicated, more thoroughly understood algebras. The chief tools we will use to assemble new algebras from those already on hand are the formation of subalgebras, the formation of homomorphic images, and the formation of direct products. The reader is probably familiar with these notions in the settings of groups, rings, and vector spaces. They fit comfortably into our general setting.

Let F be an operation of rank r on the nonempty set A, and let X be a subset of A. We say that X is **closed with respect to** F (also that F **preserves** X and that X is **invariant under** F) if and only if

$$F(a_0, a_1, \ldots, a_{r-1}) \in X \text{ for all } a_0, a_1, \ldots, a_{r-1} \in X.$$

In the event that F is constant, this means that X is closed with respect to F if and only if $F \in X$. Thus the empty set is closed with respect to every operation on A of positive rank, but it is not closed with respect to any operation of rank 0.

Taking A to be the set of integers, we see that the set of odd integers is closed with respect to multiplication but not with respect to addition.

DEFINITION 1.3. Let $\mathbf{A}$ be an algebra. A subset of the universe A of $\mathbf{A}$, which is closed with respect to each fundamental operation of $\mathbf{A}$, is called a **subuniverse** of $\mathbf{A}$. The algebra $\mathbf{B}$ is said to be a **subalgebra** of $\mathbf{A}$ if and only if $\mathbf{A}$ and $\mathbf{B}$ are similar, the universe B of $\mathbf{B}$ is a subuniverse of $\mathbf{A}$, and $Q^{\mathbf{B}}$ is the restriction to B of $Q^{\mathbf{A}}$, for each operation symbol Q of $\mathbf{A}$. Sub $\mathbf{A}$ denotes the set of all subuniverses of $\mathbf{A}$.

The ring of integers is a subalgebra of the ring of complex numbers.

"$\mathbf{B}$ is an **extension** of $\mathbf{A}$" means that $\mathbf{A}$ is a subalgebra of $\mathbf{B}$; we render this in symbols as $\mathbf{A} \subseteq \mathbf{B}$. This convenient abuse of symbols should not lead to ambiguity–if nothing else, the boldface characters convey the algebraic intent.

our system of conventions exposes us, from time to time, to the minor annoyance of subuniverses that are not universes of subalgebras. We insisted that

the universes of algebras be nonempty, but we had also insisted on empty subuniverses (exactly when there are no operations of rank 0). By accepting this incongruity we avoid the need to single out many special cases in the statements of definitions and theorems.

The notation of subalgebra defined above occasionally conflicts, at least in spirit, with common usage. Consider the case of fields. We have regarded fields as rather special sorts of rings, but the possibility of putting the function that sends each nonzero element to its multiplicative inverse on the same distinguished footing as the ring operations is certainly enticing. We have not taken this step, since the function involved is only a partial operation, and the resulting mathematical system would not fall within our definition of an algebra. This deviation from our definition may seem small–from many viewpoints it is–but to widen the definition so as to allow partial operations would result in a havoc of technical complications and substantially alter the character of the ensuing mathematics. We have the option of declaring that $0^{-1} = 0$ in order to force the operation to be defined everywhere, but this invalidates many equalities one ordinarily thinks of as holding in fields. For example,

$$x^{-1}x = 1$$

would not hold when $x = 0$ unless the field has only one element. in any case, fields generally have subalgebras that are not fields. The integers come form a subalgebra of the field of complex numbers that is not a subfield. In connection with the examples of algebras that concluded the previous section, the situation is very pleasant: Every subalgebra of a group, a ring, a vector space, etc., is again an algebra of the same sort.

Now consider similar algebras $\mathbf{A}$ and $\mathbf{B}$ and let Q be an operation symbol of rank r. A function h from A into B is said to **respect the interpretation of** Q if and only if

$$h\left(Q^{\mathbf{A}}(a_0,\cdots,a_{r-1})\right) = Q^{\mathbf{B}}\left(h(a_0),\cdots,h(a_{r-1})\right)$$

for all $a_0,\ldots,a_{r-1} \in A$.

DEFINITION 1.4. Let $\mathbf{A}$ and $\mathbf{B}$ be similar algebras. A function h from A into B is called a **homomorphism** from A into B if and only if h respects the interpretation of every operation symbol of $\mathbf{A}$. $\hom(\mathbf{A},\mathbf{B})$ denotes the set of all homomorphisms from $\mathbf{A}$ into $\mathbf{B}$.

We distinguish several kinds of homomorphisms and employ notation for them as follows. Let $\mathbf{A}$ and $\mathbf{B}$ be similar algebras. Each of

$$h : \mathbf{A} \to \mathbf{B}$$
$$\mathbf{A} \xrightarrow{h} \mathbf{B}$$
$$h \in \hom(\mathbf{A},\mathbf{B})$$

denotes that h is a homomorphism from $\mathbf{A}$ into $\mathbf{B}$. By attaching a tail to the arrow, we express the condition of one-to-oneness of h; by attaching a second

head to the arrow, we express the condition that h is onto B. Thus both

$$h : \mathbf{A} \rightarrowtail \mathbf{B}$$

and

$$\mathbf{A} \overset{h}{\rightarrowtail} \mathbf{B}$$

denote that h is a one-to-one homomorphism from $\mathbf{A}$ into $\mathbf{B}$. We call such homomorphisms **embeddings**. Likewise, both

$$h : \mathbf{A} \twoheadrightarrow \mathbf{B}$$

and

$$\mathbf{A} \overset{h}{\twoheadrightarrow} \mathbf{B}$$

denote that h is a homomorphism from $\mathbf{A}$ onto $\mathbf{B}$, and in this case we say that $\mathbf{B}$ is the **homomorphic image** of $\mathbf{A}$ under h. Further, each of

$$h : \mathbf{A} \rightarrowtail\!\!\!\!\twoheadrightarrow \mathbf{B}$$

$$\mathbf{A} \overset{h}{\rightarrowtail\!\!\!\!\twoheadrightarrow} \mathbf{B}$$

and

$$\mathbf{A} \overset{h}{\cong} \mathbf{B}$$

denote that h is a one-to-one homomorphism from $\mathbf{A}$ onto $\mathbf{B}$. We call such homomorphisms **isomorphisms**. $\mathbf{A}$ and $\mathbf{B}$ are said to be **isomorphic**, which we denote by $\mathbf{A} \cong \mathbf{B}$, iff there is an isomorphism from $\mathbf{A}$ onto $\mathbf{B}$. A homomorphism from $\mathbf{A}$ into $\mathbf{A}$ is called an **endomorphism** of $\mathbf{A}$, and an isomorphism from $\mathbf{A}$ onto $\mathbf{A}$ is called an **automorphism** of $\mathbf{A}$. $\operatorname{End}\mathbf{A}$ and $\operatorname{Aut}\mathbf{A}$ denote, respectively, the set of all endomorphisms of $\mathbf{A}$ and the set of all automorphisms of $\mathbf{A}$. The identity map 1_A belongs to each of these sets; moreover, each of these sets is closed with respect to composition of functions. In addition, each automorphism of $\mathbf{A}$ is an invertible function and its inverse is also an automorphism. Thus $\langle \operatorname{End}\mathbf{A}, \circ, 1_A \rangle$ is a monoid, which we shall designate by $\mathbf{End}\,\mathbf{A}$, and $\langle \operatorname{Aut}\mathbf{A}, \circ, ^{-1}, 1_A \rangle$ is a group, which we shall designate by $\mathbf{Aut}\,\mathbf{A}$.

$\langle \mathbb{R}^+, \cdot \rangle$ and $\langle \mathbb{R}, + \rangle$ are isomorphic, where $\mathbb{R}$ is the set of real numbers and $\mathbb{R}^+$ is the set of positive real numbers. Indeed, the natural logarithm function is an isomorphism that illustrates this fact. fact.

An isomorphism is a one-to-one correspondence between the elements of two algebras that respects the interpretation of each operation symbol. This means that with regard to a host of properties, isomorphic algebras are indistinguishable from each other. This applies to most of the properties with which we shall deal; if they are true in a given algebra, then they are true for all isomorphic images of that algebras a well. Such properties have been called "algebraic properties."

On the other hand, algebras that are isomorphic can be quite different from each other. For example, the set of all twice continuously differentiable real valued functions of a real variable that are solutions to the differential equation

$$\frac{d^2 f}{dx^2} + f = 0$$

can be given the structure of a vector space over the field of real numbers. This vector space is isomorphic to the two-dimensional space familiar from Euclidean plane geometry. Roughly speaking, the distinction between these two isomorphic vector spaces can be traced to the "internal" structure of their elements: on the one hand, functions, and on the other hand, geometric points. The notion that functions can be regarded as points with a geometric character is a key insight, not only for differential equations but also for functional analysis. Because the internal structure of the elements of an algebra can be used to establish algebraic properties, some of the most subtle and powerful theorems of algebra are those that assert the existence of isomorphisms.

Isomorphism is an equivalence relation between algebras, and the equivalence classes are called **isomorphic types**. Isomorphism is a finer equivalence relation than similarity, in the sense that if two algebras are isomorphic, then they are also similar. In fact, among equivalence relations holding between algebras, isomorphism is probably the finest we will encounter; similarity is one of the coarsest.

The formation of subalgebras and of homomorphic images seems to lead to algebras that are no more complicated than those with which the constructions started. By themselves, these constructions do not offer the means to form larger, more elaborate algebras. The direct product construction allows us to construct seemingly more elaborate and certainly larger algebras from systems of smaller ones.

Let I be any set and let A_i be a set for each $i \in I$. The system $A = \langle A_i : i \in I \rangle$ is called a **system of sets indexed by** I. By a **choice function** for A we mean a function f with domain I such that $f(i) \in A_i$ for all $i \in I$. The **direct product** of the system A is the set of all choice functions for A. The direct product of A can be designated in any of the following ways:

$$\prod A, \prod_{i \in I} A_i, \text{ or } \prod_I A_i.$$

Each set A_i for $i \in I$ is called a **factor** of the direct product. For each $i \in I$, the **i^{th} projection function**, denoted by p_i, is the function with domain $\prod A$ such that $p_i(f) = f(i)$, for all $f \in \prod A$. Sometimes we refer to members of $\prod A$ as **I-tuples** from A, and we write f_i in place of $f(i)$. Observe that if A_i is empty for some $i \in I$, then $\prod A$ is empty. Also note that if I is empty, then $\prod A$ has exactly one element: the empty function. If $A_i = B$, for all $i \in I$, then $\prod A$ is also denoted by B^I and referred to as a **direct power** of B. In the event that $I = \{0,1\}$, we use $A_0 \times A_1$ to denote $\prod A$.

Now let $|I$ be a set and let $\mathbf{A}_i$ be an algebra for each $i \in I$. Moreover, suppose that $\mathbf{A}_i$ and $\mathbf{A}_j$ are similar whenever $i, j \in I$. So $\langle \mathbf{A}_i : i \in I \rangle$ is a system of similar algebras indexed by I. We create the direct product of this system of algebras by imposing operations on $\prod A$ coordinatewise. This is the unique choice of operations on the product set for which each projection function is a homomorphism.

DEFINITION 1.5. Let $\mathbf{A} = \langle \mathbf{A}_i : i \in I \rangle$ be a system of similar algebras. The **direct product** of $\langle \mathbf{A}_i : i \in I \rangle$ is the algebra, denoted by $\prod \mathbf{A}$, with the same similarity type, with universe $\prod A$ such that for each operation symbol Q and all

$f^0, f^1, \ldots, f^{r-1} \in \prod A$, where r is the rank of Q,

$$\left(Q^{\prod A}(f^0, f^1, \cdots, f^{r-1})\right)_i = Q^{\mathbf{A}_i}\left(f_i^0, f_i^1, \cdots, f_i^{r-1}\right)$$

for all $i \in I$.

Here are some alternatives for denoting products:

$$\prod_{i \in I} \mathbf{A}_i, \text{ or } \prod_I \mathbf{A}_i.$$

If $\mathbf{B} = \mathbf{A}_i$ for all $i \in I$, we write $\mathbf{B}^I$ for the direct product and call it a **direct power** of $\mathbf{B}$. In case $I = \{0,1\}$, we write $\mathbf{A}_0 \times \mathbf{A}_1$ in place of $\prod \mathbf{A}$.

Throughout this book we have frequent need to write expressions like

$$Q^{\mathbf{A}}(a_0, a_1, \cdots, a_{r-1})$$

where Q is an operation symbol (or a more complicated expression) of the algebra $\mathbf{A}$ of rank r and $a_0, a_1, \ldots, a_{r-1} \in A$. Very often the exact rank of Q is of little significance and the expression above is needlessly complex. We replace it by

$$Q^{\mathbf{A}}(\bar{a})$$

where $\bar{a}$ stands for the tuple of elements of A of the correct length.

The formation of homomorphic images, of subalgebras, and of direct products are principal tools we will use to manipulate algebras. Frequently, these tools are used in conjunction with each other. For example, let $\mathbb{R}$ denote the ring of real numbers and let I denote the unit interval. Then $\mathbb{R}^I$ is the ring of all real-valued functions on the unit interval. Going a step further, we can obtain the ring of all continuous real-valued functions on the unit interval as a subalgebra of $\mathbb{R}^I$.

Let $\mathcal{K}$ be a class of similar algebras. We use the following notation:

$\mathbf{H}(\mathcal{K})$ is the class of all homomorphic images of members of $\mathcal{K}$.

$\mathbf{S}(\mathcal{K})$ is the class of all isomorphic images of subalgebras of members of $\mathcal{K}$.

$\mathbf{P}(\mathcal{K})$ is the class of all isomorphic images of direct products of systems of algebras belonging to $\mathcal{K}$.

We say that $\mathcal{K}$ is **closed** under the formation of homomorphic images, under the formation of subalgebras, and under the formation of direct products–provided, respectively, that $\mathbf{H}(\mathcal{K}) \subseteq \mathcal{K}$, $\mathbf{S}(\mathcal{K}) \subseteq \mathcal{K}$, and $\mathbf{P}(\mathcal{K}) \subseteq \mathcal{K}$. Observe that if $\mathcal{K}$ is closed with respect to direct products, then $\mathcal{K}$ contains all the one-element algebras of the similarity type, since, in particular, $\mathcal{K}$ must contain the direct product of the empty system of algebras.

Let $\mathcal{K}$ be a class of similar algebras. We call $\mathcal{K}$ a **variety** if and only if $\mathcal{K}$ is closed under the formation of homomorphic images, of subalgebras, and of direct products (i.e., $\mathbf{H}(\mathcal{K}) \subseteq \mathcal{K}$, $\mathbf{S}(\mathcal{K}) \subseteq \mathcal{K}$, and $\mathbf{P}(\mathcal{K}) \subseteq \mathcal{K}$). All of the classes described at the close of the last section are varieties. Varieties offer us a means

to classify algebras (that is, to organize them into classes) that is compatible with our chief means for manipulating algebras. The notion of a variety will become one of the central themes of these volumes.

Exercises 1.6

1. Let $\mathbf{A}$ and $\mathbf{B}$ be algebras. Prove that

$$\hom(\mathbf{A},\mathbf{B}) = (\mathrm{Sub}\,\mathbf{B} \times \mathbf{A}) \cap \{h : h \text{ is a function from } A \text{ into } B\}.$$

2. Let $\mathbf{A} = \langle \mathbf{A}_i : i \in I \rangle$ be a system of similar algebras. Prove that p_i is a homomorphism from $\prod \mathbf{A}$ onto $\mathbf{A}_i$ for each $i \in I$.

3. Let $\mathbf{A} = \langle \mathbf{A}_i : i \in I \rangle$ be a system of similar algebras and assume that $\mathbf{B}$ is an algebra of the same type with $B = \prod A$. Prove that if p_i is a homomorphism from $\mathbf{B}$ onto $\mathbf{A}_i$ for each $i \in I$, then $\mathbf{B} = \prod \mathbf{A}$.

4. Let $\mathbf{A} = \langle \mathbf{A}_i : i \in I \rangle$ be a system of similar algebras. Let $\mathbf{B}$ be an algebra of the same type and h_i be a homomorphism from $\mathbf{B}$ into $\mathbf{A}_i$ for each $i \in I$. Prove that there is a homomorphism g from $\mathbf{B}$ into $\prod \mathbf{A}$ such that $h_i = p_i \circ g$ for each $i \in I$.

5. Prove that every semigroup is isomorphic to a semigroup of functions from X into X, where the operation is composition of functions and X is some set.

6. Prove that every ring is isomorphic to a ring of endomorphisms of some Abelian group. [Hint: Let $\mathbf{A}$ be an Abelian group. The sum h of endomorphisms f and g of $\mathbf{A}$ is defined so that that

$$h(a) = f(a) + g(a) \text{ for all } a \in A,$$

where $+$ is the basic binary operation of $\mathbf{A}$. The product of a pair of endomorphisms is their composition.]

7. Describe all the three-element homomorphic images of $\langle \omega, + \rangle$, where ω is the set of natural numbers.

1.3 Generation of Subalgebras

Let $\mathbf{a}$ be an algebra and let X be an arbitrary subset of the universe A of $\mathbf{A}$. X is unlikely to be a subuniverse of $\mathbf{A}$, since quite possibly there is a basic operation that, when applied to certain elements of X, produces a value outside X. So X may fail to be a subuniverse because it lacks certain elements. As a first step toward extending X to a subuniverse, one might gather into a set Y all of

those elements that result from applying the operations to the elements of X. Then $X \cup Y$ is no longer deficient in the way X was. The new elements from Y, however, may now be taken as arguments for the basic operations, and $X \cup Y$ may not be closed under all these operations. But by repeating this process, perhaps countably infinitely many times, X can be extended to a subuniverse of $\mathbf{A}$. With respect to the subset relation, this subuniverse will be the smallest subuniverse of $\mathbf{A}$ that includes X, since only those elements required by the closure conditions in the definition of subuniverse are introduced in the process. The subuniverse obtained in this way must be included in every subuniverse of which X is a subset. Thus it may be obtained as the intersection of all such subuniverses. The finitary character of the fundamental operations of the algebra ensures that this iterative process succeeds after only countably many steps.

THEOREM 1.7. *Let* $\mathbf{A}$ *be an algebra and let* S *be any nonempty collection of subuniverses of* $\mathbf{A}$. *Then* $\bigcap S$ *is a subuniverse of* $\mathbf{A}$.

Proof. Evidently $\bigcap S$ is a subset of A. Let F be any basic operation of $\mathbf{A}$ and suppose that r is the rank of F. To see that $\bigcap S$ is closed under F, pick any $a_0, a_1, \ldots, a_{r-1} \in \bigcap S$. For all $B \in S$ we know that $a_0, a_1, \ldots, a_{r-1} \in B$; but then $F(a_0, a_1, \cdots, a_{r-1}) \in B$, since B is a subuniverse. Therefore $F(a_0, a_1, \cdots, a_{r-1}) \in \bigcap S$ and $\bigcap S$ is closed under F. ■

DEFINITION 1.8. Let $\mathbf{A}$ be an algebra and let $X \subseteq A$. The **subuniverse of A generated by** X is the set $\bigcap\{B : X \subseteq B$ and B is a subuniverse of $\mathbf{A}\}$. $\mathrm{Sg}^{\mathbf{A}}(X)$ denotes the subuniverse of $\mathbf{A}$ generated by X.

Since $X \subseteq A$ and A is a subuniverse of $\mathbf{A}$, Theorem 1.7 justifies calling $\mathrm{Sg}^{\mathbf{A}}(X)$ a subuniverse of $\mathbf{A}$.

Now we can formalize the discussion that opened this section.

THEOREM 1.9. *Let* $\mathbf{A}$ *be an algebra and* $X \subseteq A$. *Define* X_n *by the following recursion:*

$$X_0 = X$$
$$X_{n+1} = X_n \cup \{F(\bar{a}) : F \text{ is a basic operation of } \mathbf{A} \text{ and } \bar{a} \text{ is a tuple from } X_n\}.$$

Then $\mathrm{Sg}^{\mathbf{A}}(X) = \bigcup\{X_n : n \in \omega\}$.

Proof. The proof consists of two claims.

CLAIM 1. $\mathrm{Sg}^{\mathbf{A}}(X) \subseteq \bigcup\{X_n : n \in \omega\}$.

Since $X \subseteq \bigcup\{X_n : n \in \omega\}$, we need only show that $\bigcup\{X_n : n \in \omega\}$ is a subuniverse. Let F be a basic operation and let $\bar{a}$ be a tuple from $\bigcap\{X_n : n \in \omega\}$. Since F has some finite rank and $X_0 \subseteq X_1 \subseteq \cdots$, we can easily see that $\bar{a}$ is a tuple from X_m for some large enough m. But then $F(\bar{a}) \in X_{m+1} \subseteq \bigcap\{X_n : n \in \omega\}$. So this latter set is a subuniverse, as desired.

CLAIM 2. $\bigcup\{X_n : n \in \omega\} \subseteq \mathrm{Sg}^{\mathbf{A}}(X)$.

Since $\mathrm{Sg}^{\mathbf{A}}(X)$ is the intersection of all subuniverses that include X, it suffices to show that $X_n \subseteq B$ for every subuniverse that includes X and for every natural number n. This can be immediately accomplished by induction on n. ■

COROLLARY 1.10. *Let* $\mathbf{A}$ *be an algebra and* $X \subseteq \mathbf{A}$. *If* $a \in \mathrm{Sg}^{\mathbf{A}}(X)$, *then there is a finite set* Y *such that* $Y \subseteq X$ *and* $a \in \mathrm{Sg}^{\mathbf{A}}(Y)$.

Proof. We will prove by induction on n that

$$\text{If } a \in X_n, \text{ then } a \in \mathrm{Sg}^{\mathbf{A}}(Y) \text{ for some finite } Y \subseteq X. \qquad (\star)$$

Initial step: $n = 0$. Take $Y = \{a\}$.

Inductive Step: $n = m + 1$, and we assume without loss of generality that $a = F(\bar{b})$ where F is a basic operation and $\bar{b}$ is a tuple from X_m. Letting Y be the union of the finite sets obtained by applying the inductive hypothesis to each element of $\bar{b}$, we see that $\bar{b}$ is a tuple from $\mathrm{Sg}^{\mathbf{A}}(Y)$. Thus $a \in \mathrm{Sg}^{\mathbf{A}}(Y)$ as desired. ■

COROLLARY 1.11. *Let* $\mathbf{A}$ *be an algebra and* $X, Y \subseteq \mathbf{A}$. *Then*

i. $X \subseteq \mathrm{Sg}^{\mathbf{A}}(X)$.

ii. $\mathrm{Sg}^{\mathbf{A}}\left(\mathrm{Sg}^{\mathbf{A}}(X)\right) = \mathrm{Sg}^{\mathbf{A}}(X)$.

iii. *If* $X \subseteq Y$, *then* $\mathrm{Sg}^{\mathbf{A}}(X) \subseteq \mathrm{Sg}^{\mathbf{A}}(Y)$.

iv. $\mathrm{Sg}^{\mathbf{A}}(X) = \bigcup\{\mathrm{Sg}^{\mathbf{A}}(Z) : Z \subseteq X \text{ and } Z \text{ is finite}\}$.

The properties of $\mathrm{Sg}^{\mathbf{A}}$, considered as a unary operation on the power set of A, which have been gathered in this last corollary, are so frequently used that usually no reference will be given. Subuniverses of the form $\mathrm{Sg}^{\mathbf{A}}(Z)$, where Z is finite, are said to be **finitely generated**. Part 4 of this corollary entails that the universe of any algebra is the union of its finitely generated subuniverses.

The set-inclusion relation is a partial order on the collection of all subuniverses of $\mathbf{A}$. This order induces lattice operations of join and meet on the collection of all subuniverses. Some of the fundamental facts concerning this order are easily deduced. They have been gathered in the next corollary.

COROLLARY 1.12. *Let* $\mathbf{A}$ *be any algebra,* S *be any nonempty collection of subuniverses of* $\mathbf{A}$, *and* B *be any subuniverse of* $\mathbf{A}$. *Then*

i. *With respect to set-inclusion,* $\bigcap S$ *is the largest subuniverse included in each member of* S.

ii. *With respect to set-inclusion,* $\mathrm{Sg}^{\mathbf{A}}(\bigcup S)$ *is the smallest subuniverse including each member of* S.

iii. *B is finitely generated if and only if whenever* $B \subseteq \mathrm{Sg}^{\mathbf{A}}(\bigcup T)$ *for any set T of subuniverses of* $\mathbf{A}$*, then* $B \subseteq \mathrm{Sg}^{\mathbf{A}}(\bigcup T')$ *for some finite* $T' \subseteq T$.

iv. *Suppose that for all* $B, C \in S$ *there is* $D \in S$ *such that* $B \cup C \subseteq D$*. Then* $\bigcup S$ *is a subuniverse of* $\mathbf{A}$.

Parts 1 and 2 describe, respectively, the meet and join in the lattice of subuniverses. In fact, they describe how to form meets and joins of arbitrary collections of subuniverses rather than just the meets and joins of subuniverses two at a time. The import of 3 is that the notion of finite generation, which on its face appears to be something "internal" to a subuniverse, can be characterized in terms of the order-theoretic properties of the set of all subuniverses. This last corollary can be deduced from the preceding material with help of only the following fact:

> The subuniverses of $\mathbf{A}$ are precisely those subsets X of A such that $X = \mathrm{Sg}^{\mathbf{A}}(X)$.

Suppose B and C are any two subuniverses of $\mathbf{A}$. We define the join of B and C (denoted $B \vee C$) by

$$B \vee C = \mathrm{Sg}^{\mathbf{A}}(B \cup C)$$

and the meet of B and C (denoted $B \wedge C$) by

$$B \wedge C = B \cap C.$$

It is not hard to prove, using the last corollary, that the collection of all of $\mathbf{A}$ endowed with these two operations is a lattice. We call this lattice the **lattice of subuniverses** of $\mathbf{A}$ and denote it by $\mathbf{Sub}\,\mathbf{A}$.

Exercises 1.13

1. Prove that every subuniverse of $\langle \omega, + \rangle$ is finitely generated, where ω is the set $\{0, 1, 2, \ldots\}$ of natural numbers.

2. Supply proofs for Corollaries 1.11 and 1.12.

3. A collection C of sets is said to be **directed** iff for all $B, D \in C$ there is $E \in C$ such that $B \subseteq E$ and $D \subseteq E$. C is called a **chain** of sets provided $\subseteq$ is a linear ordering of C. Let $\mathbf{A}$ be an algebra. Prove that the following statements are equivalent:

i. B is a finitely generated subuniverse of $\mathbf{A}$.

ii. If C is any nonempty directed collection of subuniverses of $\mathbf{A}$ and $B \subseteq \bigcup C$, then there is $D \in C$ such that $B \subseteq D$.

iii. If C is any nonempty chain of subuniverses of $\mathbf{A}$ and $B \subseteq \bigcup C$, then there is $D \in C$ such that $B \subseteq D$.

iv. If C is any nonempty chain of subuniverses of $\mathbf{A}$ and $B = \bigcup C$, then $B \in C$.

(HINT: It may help to prove first that every infinite set M is the union of a chain of its subsets, each of which has cardinality less than the cardinality $|M|$ of M. Zorn's Lemma or some other variant of the Axiom of Choice would be useful at this point.)

4. An algebra $\mathbf{A}$ is called **mono-unary** if it has only one basic operation and that operation is unary. Prove that any infinite mono-unary algebra has a proper subalgebra.

1.4 Congruence Relations and Quotient Algebras

Unlike the formation of subalgebras, the formation of homomorphic images of an algebra apparently involves external considerations. But there is a notion of quotient algebra that captures all homomorphic images, at least up to isomorphism. The constructions using normal subgroups and ideals familiar from the theories of groups and rings provide a clue as to how to proceed in the general setting.

Let h be a homomorphism from $\mathbf{A}$ onto $\mathbf{B}$. Define

$$\theta = \left\{ \langle a, a' \rangle : a, a' \in A \text{ and } h(a) = h(a') \right\}.$$

So θ is a binary relation on the universe of $\mathbf{A}$. It is convenient to write $a\ \theta\ a'$ in place of $\langle a, a' \rangle \in \theta$. Now θ is easily seen to be an equivalence relation on A, since h is a function with domain A. Because h is a homomorphism, θ has an additional property called the **substitution property** for $\mathbf{A}$:

> Let F be any basic operation of $\mathbf{A}$ and let $a_0, b_0, a_1, b_1, \ldots \in A$. If $a_i\ \theta\ b_i$ for all i less than the rank of F, then $F(a_0, a_1, \cdots)\ \theta\ F(b_0, b_1 \cdots)$.

We call the relation θ just described the **kernel** of h and denote it by $\ker h$.

DEFINITION 1.14. Let $\mathbf{A}$ be an algebra. By a **congruence relation on A** we mean an equivalence relation on the universe of $\mathbf{A}$ that has the substitution property for $\mathbf{A}$. Con $\mathbf{A}$ denotes the set of all congruence relations on $\mathbf{A}$.

The kernels of homomorphisms are always congruence relations.

Now congruence relations on $\mathbf{A}$, being special kinds of equivalence relations, induce partitions of A. Let θ be a congruence relation on the algebra $\mathbf{A}$. We use the following notation:

$$a/\theta = \{b : a\ \theta\ b \text{ and } b \in A\} \quad \text{for all } a \in A$$
$$A/\theta = \{a/\theta : a \in A\}.$$

For $a \in A$, the set a/θ is called the **congruence class** of a modulo θ. A/θ is the partition of A into congruence classes modulo θ. There is a natural map g, called the **quotient map**, from A onto A/θ defined by

$$g(a) = a/\theta \quad \text{for all } a \in A.$$

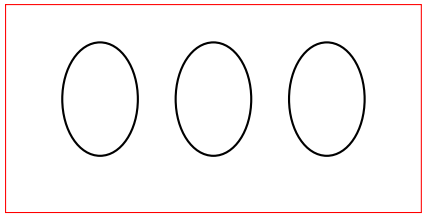


Figure 1.1:

We can impose the basic operations on A/θ in such a way that the quotient map becomes a homomorphism. Let F be a basic operation of A and let r be its rank. We will define an operation F_θ on A/θ that will correspond to F. The following condition is crucial if g is to be a homomorphism:

$$g(F(a_0,a_1,\cdots,a_{r-1})) = F_\theta(g(a_0),g(a_1),\cdots,g(a_{r-1}))$$

for all $a_0,a_1,\ldots \in A$. This looks very much like an adequate definition of F_θ, except that the elements $a_0,a_1,\ldots$ seem to play a rather privileged role on the left-hand side, whereas $g(a_0) = a_0/\theta$, $g(a_1) = a_1/\theta,\ldots$ of the right-hand side are congruence classes that are represented, as it were, accidentally by $a_0,a_1,\ldots$. But the substitution property is exactly the statement that any other choice of representatives would lead to the same value of either sided. Thus the equation displayed above can be used as a definition of an operation on A/θ.

DEFINITION 1.15. Let $\mathbf{A}$ be an algebra and θ be a congruence relation on $\mathbf{A}$. The **quotient algebra** $\mathbf{A}/\theta$ is the algebra similar to $\mathbf{A}$ with universe A/θ in which $Q_\theta^{\mathbf{A}}$ is the interpretation of Q, for each operation symbol Q.

Since the congruence θ is obviously the kernel of the quotient map from $\mathbf{A}$ onto $\mathbf{A}/\theta$, we conclude that the congruence relations on $\mathbf{A}$ are exactly the kernels of the homomorphisms with domain A. This begs the question of the connection between $\mathbf{B}$ and $\mathbf{A}/\theta$ where $\theta = \ker h$ and $h : \mathbf{A} \twoheadrightarrow \mathbf{B}$. The answer is contained in the next theorem, where we record the result of this discussion as well.

THEOREM 1.16 (THE HOMOMORPHISM THEOREM)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be similar algebras, let* h *be a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{B}$*, and let* θ *be a congruence relation on* $\mathbf{A}$ *and* g *be the quotient map from* $\mathbf{A}$ *onto* $\mathbf{A}/\theta$*. Then*

- ***i.*** *The kernel of* h *is a congruence relation on* $\mathbf{A}$*.*
- ***ii.*** *The quotient map* $g : \mathbf{A} \twoheadrightarrow \mathbf{A}/\theta$ *is a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{A}/\theta$*.*
- ***iii.*** *If* $\theta = \ker h$*, then the unique function* f *from* A/θ *onto* B *satisfying* $f \circ g = h$ *is an isomorphism from* $\mathbf{A}/\theta$ *onto* $\mathbf{B}$*.*

Proof. As the various definitions were virtually designed to make 1 and 2 true, we will only look at 3. Since we want $f \circ g = h$ and since g is the quotient map, the only option is to define f by

$$f(a/\theta) = h(a) \quad \text{for all } a \in A.$$

To see that this definition is sound, suppose that $a\ \theta\ a'$. Then $h(a) = h(a')$, since $\theta = \ker h$. Thus $f(a/\theta) = h(a) = h(a') = f(a'/\theta)$, as desired. f is one-to-one, since

$$f(a/\theta) = f(a'/\theta) \text{ implies } h(a) = h(a')$$

and, as $\ker h = \theta$, we have $a/\theta = a'/\theta$. Finally, to demonstrate that f is a homomorphism, let Q be an operation symbol and $\bar{a}$ be a tuple from A. Then

$$\begin{aligned} f\left(Q^{\mathbf{A}}_{\theta}(\bar{a}/\theta)\right) &= f\left(Q^{\mathbf{A}}_{\theta}(g(\bar{a}))\right) \\ &= f\left(g\left(Q^{\mathbf{A}}(\bar{a})\right)\right) \\ &= h\left(Q^{\mathbf{A}}(\bar{a})\right) \\ &= Q^{\mathbf{B}}\left(h(\bar{a})\right) \\ &= Q^{\mathbf{B}}\left(f(g(\bar{a}))\right) \\ &= Q^{\mathbf{B}}\left(f(\bar{a}/\theta)\right). \end{aligned}$$

Thus, f is a homomorphism and hence an isomorphism from $\mathbf{A}/\theta$ onto $\mathbf{B}$. ■

This theorem, specialized to groups and rings, becomes a familiar result. In these settings, it is possible to go a step further and distinguish normal subgroups, in the case of groups, and ideals, in the case of rings. These are, respectively, the congruence classes containing the unit element of the group and the zero element of the ring. For these kinds of algebras, it can be argued that these special congruence classes determine the whole congruence relation. In fact, for groups and all their expansions, any single congruence class determines the whole congruence relation. This is a rather unusual property for an algebra to have, as some of the examples below illustrate. This property and others like it are examined in some detail in Volume 2.

EXAMPLE 1.17. 1. Let $\mathbb{Z} = \langle \mathbb{Z}, +, \cdot, -, 0, 1 \rangle$ denote the ring (with unit) of integers. For each integer q, define the equivalence relation $\equiv_q$ on $\mathbb{Z}$ by

$$r \equiv_q s \text{ iff } q \text{ is a factor of } r - s.$$

It is easy to see that $\equiv_0$ is the identity relation whereas $r \equiv_1 s$ holds for all integers r and s. Since $\equiv_{-q}$ is the same relation as $\equiv_q$, we ignore negative q's. In view of the distributive law, we conclude that $\equiv_q$ is a congruence relation on the ring of integers. Of all the $\equiv_q$'s, we see that $\equiv_0$ is the smallest, that $\equiv_1$ is the largest, and that

$$\equiv_q \subseteq \equiv_t \text{ iff } t \text{ is a factor of } q.$$

The ring $\mathbb{Z}/\equiv_0$ is isomorphic to $\mathbb{Z}$ and the ring $\mathbb{Z}/\equiv_1$ has only one element. For $q > 1$, $\mathbb{Z}/\equiv_q$ is the ring of "residues modulo q." It is isomorphic to the ring with universe $\{0,1,\ldots,q-1\}$ where addition and multiplication are defined modulo q; that is, the operation is performed as in $\mathbb{Z}$, and then the remainder modulo q is extracted to obtain the value in $\{0,1,\ldots,q-1\}$.

Now let θ be an arbitrary congruence relation on $\mathbb{Z}$. We can show that θ is $\equiv_q$ for some q. Consider the case when $0/\theta = \{0\}$. In this case we have

$$\begin{aligned} r\ \theta\ s &\iff (r-s)\ \theta\ (s-s) = 0 \\ &\iff (r-s) = 0 \\ &\iff r = s \end{aligned}$$

and so θ is $\equiv_0$. In the case that $0/\theta$ is properly larger than $\{0\}$, it is easy to prove that $0/\theta$ has a positive number; let q be the least positive number of $0/\theta$. Let $r,s \in \mathbb{Z}$ and pick $d,t \in \mathbb{Z}$ so that

$$r - s = qd + t \text{ and } t \in \{0,1,\ldots,q-1\}.$$

Then

$$\begin{aligned} r\ \theta\ s &\iff (r-s)\ \theta\ 0 \\ &\iff (qd+t)\ \theta\ 0 \\ &\iff t\ \theta\ 0 (\text{since } q\ \theta\ 0 \text{ implies } qd\ \theta\ 0) \\ &\iff t = 0 (\text{ since } 0 \le t < q) \\ &\iff q \text{ is a factor of } r-s \\ &\iff r \equiv_q s. \end{aligned}$$

So the congruence relations on the ring of integers are exactly the relations $\equiv_q$ where q is a non-negative integer. The homomorphic images of $\mathbb{Z}$ are, up to isomorphism, $\mathbb{Z}$ itself, the one element ring, and the ring of residues modulo integers greater than 1.

2. Let $\mathbb{R} = \langle \mathbb{R}, \wedge, \vee \rangle$ where $\mathbb{R}$ is the set of real numbers and

$$\begin{aligned} r \wedge s &= \min(r,s) \\ r \vee s &= \max(r,s) \end{aligned}$$

$\mathbb{R}$ is a lattice. We wish to describe all the congruence relations on $\mathbb{R}$. So let θ be a congruence relation. Suppose $r\ \theta\ s$ and $r \le t \le s$. Then $r = (r \wedge t)\ \theta\ (s \wedge t) = t$, and so $r\ \theta\ t$. This means that the congruence classes of θ are convex–that is, they are intervals, perhaps infinite or even degenerate. Pick an arbitrary element from each congruence class. It is evident that the set of selected elements forms a subalgebra of $\mathbb{R}$ isomorphic to $\mathbb{R}/\theta$.

Now let θ be any equivalence relation on $\mathbb{R}$ such that each θ-equivalence class is a convex set of real numbers. To verify the substitution property, let $r\ \theta\ r'$ and $s\ \theta\ s'$.

CASE 0: $r\ \theta\ s$.

The substitution property is immediate, since $r \wedge s$, $r \vee s$, $r' \wedge s'$, and $r' \vee s'$ all belong to $\{r, s, r', s'\} \subseteq r/\theta$.

CASE 1: r and s lie in different equivalence classes modulo θ.

The two θ-classes, which are intervals, cannot overlap. Thus, without loss of generality, we assume that every element of r/θ is less than every element of s/θ. Hence

$$\begin{aligned} r \wedge s &= r \\ r' \wedge s' &= r' \\ r \vee s &= s \\ r' \vee s' &= s'. \end{aligned}$$

Therefore $(r \wedge s)\ \theta\ (r' \wedge s')$ and $(r \vee s)\ \theta\ (r' \vee s')$, as desired, and so θ is a congruence relation.

Thus the congruence relations of $\mathbb{R}$ are exactly those equivalence relations whose equivalence classes are convex sets of real numbers. Since any proper convex subset of $\mathbb{R}$ is a congruence class of infinitely many congruence relations, $\mathbb{R}$ provides a strong contrast with the behavior of congruence relations on groups.

For most algebras, the task of describing all the congruence relations is hopelessly difficult, so these two examples have a rather special character. The collection of all congruence relations of an algebra is a rich source of information about the algebra; discovering the properties of this collection often leads to a deeper understanding of the algebra.

Just as the subuniverses of an algebra form a lattice, so do the congruence relations. Roughly the same analysis can be used. Let $\mathbf{A}$ be an algebra and let X be a binary relation on A. Now X may fail to be a congruence relation, either because it is not an equivalence relation on A or because it does not have the substitution property. In either case, the failure can be traced to the existence of an ordered pair that fails to belong to X but that must belong to any congruence relation that includes X. All such necessary ordered pairs can be gathered into a set Y, and $X \cup Y$ is at least not subject to the same deficiencies as X. Yet $X \cup Y$ may fail transitivity or the substitution property. But by repeating the process, perhaps countably infinitely often, a congruence relation will be built; with respect to set inclusion, it will be the smallest congruence relation on $\mathbf{A}$ that includes X.

THEOREM 1.18. *Let* $\mathbf{A}$ *be an algebra and* C *be any nonempty collection of congruence relations on* $\mathbf{A}$. *Then* $\bigcap C$ *is a congruence relation on* $\mathbf{A}$.

The routine proof of this theorem is left as an exercise. This theorem allows us to proceed as we did with subuniverses.

DEFINITION 1.19. Let $\mathbf{A}$ be an algebra and let $X \subseteq A \times A$. The **congruence relation on A generated by** X is the set

$$\bigcap\{\theta : X \subseteq \theta \text{ and } \theta \text{ is a congruence relation on } A\}.$$

$\mathrm{Cg}^{\mathbf{A}}(X)$ denotes the congruence relation on $\mathbf{A}$ generated by X.

As we did with subuniverses, we can formalize the discussion above to obtain a description of how to extend a binary relation to obtain the congruence relation it generates. This is complicated by the necessity of arriving at an equivalence relation. For the purposes of convenience in dealing with congruences, both here and in general, we introduce some notation. Let $\mathbf{A}$ be an algebra and θ be a binary relation on A. Let $\bar{a}$ and $\bar{a}'$ be tuples from A of the same length, say r. So

$$\bar{a} = \langle a_0, a_1, \ldots, a_{r-1}\rangle \text{ and } \bar{a}' = \langle a'_0, a'_1, \ldots, a'_{r-1}\rangle.$$

We use

$$\bar{a}\ \theta\ \bar{a}'$$

in place of the more elaborate

$$\begin{array}{ccc} a_0 & \theta & a'_0 \\ a_1 & \theta & a'_1 \\ & \vdots & \\ a_{r-1} & \theta & a'_{r-1}. \end{array}$$

Thus $\bar{a}\ \theta\ \bar{a}'$ stands for "$a_i\ \theta\ a'_i$ for all $i < r$." Using this notation, we can rephrase the substitution property as:

> $\bar{a}\ \theta\ \bar{a}'$ implies $F(\bar{a})\ \theta\ F(\bar{a}')$, for all basic operations F and all tuples $\bar{a}$ and $\bar{a}'$.

THEOREM 1.20. *Let* $\mathbf{A}$ *be an algebra and* $X \subseteq A \times A$. *Define* X_n *by the following recursion:*

$$\begin{aligned} X_0 &= X \cup \big\{\langle a, a'\rangle : \langle a', a\rangle \in X\big\} \cup \{\langle a, a\rangle : a \in A\} \\ X_{n+1} &= X_n \cup T_n \cup Q_n, \end{aligned}$$

where $Q_n = \{\langle F(\bar{a}), F(\bar{a}')\rangle : F$ *is a basic operation and* $\bar{a}$ *and* $\bar{a}'$ *are tuples such that* $\bar{a}\ X_n\ \bar{a}'\}$ *and* $T_n = \{\langle a, c\rangle : a\ X_n\ b\ X_n\ c$ *for some* $b \in A\}$.

Then $\mathrm{Cg}^{\mathbf{A}}(X) = \bigcup_{n\in\omega} X_n$.

The proof of this theorem is much like the proof of Theorem 1.9, the only new element being an easy argument about transitive closures.

COROLLARY 1.21. *Let* $\mathbf{A}$ *be an algebra and* $X \subseteq A \times A$. *If* $\langle a, a'\rangle \in \mathrm{Cg}^{\mathbf{A}}(X)$, *then there is a finite set* $Y \subseteq X$ *such that* $\langle a, a'\rangle \in \mathrm{Cg}^{\mathbf{A}}(Y)$.

COROLLARY 1.22. *Let* $\mathbf{A}$ *be an algebra and* $X, Y \in A \times A$. *Then*

i. $X \subseteq \mathrm{Cg}^{\mathbf{A}}(X)$.

ii. $\mathrm{Cg}^{\mathbf{A}}\left(\mathrm{Cg}^{\mathbf{A}}(X)\right) = \mathrm{Cg}^{\mathbf{A}}(X)$.

iii. *If* $X \subseteq Y$, *then* $\mathrm{Cg}^{\mathbf{A}}(X) \subseteq \mathrm{Cg}^{\mathbf{A}}(Y)$.

iv. $\mathrm{Cg}^{\mathbf{A}}(X) = \bigcup\{\mathrm{Cg}^{\mathbf{A}}(Z) : Z \subseteq X \text{ and } Z \text{ is finite}\}$.

Congruence relations of the form $\mathrm{Cg}^{\mathbf{A}}(Z)$, where Z is finite, are said to be **finitely generated**. Those of the form $\mathrm{Cg}^{\mathbf{A}}(\{\langle a, a'\rangle\})$ are called **principal congruence relations**. This last piece of notation is obviously awkward. We replace it by $\mathrm{Cg}^{\mathbf{A}}(a, a')$. Evidentally, $\mathrm{Cg}^{\mathbf{A}}(a, a')$ is the smallest congruence relation that places a and a' in the same congruence class–or, as we shall say more suggestively, $\mathrm{Cg}^{\mathbf{A}}(a, a')$ is the smallest congruence collapsing $\langle a, a'\rangle$.

COROLLARY 1.23. *Let* $\mathbf{A}$ *be an algebra,* C *be a nonempty collection of congruence relations on* $\mathbf{A}$, *and* θ *be any congruence relation on* $\mathbf{A}$. *Then*

i. *With respect to set-inclusion,* $\bigcap C$ *is the largest congruence relation on* $\mathbf{A}$ *included in each member of* C.

ii. *With respect to set-inclusion,* $\mathrm{Cg}^{\mathbf{A}}(\bigcup C)$ *is the smallest congruence relation including each member of* C.

iii. θ *is finitely generated if and only if whenever* D *is a set of congruences on* $\mathbf{A}$ *and* $\theta \subseteq \mathrm{Cg}^{\mathbf{A}}(\bigcup D)$, *then* $\theta \subseteq \mathrm{Cg}^{\mathbf{A}}(\bigcup D')$ *for some finite* $D' \subseteq D$.

iv. *Suppose that for each* $\phi, \psi \in C$ *there is* $\eta \in C$ *such that* $\phi \cup \psi \subseteq \eta$. *Then* $\bigcup C$ *is a congruence relation on* $\mathbf{A}$.

The proofs of these three corollaries do not differ significantly from the proofs of the three corollaries to Theorem 1.9.

Suppose ϕ and ψ are congruence relations on the algebra $\mathbf{A}$. We can define the **join** (designated by $\vee$) and the **meet** (designated by $\wedge$) of ϕ and ψ so that $\phi \vee \psi = \mathrm{Cg}^{\mathbf{A}}(\phi \cup \psi)$ and $\phi \wedge \psi = \phi \cap \psi$. With these operations, the collection of all congruence relations on $\mathbf{A}$ becomes a lattice, which we shall call the **congruence lattice** of $\mathbf{A}$ and denote by **Con A**. Every congruence relation of $\mathbf{A}$ is an equivalence relation on A and, as we saw in Exercise 1.2(7), the equivalence relations on A constitute a lattice $\mathbf{Eqv}A$. In fact, **Con A** is a sublattice of $\mathbf{Eqv}A$. The meet operation in both **Con A** and $\mathbf{Eqv}A$ is just set-theoretic intersection. To establish the contention that **Con A** is a sublattice of $\mathbf{Eqv}A$, it is necessary to prove that the join operation in **Con A** is the restriction of the join operation in $\mathbf{Eqv}A$. This is the import of the next theorem, and it even applies to joins of arbitrary subsets of Con $\mathbf{A}$. Note that we write Con $\mathbf{A}$ for the set of all congruence relations on $\mathbf{A}$–the universe of the lattice **Con A**. By the same token, we write Sub $\mathbf{A}$ for the set of all subuniverses of $\mathbf{A}$, and $\mathbf{Eqv}A$ for the set of all equivalence relations over the set A.

THEOREM 1.24. *Let* $\mathbf{A}$ *be an algebra and let* $C \subseteq \operatorname{Con} \mathbf{A}$.

i. $\operatorname{Con} \mathbf{A} = (\operatorname{Sub} \mathbf{A} \times \mathbf{A}) \cap \operatorname{Eqv} A$.

ii. $\operatorname{Cg}^{\mathbf{A}}(\bigcup C) = \bigcap \{R : \bigcup C \subseteq R \text{ and } R \in \operatorname{Eqv} A\}$.

Proof. **i.** This is just a restatement of the definition of a congruence relation, using the language of subuniverses and direct products.

ii. Let $\theta = \bigcap \{R : \bigcup C \subseteq R \text{ and } R \in \operatorname{Eqv} A\}$. Thus θ is the smallest equivalence relation on A that includes $\bigcup C$. Since it is clear that $\bigcup C \subseteq \theta \subseteq \operatorname{Cg}^{\mathbf{A}}(\bigcup C)$, we need only establish that θ is a congruence on A. In view of part 1 of the theorem, it remains to establish that θ is a subuniverse of $\mathbf{A} \times \mathbf{A}$.

The transitive closure of relations was described in the Preliminaries. A more detailed analysis is used here. Let D be any collection of relations on A–that is, subsets of $A \times A$. Let D^* denote the set of all those relations that can be obtained as compositions (i.e., relational products) of finite nonempty sequences of relations belonging to D. Thus $\phi_0 \circ \phi_1 \circ \cdots \circ \phi_{n-1}$, where $\phi_i \in D$ for each $i < n$ is a typical element of D^*. Checking that $\bigcup D^*$ is actually the transitive closure of $\bigcup D$ presents no difficulties. This analysis of the transitive closure leads immediately to the following conclusion: If D consists entirely of symmetric reflexive relations on A, then the transitive closure $\bigcup D$ is also symmetric and reflexive and is, therefore, the smallest equivalence relation including $\bigcup D$. In particular, θ is the transitive closure $\bigcup C^*$ of $\bigcup C$.

Now consider any two subuniverses ϕ and ψ of $\mathbf{A} \times \mathbf{A}$. $\phi \circ \psi$ must be a subuniverse of $\mathbf{A} \times \mathbf{A}$ as well, since if F is any basic operation of $\mathbf{A}$ and n is the rank of F and if $a_i \phi b_i \psi c_i$ for all $i < n$, then

$$F(a_0, a_1, \cdots, a_{n-1})\, \phi F(b_0, b_1, \cdots, b_{n-1})\, \psi F(c_0, c_1, \cdots, c_{n-1}).$$

By the obvious inductive extension of this fact, if D consists entirely of subuniverses of $\mathbf{A} \times \mathbf{A}$, then so does D^*. Moreover, if every member of D is reflexive, then D^* is directed upward by set-inclusion (see Exercise 1.13(3)). In particular, this means that C^* is a collection of subuniverses of $\mathbf{A} \times \mathbf{A}$ and that for any ϕ and ψ in C^* there is η in C^* such that $\phi \cup \psi \subseteq \eta$. Hence, by Corollary 1.12 (4), $\bigcup C^*$ is a subuniverse of $\mathbf{A} \times \mathbf{A}$. That is, θ is a subuniverse of $\mathbf{A} \times \mathbf{A}$, as desired.

■

Let $\mathbf{A}$ be any algebra. With $\mathbf{A}$ we can associate four other algebras: $\mathbf{End}\,\mathbf{A}$, which is the monoid of all endomorphisms of $\mathbf{A}$; $\mathbf{Sub}\,\mathbf{A}$, which is the lattice of all subuniverses of $\mathbf{A}$; and $\mathbf{Con}\,\mathbf{A}$, which is the lattice of all congruence relations on $\mathbf{A}$. Chapter 2 is devoted to the rudiments of the abstract theory of lattices, and Chapter 3 takes up some aspects of the theories of monoids and of groups. These four algebras related to $\mathbf{A}$ contain significant information about $\mathbf{A}$.

Exercises 1.25

1. Let $h \in \hom(\mathbf{A}, \mathbf{B})$. Prove that $\ker h$ is a congruence relation on $\mathbf{A}$.

2. Let $\theta \in \text{Con } \mathbf{A}$. Prove that the quotient map from A onto A/θ is a homomorphism from $\mathbf{A}$ onto $\mathbf{A}/\theta$ and that its kernel is θ.

3. Let $\mathbf{G}$ be a group, $\theta \in \text{Con } \mathbf{G}$, and N be a normal subgroup of $\mathbf{G}$. Prove that e/θ is a normal subgroup of $\mathbf{G}$, where e is the unit of the group. Prove that $\{\langle a, b\rangle : a \cdot b^{-1} \in N\}$ is a congruence relation on $\mathbf{G}$. Finally, prove that if $\phi \in \text{Con } \mathbf{G}$, then $\phi = \theta$ iff $e/\phi = e/\theta$.

4. Verify that $\equiv_q$ is a congruence relation on the ring $\mathbb{Z}$ of integers for every natural number q.

5. Suppose $\theta \in \text{Con } \mathbf{A}$. Prove that $\theta = \bigcup \{\text{Cg}^{\mathbf{A}}(a, a') : a \,\theta\, a'\}$. Is every subuniverse the join of 1-generated subuniverses?

6. Let $\mathbf{A}$ be an algebra and $h \in \hom(\mathbf{A}, \mathbf{A})$. Prove that h is one-to-one iff $\ker h = 0_{\mathbf{A}}$, where $0_{\mathbf{A}}$ dentoes the least congruence relation on $\mathbf{A}$, namely the identity relation $\{\langle a, a\rangle : a \in A\}$.

7. Let $\mathbf{A}$ be an algebra. Define F to be the function with domain $\text{End}\,\mathbf{A}$ such that

$$F(h) = h^{-1} \circ h \text{ for all } h \in \text{End}\,\mathbf{A}.$$

Prove that F maps $\text{End}\,\mathbf{A}$ into $\text{Con } \mathbf{A}$.

***8.** (Burris and Kwatinetz) Let $\mathbf{A}$ be an algebra that is countable (i.e., finite or countably infinite) and has only countably many basic operations. Prove each of the following:

i. $|\text{Aut}\,\mathbf{A}| \leq \omega$ or $|\text{Aut}\,\mathbf{A}| = 2^{\omega}$

ii. $|\text{Sub}\,\mathbf{A}| \leq \omega$ or $|\text{Sub}\,\mathbf{A}| = 2^{\omega}$

iii. $|\text{End}\,\mathbf{A}| \leq \omega$ or $|\text{End}\,\mathbf{A}| = 2^{\omega}$

iv. $|\text{Con } \mathbf{A}| \leq \omega$ or $|\text{Con } \mathbf{A}| = 2^{\omega}$

where ω is the cardinality of the set of natural numbers and 2^{ω} is the cardinality of the set of real numbers.

C H A P T E R T W O

Lattices

2.1 Fundamental Concepts

Lattices arise often in algebraic investigations. We have already seen that **Sub A** and **Con A** are lattices for every algebra **A**. Knowing the significance of normal subgroups in group theory and ideals in ring theory, we should not be surprised that lattices of congruences have an important role to play. By itself, this would justify a detailed development of lattice theory. It turns out that, in addition to congruence lattices, many other lattices prove useful in developing a general theory of algebras.

This chapter is an introduction to lattice theory, focusing on the results that will be put to use in this volume. §§ 4.6 and 4.8 further elaborate some aspects of lattice theory introduced here. Lattice theory is a rich subject in its own right. We can highly recommend [Birkhoff 1967], [Crawley and Dilworth 1973], and [Grätzer 1978] for fuller expositions of lattice theory.

Lattices were defined in Chapter 1 as algebras of the form $\langle L, \wedge, \vee \rangle$, with two binary operations called **meet** (designated by $\wedge$) and **join** (designated by **join**), for which the following equalities hold true for all $a, b, c \in L$:

$$\begin{array}{rclcrcl} a \wedge b & = & b \wedge a & & a \vee b & = & b \vee a \\ a \wedge (b \wedge c) & = & (a \wedge b) \wedge c & & a \vee (b \vee c) & = & (a \vee b) \vee c \\ a \wedge a & = & a & & a \vee a & = & a \\ a \wedge (a \vee b) & = & a & & a \vee (a \wedge b) & = & a \end{array}$$

The equalities on the first line express commutativity, those on the second line associativity, those on the third line idempotency, and those on the last line absorption. These equalities are called the **axioms of lattice theory**. An alternative system of notation in common use denotes join by "$+$" and meet by "$\cdot$" (or simply by juxtaposition).

Lattices can also be viewed as special ordered sets. Let L be a nonempty set and $\leq$ be a binary relation on L. The relation $\leq$ is said to be an **order** on A if and only if for all $a, b, c \in L$

i. (Reflexivity) $a \leq a$;

ii. (Anti-symmetry) If $a \leq b$ and $b \leq a$, then $a = b$;

iii. (Transitivity) If $a \leq b$ and $b \leq c$, then $a \leq c$.

Orders have frequently been referred to as partial orders in the literature. Let $\leq$ be an order on L and let $X \subseteq L$. An element $a \in L$ is called an **upper (lower) bound** of X if and only if $x \leq a$ ($a \leq x$) for all $x \in X$; a is called a **least upper bound** of X if and only if a is an upper bound of X and $a \leq b$ for all upper bounds b of X. Dually, a is called a **greatest lower bound** of X if and only if a is a lower bound of X and $b \leq a$ for all lower bounds b of X. Since $\leq$ is anti-symmetric, least upper bounds and greatest lower bounds are unique, when they exist. The least upper bound of X is called the **supremum** of X and is denoted by $\sup X$; the greatest lower bound of X is also called the **infimum** of X and is denoted by $\inf X$.

DEFINITION 2.1. Let L be a nonempty set. A **lattice order** on L is an order on L with respect to which every subset of L with exactly two elements has both a least upper bound and a greatest lower bound. The relational structure $\langle L, \leq \rangle$ is called a **lattice ordered set** if $\leq$ is a lattice order on L.

Each lattice has an underlying lattice order, from which the lattice operations of join and meet can be recovered. More precisely, let $\mathbf{L}$ be the lattice $\langle L, \wedge, \vee \rangle$ and define $\mathbf{L}^o$ to be $\langle L, \leq \rangle$ where $\leq$ is the binary relation on L specified, for all $a, b \in L$, by

$$a \leq b \text{ if and only if } a = a \wedge b.$$

Routine calculations reveal that $\leq$ is a lattice order on L. Now suppose $\mathbf{L}$ is some lattice ordered set $\langle L, \leq \rangle$ and define $\mathbf{L}^a$ to be $\langle L, \wedge, \vee \rangle$ where the two binary operations are specified, for all $a, b \in A$, by

$$a \wedge b = \inf\{a, b\}$$
$$a \vee b = \sup\{a, b\}.$$

Again, routine calculations reveal that $\mathbf{L}^a$ is, indeed, a lattice. Moreover, $\mathbf{L}^{oa} = \mathbf{L}$ for every lattice $\mathbf{L}$ and $\mathbf{L}^{ao} = \mathbf{L}$ for every lattice ordered set $\mathbf{L}$.

It is useful to gain some skill at manipulating lattice equations and inclusions. Some of the most frequently used manipulations come up in the following exercises.

Exercises 2.2

1. Let $\mathbf{L}$ be a lattice. Prove that for all $a, b \in L$, $a = a \wedge b$ if and only if $b = a \vee b$.

2. Verify the claims made above.

i. If $\mathbf{L}$ is a lattice, then $\mathbf{L}^o$ is a lattice ordered set.

ii. If $\mathbf{L}$ is a lattice ordered set, then $\mathbf{L}^a$ is a lattice.

iii. If $\mathbf{L}$ is a lattice, then $\mathbf{L}^{oa} = \mathbf{L}$.

iv. If $\mathbf{L}$ is a lattice ordered set, then $\mathbf{L}^{ao} = \mathbf{L}$.

3. Let $\mathbf{L}$ be a lattice and let $a,b,c,d \in L$. Prove that if $a \leq b$ and $c \leq d$, then $a \wedge c \leq b \wedge d$ and $a \vee c \leq b \vee d$.

4. Let $\mathbf{L}$ be a lattice and let $a,b,c \in \mathbf{L}$. Prove that

i. $a \leq c$ and $b \leq c$ if and only if $a \vee b \leq c$.

ii. $c \leq a$ and $c \leq b$ if and only if $c \leq a \wedge b$.

iii. $a \wedge b \leq a \vee b$.

5. Prove that the two axioms of lattice theory that express idempotency are derivable from the other six axioms.

6. Let $\mathbf{L}$ be a lattice and let $a,b,c \in \mathbf{L}$. Prove that

i. $(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee c)$.

ii. $a \vee (b \wedge c) \leq (a \vee b) \wedge (a \vee c)$.

iii. $(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) \leq (a \vee b) \wedge (b \vee c) \wedge (c \vee a)$.

iv. $(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee (a \wedge c))$.

7. Let $\mathbf{L} = \langle L, \wedge, \vee \rangle$ be a lattice. Prove that $\langle L, \wedge, * \rangle$ is a lattice if and only if $a * b = a \vee b$ for all $a,b \in L$.

Let $\mathbf{L}$ be a lattice, or more generally, an ordered set. For $a,b \in L$, we say that b **covers** a (and that b is an **upper cover** of a and that a is a **lower cover** of b), and we write $a \prec b$ if and only if $a < b$ and $\{c : a < c < b, c \in L\}$ is empty. $\mathbf{L}$ is said to be **bounded** provided $\mathbf{L}$ has both a greatest element and a least element. We use 1 to denote the greatest element of $\mathbf{L}$ and 0 to denote the least element of $\mathbf{L}$, whenever they exist. If $\mathbf{L}$ has a least element 0, then the upper covers of 0 are called the **atoms** of $\mathbf{L}$. Dually, if $\mathbf{L}$ has a greatest element 1, then the lower covers of 1 are called the **coatoms** of $\mathbf{L}$. Elements $a,b \in L$ are said to be **comparable** if $a \leq b$ or $b \leq a$; $a \parallel b$ denotes that a and b are **incomparable**. A subset of L in which any two elements are comparable is called a **chain**, whereas a subset in which no two elements are comparable is called an **antichain**. $I[a,b]$ denotes the **interval** from a to b–that is, the set $\{c : c \in L \text{ and } a \leq c \leq b\}$. We also use $I(a]$ to denote $\{c : c \in L \text{ and } c \leq a\}$ and $I[a)$ to denote $\{c : c \in L \text{ and } a \leq c\}$.

By using the covering relation, it is possible to draw diagrams of finite lattices and of certain infinite lattices. The practical usefulness of these diagrams is great, and the reader is encouraged to draw lattice diagrams whenever that may seem helpful. A **Hasse diagram**Hasse, H. of the lattice $\mathbf{L}$ is obtained by arranging

the elements of $\mathbf{L}$ as points on a plane in such a way that if $a < b$, then the point representing b is above the point representing a. lattice diagram|seeHasse diagram of a lattice Hasse diagram of a lattice lattice(s)!Hasse diagrams of Then a line segment is drawn between any two points that constitute a covering in $\mathbf{L}$. For those lattices for which the lattice ordering is the transitive closure of the covering relation (this includes all finite lattices), a Hasse diagram completely determines the lattice. The ability to draw revealing Hasse diagrams of lattices and other ordered sets is an acquired skill. A lattice does not have to be very large before many strikingly different ways of drawing its Hasse diagrams becomes available. Generally speaking, a diagram is most useful when it is spread out and reduces line crossings to a minimum. Figure 2.1 consists of just a few of the lattice diagrams we will use at various places in these volumes. One of the most elaborate lattice diagrams in this work is given in the second volume, where the bulk of two sections is essentially devoted to providing instructions for the drawing of the diagram.

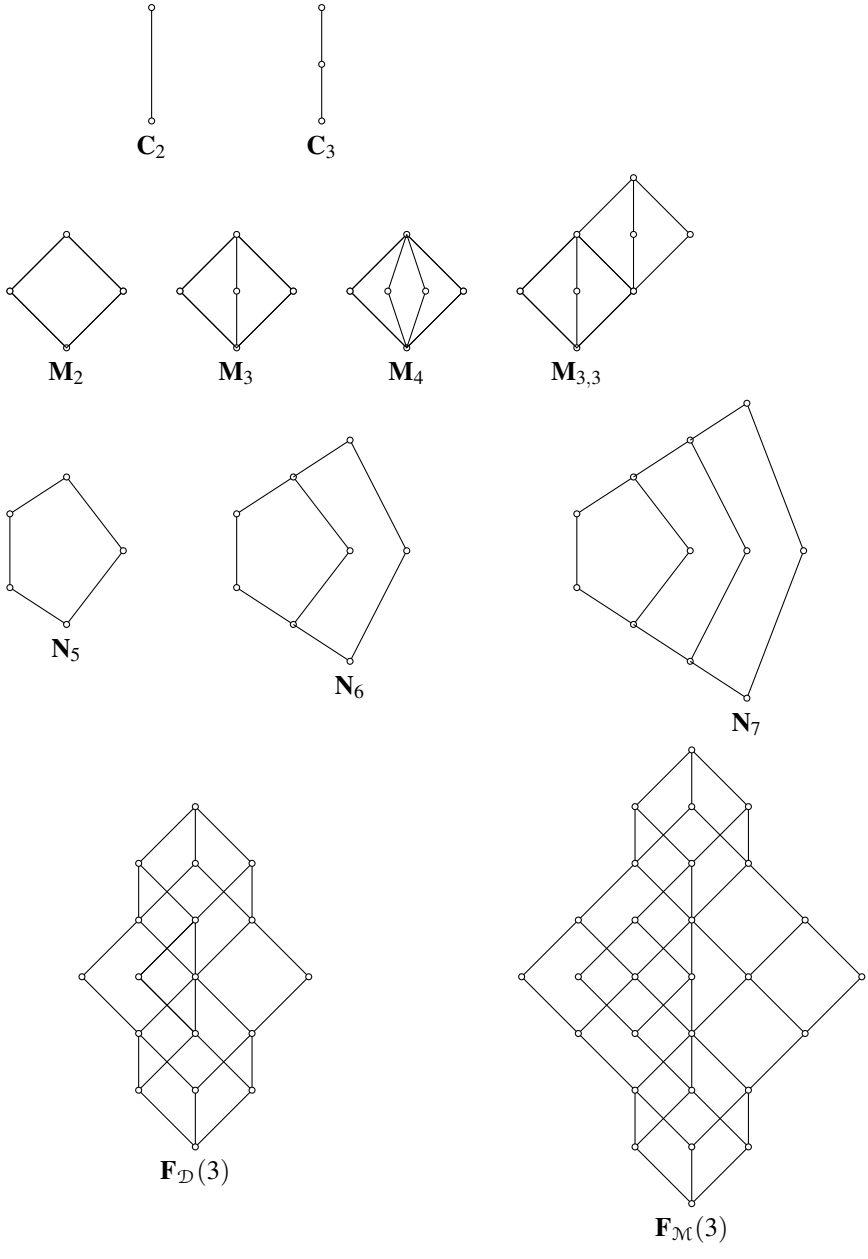

$\mathbf{C}_2$
$\mathbf{C}_3$
$\mathbf{M}_2$
$\mathbf{M}_3$
$\mathbf{M}_4$
$\mathbf{M}_{3,3}$
$\mathbf{N}_5$
$\mathbf{N}_6$
$\mathbf{N}_7$
$\mathbf{F}_{\mathcal{D}}(3)$
$\mathbf{F}_{\mathcal{M}}(3)$

Figure 2.1:

Some caution needs to be exercised with Hasse diagrams. Figure 2.2 is a diagram that looks very much like a lattice diagram but is not:

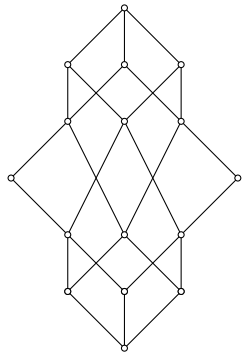


Figure 2.2:

It is apparent that when a lattice diagram is turned upside down, another lattice diagram is obtained. Considering the definitions, we see that $\leq$ becomes $\geq$ and that $\wedge$ and $\vee$ have been interchanged. Of course, this is a consequence of the obvious symmetry of the axioms of lattice theory. More formally, we have the **principle of duality for lattices**:

If $\langle L, \wedge, \vee \rangle$ is a lattice, then $\langle L, \vee, \wedge \rangle$ is a lattice.

If $\langle L, \leq \rangle$ is a lattice ordered set, then $\langle L, \geq \rangle$ is a lattice ordered set.

As a consequence, if σ is a statement that is true in every lattice and σ' is a statement obtained from σ by interchanging $\leq$ and $\geq$ and interchanging $\wedge$ and $\vee$, then σ' is true in every lattice.

If $\mathbf{L} = \langle L, \wedge, \vee \rangle$ is a lattice, then $\mathbf{L}^{\partial}$ is the lattice $\langle L, \vee, \wedge \rangle$ and it is called the **dual** of $\mathbf{L}$. Similar notation applies to lattice ordered sets.

Now let $\mathbf{L} = \langle L, \leq \rangle$ and $\mathbf{L}' = \langle L', \geq' \rangle$ be two lattice ordered sets. A function f from L into L' is said to be **isotone** or **order preserving** if and only if, for all $a, b \in L$, $a \leq b$ implies $f(a) \leq' f(b)$. It is easy to check that every homomorphism between two lattices is isotone. But not every isotone map is a homomorphism, as Figure 2.3 reveals.

On the other hand, if h is a one-to-one isotone map from L onto L', and h^{-1} is also isotone, then h is an isomorphism from the lattice $\mathbf{L}$ onto the lattice $\mathbf{L}'$.

It is also important to realize that while $\leq$ may be a lattice order on L and L' may be a subset of L on which $\leq$ induces a lattice order, it can happen that L' is not a subuniverse of the lattice $\langle L, \wedge, \vee \rangle$. This phenomenon can be traced to the global nature of the definition of join and meet in terms of the ordering. Figure 2.4 is an example. L' consists of the points denoted by $*$.

Exercises 2.3

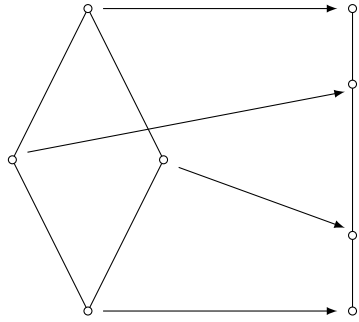


Figure 2.3:

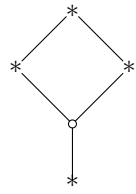


Figure 2.4:

1. Draw a Hasse diagram for the lattice of all subgroups of the symmetric group on $\{0,1,2\}$.

2. Draw the Hasse diagram of the lattice of all subsets of the set $\{0,1,2,3\}$.

3. Prove that every isotone one-to-one function from a lattice $\mathbf{L}$ onto a lattice $\mathbf{L}'$ that preserves incomparability is an isomorphism. Provide an example to show that an isotone one-to-one function from a lattice $\mathbf{L}$ into a lattice $\mathbf{L}'$ need not be a homomorphism.

In the study of lattices, it is helpful to single out individual elements of lattices that have special properties with respect to the ordering of the basic operations.

DEFINITION 2.4. Let $\mathbf{L}$ be a lattice and $a \in \mathbf{L}$. a is **join irreducible** iff $a = b \vee c$ always implies $a = b$ or $a = c$. join!irreducible element join!irreducible element!strictly Dually, a is said to be **meet irreducible** iff $a = b \wedge c$ always implies $a = b$ or $a = c$. meet!irreducible element meet!irreducible element!strictly $J(\mathbf{L})$ denotes the set of all join irreducible members of $\mathbf{L}$ and $M(\mathbf{L})$ denotes the set of all meet irreducible elements of $\mathbf{L}$. a is said to be **join prime**join!prime element

iff $a \leq b \vee c$ always implies $a \leq b$ or $a \leq c$. Dually, a is **meet prime**meet!irreducible prime iff $a \geq b \wedge c$ always implies $a \geq b$ or $a \geq c$.

In the lattice $\mathbf{N}_5$ (see Figure 2.1), every element is either join prime or meet prime. In $\mathbf{M}_3$ only 1 is meet prime and only 0 is join prime, but every element is either join irreducible or meet irreducible. These four properties of an element are preserved in passing to a sublattice that contains the element. Every join prime element is join irreducible and every meet prime element is meet irreducible.

THEOREM 2.5. *Let* $\mathbf{L}$ *be a finite lattice. The following conditions are equivalent:*

i. *The two-element lattice is a homomorphic image of* $\mathbf{L}$.

ii. $\mathbf{L}$ *has a join prime element different from* 0.

iii. $\mathbf{L}$ *has a meet prime element different from* 1.

Proof. Since (i) is a selfdual property and (ii) is the dual of (iii), we need only show that (i) and (ii) are equivalent. Let $\mathbf{C}_2$ be the two-element lattice with $C_2 = \{0,1\}$ and $0 < 1$. Suppose that (i) holds and that $h : \mathbf{L} \twoheadrightarrow \mathbf{C}_2$. Let $h^{-1}(1) = \{a_0, a_1, \ldots, a_n)\}$ and set

$$a = a_0 \wedge a_1 \wedge \cdots \wedge a_n.$$

So $h(a) = h(a_0) \wedge h(a_1) \wedge \cdots \wedge h(a_n) = 1$. To see that a is join prime, suppose $a \leq b \vee c$. Then $h(a) \leq h(b) \vee h(c)$ and so $1 = h(b) \vee h(c)$. Since $h(b), h(c) \in \{0,1\}$, we see that either $h(b) = 1$ or $h(c) = 1$. Thus either $a \leq b$ or $a \leq c$, and a is join prime. $a \neq 0$, since h is onto $\mathbf{C}_2$.

For the converse, suppose that a is a nonzero join prime element of L. Define $h : L \twoheadrightarrow C_2$ by: $h(u) = 1$ iff $a \leq u$. Then, since a is join prime, $h^{-1}(0)$ is closed under join. It is clear that $h^{-1}(1)$ is closed under meet. Moreover, if $h(u) = 0$ and $h(v) = 1$, then $a \leq u \vee v$ and $a \not\leq u \wedge v$, so $h(u \vee v) = 1$ and $h(u \wedge v) = 0$. Thus h is a homomorphism from $\mathbf{L}$ onto $\mathbf{C}_2$, as desired. ■

The condition in the previous theorem that $\mathbf{L}$ be finite cannot be omitted. In the demonstration that (i) $\Rightarrow$ (ii), it played a crucial role. The condition can be weakened. The **ascending chain condition** holds for the lattice $\mathbf{L}$ provided $\mathbf{L}$ has no sublattice isomorphic to the lattice of natural numbers under their usual ordering–that is, provided that every ascending chain in $\mathbf{L}$ is finite. The dual property is referred to as the **descending chain condition**. These conditions have been applied to the ideal lattices of rings.

THEOREM 2.6. *The following conditions are equivalent for any lattice* $\mathbf{L}$.

i. *Every nonempty subset of L has a maximal element.*

ii. $\mathbf{L}$ *satisfies the ascending chain condition.*

Dually, the following conditions are equivalent for any lattice $\mathbf{L}$.

i'. *Every nonempty subset of L has a minimal element.*

ii'. **L** *satisfies the descending chain condition.*

Proof. In view of the duality, we need only prove that (i') is equivalent to (ii').

(i') $\Rightarrow$ (ii') To argue the contrapositive, just notice that any sublattice of **L** ordered like the negative numbers has no smallest element.

(ii') $\Rightarrow$ (i') Again, we argue the contrapositive. Let X be a nonempty subset of L without minimal elements. According to the Axiom of Choice, there is a choice function F for the collection of nonempty subsets of X (i.e., $F(Y) \in Y$ for every nonempty $Y \subseteq X$). Since X has no minimal elements, $\{x : x \in X, \text{ and } x < z\}$ is nonempty, for all $z \in X$. This allows us to define a function g from the set of natural numbers into X by the following recursion:

$$\begin{aligned} g(0) &= F(X) \\ g(n+1) &= F\left(\{x : x \in X \text{ and } x < g(n)\}\right) \text{ for all natural numbers } n. \end{aligned}$$

Evidentally, $\{g(n) : n \text{ is a natural number}\}$ is a subset of L ordered like the negative integers. ■

The theorem above, which is actually very straightforward, nevertheless invokes the Axiom of Choice. One more or less immediate consequence of this theorem is that the descending chain condition is sufficient for the representation of elements of a lattice as joins of finitely many join irreducible elements.

THEOREM 2.7. *If* **L** *is a lattice with the ascending chain condition, then every element of L is a meet of finitely many meet irreducible elements; dually, if* **L** *is a lattice with the descending chain condition, then every element of L is a join of finitely many join irreducible elements.*

Proof. Suppose that **L** has the descending chain condition. Let X be the set of all elements of L that cannot be written as the join of finitely many join irreducible elements. If X is nonempty, then it would have a minimal element x. In this case, x cannot be join irreducible, so there are elements y and z, each properly less than x, such that $x = y \vee z$. Since y and z are both properly less than x, they are not in X. Consequently, y and z can be expressed as joins of join irreducible elements. Thus x can be expressed in the same way. But this means that x cannot belong to X. Thus the supposition that X is nonempty is not tenable. So every element of L can be expressed as the join of join irreducible elements. ■

This theorem resembles the familiar theorem of arithmetic that every natural number can be written as the product of prime numbers. Here, multiplication is replaced by join and primeness replaced by join irreducibility. Actually, the connection is closer than it appears at first. The set of natural numbers, endowed with the operations of forming greatest common divisors and least common multiples, is a lattice with the descending chain condition. The join irreducible elements in this lattice are the powers of prime numbers. Of course, a powerful aspect of

factorization of numbers into primes is the uniqueness of the factorization. For lattices in general, there may be elements that can be expressed as the join of join irreducible elements in many different ways. Later in this chapter, we will return to this topic and demonstrate that uniqueness can be obtained for some interesting classes of lattices.

Exercises 2.8

1. Construct a lattice that has the two-element lattice as a homomorphic image but has no join prime elements.

2. Prove that a lattice with the ascending chain condition has a meet prime element different from 1 iff it has the two-element lattice as a homomorphic image.

3. Construct a finite lattice that has an element that can always be expressed as the join of join irreducible elements in two distinct ways.

4. Prove that if $\mathbf{L}$ satisfies the ascending chain condition, then every chain in $\mathbf{L}$ has a largest element.

2.2 Complete Lattices and Closure Systems

The lattices that prove most significant in the development of the general theorem of algebras are congruence lattices, lattices of clones, subuniverse lattices, lattices of equational theories (and their duals, the lattices of varieties), and interpretability lattices. We have already introduced congruence lattices and subuniverse lattices; the reader will meet the other lattices later in this work. All these lattices have important properties in common that do not hold for all lattices.

A lattice $\mathbf{L}$ is said to be **complete** if and only if every subset of L has both a least upper bound and a greatest lower bound.

A complete lattice $\mathbf{L}$ always has a largest element, usually designated by 1, which is the least upper bound of L, and a smallest element, usually designated by 0, which is the greatest lower bound of L. If $\mathbf{L}$ is a complete lattice and $X \subseteq L$, we use $\bigwedge X$ to denote the greatest lower bound of X and $\bigvee X$ to denote the least upper bound of X. (Thus, in the notation introduced just prior to Definition 2.1, $\sup X = \bigvee X$ and $\inf X = \bigwedge X$.) If $X = \{x_i : i \in I\}$ we also write

$$\bigwedge_I x_i \text{ for } \bigwedge X \text{ and } \bigvee_I x_i \text{ for } \bigvee X.$$

Theorem 1.7 and Theorem 1.18 have corollaries that assert that $\mathbf{Sub}\,\mathbf{A}$ and $\mathbf{Con}\,\mathbf{A}$ are complete lattices, for every algebra $\mathbf{A}$. These two conclusions were established virtually in the same manner, which we now formalize.

DEFINITION 2.9. F is a **closed set system** on the set A if and only if

i. F is a collection of subsets of A,

ii. $A \in F$, and

iii. $\bigcap G \in F$ for every nonempty $G \subseteq F$.

The collection of all subuniverses of an algebra, the collection of all congruence relations of an algebra, the collection of all equivalence relations on a set, and the set of all closed subsets of a topological space are all examples of a closed set system.

DEFINITION 2.10. Let A be a set. C is a **closure operator** on A if and only if C is the function from the power set of A into the power set of A such that

i. $X \subseteq C(X)$ for all $X \subseteq A$,

ii. $C(C(X)) = C(X)$ for all $X \subseteq A$, and

iii. If $X \subseteq Y \subseteq A$, then $C(X) \subseteq C(Y)$.

$\mathrm{Sg}^{\mathbf{A}}$ and $\mathrm{Cg}^{\mathbf{A}}$, for any algebra $\mathbf{A}$, are examples of closure operators, as the familiar operations of forming topological closure and of forming the closed convex hull in a topological vector space.

The connection between closed set systems and closure operators is much like the connection between lattice orderings and lattices, discussed in the previous section. Given a closed set system on A, one may define a closure operator on A; given a closure operator on A, one may define a closed set system on A. Moreover, these two processes are inverses of each other. More precisely, let F be a closed set system on A. Define the function C on the power set of A by

$$C(X) = \bigcap \{K : X \subseteq K \text{ and } K \in F\}$$

for all $X \subseteq A$. C turns out to be a closure operator on A. For the reverse definition, let C be any closure operator on A. Define

$$F = \{C(X) : X \subseteq A\}.$$

In the next set of exercises, the reader is asked to check that F is a closed set system on A and that the two procedures just described reverse each other. The distinction between closed set systems and closure operators is only one viewpoint, not essence.

THEOREM 2.11. *Let C be a closure operator on the set A and let F be its closed set system. Then set-inclusion on F is a lattice ordering with respect to which F becomes a complete lattice, with the lattice operations defined so that for all $G \subseteq F$,*

$$\bigwedge G = A, \text{ if } G \text{ is empty,}$$
$$\bigwedge G = \bigcap G, \text{ if } G \text{ is not empty, and}$$
$$\bigvee G = C\left(\bigcup G\right)$$

Proof. Evidentally, $\bigwedge G$ is the greatest lower bound of G, with respect to the set-inclusion relation.

$\bigvee G$ is an upper bound of G, since if $K \in G$, then $K \subseteq \bigcup G$ and $\bigcup G \subseteq C(\bigcup G)$, since C is a closure operator. Now suppose that H is a closed set such that $K \subseteq H$ for all $K \in G$. Then $\bigcup G \subseteq H$. This implies that $C(\bigcup G) \subseteq C(H) = H$, since C is a closure operator and H is a closed set. Thus $C(\bigcup G)$ is the upper bound of G. So set-inclusion is a complete lattice ordering on F, and the lattice operations are as desired. ■

The converse of this theorem is true as well, at least up to isomorphism, as shown in Theorem 2.12.

THEOREM 2.12. *Every complete lattice is isomorphic to the lattice of all closed sets of some closed set system.*

Proof. Let $\mathbf{L}$ be a complete lattice. Take F from the collection of all principal ideals of $\mathbf{L}$–that is, the all sets of the form:

$$\{a : a \in L \text{ and } a \leq b\} = D_b.$$

It is easy to check that F is a closed set system and that the function that takes b to D_b for all $b \in L$ is an isomorphism. ■

This theorem is a representation theorem for complete lattices, since it renders every abstract complete lattice isomorphic to some concrete complete lattice of closed sets.

The closure operators $\mathrm{Sg}^{\mathbf{A}}$ and $\mathrm{Cg}^{\mathbf{A}}$, where $\mathbf{A}$ is an algebra, were seen in Corollaries 1.11 and 1.22 to have an additional property: The closure of a set is the union of the closures of its finite subsets. This property, which is distinctively algebraic, is not enjoyed by the familiar topological closure operation on the real line.

DEFINITION 2.13. Let C be a closure operator on a set A. C is said to be **algebraic** if and only if $C(X) = \bigcup\{C(Z) : Z \subseteq X \text{ and } Z \text{ is finite}\}$, for all $X \subseteq A$.

Let $\leq$ be an ordering of the set A. A subset B of A is said to be **directed upward by** $\leq$ iff for all $a, b \in B$, there is $c \in B$ such that $a \leq c$ and $b \leq c$.

THEOREM 2.14. *A closure operator is algebraic if and only if the union of any collection of closed sets that is directed upward by $\subseteq$ is itself a closed set.* ■

Knowing the close connection between closure operators and the corresponding complete lattices of closed sets, we can hope for an order-theoretic property consisting of the property of being algebraic for closure operators. This property is suggested by the last corollary to Theorem 1.20.

DEFINITION 2.15. Let $\mathbf{L}$ be a complete lattice. An element $a \in L$ is called **compact** if and only if for all $X \subseteq L$, if $a \leq \bigvee X$, then $a \leq \bigvee Y$ for some finite

$Y \subseteq X$. **L** is said to be **algebraic** if and only if every element of L is the join of a set of compact elements of L.

If C is a closure operator and K is a closed set, we say K is **finitely generated** iff $K = C(Z)$ for some finite set Z. In the lattice of closed sets arising from an algebraic closure operator, the formation of the join of an arbitrary collection M of closed sets can be reduced to the union of the collection of the joins of all the finite subcollections of M.

THEOREM 2.16. *If C is an algebraic closure operator, then its lattice of closed sets is an algebraic lattice, and the compact elements of this lattice are exactly the finitely generated closed sets. Conversely, every algebraic lattice is isomorphic of the lattice of closed sets for some algebraic closure operator.*

Proof. Let C be an algebraic closure operator on A. We first argue that the compact elements of the lattice of closed sets coincide with the finitely generated closed sets. So suppose that Z is a finite subset of A and let $H = C(Z)$. Let G be any collection of closed sets such that $H \subseteq \bigvee G$. Since C is algebraic and Z is finite, there is a finite set $Y \subseteq \bigcup G$ such that $Z \subseteq C(Y)$. Thus $H \subseteq C(Y) \subseteq C(\bigcup G) = \bigvee G$. Now pick $G' \subseteq G$ so that G' is finite and $Y \subseteq \bigcup G'$. So $H \subseteq \bigvee G'$. Therefore, every finitely generated closed set is compact. Now let H be any compact member of the lattice of closed sets. Evidently,

$$H \subseteq \bigvee \{C(Z) : Z \subseteq H \text{ and } Z \text{ is finite}\}.$$

So there are finitely many finite subsets $Z_0, Z_1, \ldots, Z_k$ of H such that

$$\begin{aligned} H &\subseteq \bigvee \{C(Z_0), C(Z_1), \ldots, C(Z_k)\} \\ &= C\left(C(Z_0) \cup C(Z_1) \cup \cdots \cup C(Z_k)\right) \\ &= C(Z_0 \cup Z_1 \cup \cdots Z_k). \end{aligned}$$

By letting $Y = Z_0 \cup Z_1 \cup \cdots \cup Z_k$, we see that

$$Y \subseteq H \subseteq C(Y).$$

Since H is a closed set, we conclude that $H = C(Y)$. Therefore, every compact member of the lattice of closed sets is finitely generated.

The lattice of closed sets is algebraic, since (for any closure operator) every closed set is the union of the closures of its finite subsets.

For the converse, suppose that **L** is an algebraic lattice, and let A be the set of all compact elements of L. For each $b \in L$ define

$$D_b = \{a : a \in A \text{ and } a \leq b\}.$$

It is easy to check that $F = \{D_b : b \in L\}$ is a closed set system and that the map f from L onto F defined by $f(b) = D_b$, for all b, is an isomorphism of **L** onto the lattice **F** of closed sets (that f is one-to-one follows from the fact that **L** is algebraic). To see that the associated closure operator is algebraic, we apply Theorem 2.14.

Let $G = \{D_b : b \in L\}$ be a collection of closed sets directed by $\subseteq$. Notice that for any elements $b, c \in L$, we have $D_c \subseteq D_b$ iff $c \leq b$. So the set X is directed by $\leq$. Let $g = \bigvee X$. Now observe

$$\begin{aligned} a \in D_g \quad & \text{iff } a \text{ is compact and } a \leq \bigvee X \\ & \text{iff } a \text{ is compact and } a \leq \bigvee X' \text{ for some finite } X' \subseteq X \\ & \text{iff } a \text{ is compact and } a \leq b \text{ for some } b \in X \\ & \text{iff } a \in D_b \text{ for some } b \in X \\ & \text{iff } a \in \bigcup G. \end{aligned}$$

So $D_g = \bigcup G$, making $\bigcup G$ a closed set. Thus the closure operator associated with $\mathbf{F}$ is algebraic. ■

On any lattice $\mathbf{L}$, $\mathrm{Sg}^{\mathbf{L}}$ and $\mathrm{Cg}^{\mathbf{L}}$ are algebraic closure operators. There are three other algebraic closure operators connected with $\mathbf{L}$, which we introduce next.

DEFINITION 2.17. Let $\mathbf{L}$ be a lattice and $U \subseteq L$. U is said to be an **ideal** of $\mathbf{L}$ iff U is nonempty, $b \leq a \in U$ implies $b \in U$, and $a, b \in U$ implies $a \vee b \in U$. Dually, U is a **filter** of $\mathbf{L}$ iff U is nonempty, $b \geq a \in U$ implies $b \in U$, and $a, b \in U$ implies $a \wedge b \in U$. U is called **convex** iff whenever $a, c \in U$ and $a \leq b \leq c$, then $b \in U$.

This definition of ideal has much in common with the definition in ring theory. However, the correspondence between ideals in rings and congruences in rings does not carry over to lattices. For example, $\mathbf{M}_3$ (see Figure 2.1) has five ideals but only two congruence relations. One intuition about ideals in lattices is that an ideal specifies a notion of "small elements." The members of the ideal are "small," whereas the members of the lattice that lie outside the ideal are not "small." For example, in the lattice of all subsets of the unit interval, the sets with Lebesgue measure zero constitute an ideal. Ideals of the form $I(a] = \{b : b \leq a\}$ are called **principal ideals**; dually, filters of the form $I[a)$ are called **principal filters**.

It is easy to verify that the intersection of any nonempty collection of ideals of a lattice $\mathbf{L}$ is once more an ideal of $\mathbf{L}$ or it is empty. Thus the collection consisting of the empty set and all the ideals of $\mathbf{L}$ is a closed set system. With respect to $\subseteq$, we obtain a complete lattice. Actually, the collection of ideals of $\mathbf{L}$ (without the empty set) constitutes a sublattice that can fail to be complete. If $\mathbf{L}$ has a least element, then the ideals of $\mathbf{L}$ form a complete lattice. Conversely, if the ideals of $\mathbf{L}$ for a complete lattice, then $\mathbf{L}$ will have a smallest ideal, which is easily seen to be a singleton set whose element must be the least element in the lattice. We adopt the convention that $\mathrm{Idl}\,\mathbf{L}$, which we call the **lattice of ideals of** $\mathbf{L}$, is the lattice of ideals of $\mathbf{L}$ if $\mathbf{L}$ has a least element and is the lattice consisting of the empty set and all the ideals of $\mathbf{L}$ otherwise. The situation for filters is dual to that for ideals. We adopt a similar convention for $\mathrm{Fil}\,\mathbf{L}$, the lattice of filters of

$\mathbf{L}$. Cvx $\mathbf{L}$ denotes the lattice of convex sets in $\mathbf{L}$; this lattice always includes the empty set. $\mathrm{Ig}^{\mathbf{L}}$, $\mathrm{Fg}^{\mathbf{L}}$, and $\mathrm{Cv}^{\mathbf{L}}$ denote the corresponding closure operators.

THEOREM 2.18. *Let* $\mathbf{L}$ *be a lattice.* Idl $\mathbf{L}$, Fil $\mathbf{L}$, *and* Cvx $\mathbf{L}$ *are algebraic lattices and* $\mathbf{L}$ *is isomorphic to a sublattice of both* Idl $\mathbf{L}$ *and* Fil $\mathbf{L}^{\partial}$. *Moreover, if L is finite then* Idl $\mathbf{L} \cong \mathbf{L} \cong$ Fil $\mathbf{L}^{\partial}$. *The intersection of any filter on* $\mathbf{L}$ *with any ideal on* $\mathbf{L}$ *is always a convex subuniverse of* $\mathbf{L}$, *and every convex subuniverse of* $\mathbf{L}$ *arises in this way.*

Proof. The following descriptions of $\mathrm{Ig}^{\mathbf{L}}$, $\mathrm{Fg}^{\mathbf{L}}$, and $\mathrm{Cv}^{\mathbf{L}}$ reveal that they are algebraic closure operators; in view of Theorem 2.16, the corresponding lattices are algebraic lattices. Let $X \subseteq L$. Then

$$\mathrm{Ig}^{\mathbf{L}}(X) = \left\{a : a \leq \bigvee Y \text{ for some finite } Y \subseteq X\right\}$$
$$\mathrm{Fg}^{\mathbf{L}}(X) = \left\{a : \bigwedge Y \leq a \text{ for some finite } Y \subseteq X\right\}$$
$$\mathrm{Cv}^{\mathbf{L}}(X) = \{c : a \leq c \leq b \text{ for some } a, b \in X\}.$$

The reader can easily supply the proofs that these sets are, respectively, an ideal, a filter, and a convex set.

The two maps

$$a \mapsto \{b : b \leq a\} = I(a]$$
$$a \mapsto \{b : a \leq b\} = I[a)$$

are the desired embeddings, and in case L is finite they are easily seen to be surjective. Since filters and ideals are convex subuniverses, the intersection between a filter with an ideal is again a convex subuniverse. Finally, let B be a convex subuniverse. Set

$$F = \{a : b \leq a \text{ for some } b \in B\}$$

and

$$I = \{a : a \leq b \text{ for some } b \in B\}.$$

F is a filter and I is an ideal, since B is a subuniverse. $B = F \cap I$, since B is convex. ■

As a consequence, every lattice is embeddable in an algebraic lattice. On the other hand, some infinite joins or meets that exist in $\mathbf{L}$ may not be preserved by these embeddings. For example, the integers used under their usual ordering, with top and bottom elements adjoined, comprise a complete lattice. Neither of the embeddings described above preserve both infinite joins and meets. in this case, the embeddings can be chosen differently so that arbitrary meets and joins are preserved. In general, this is not possible (see Exercise 2.20(4) below). But, as discussed in Example 2.22(6) below, every lattice is embeddable in a complete lattice in such a way that whatever infinite joins and meets exist will be preserved. Unfortunately, the complete lattices one obtains are no longer algebraic.

In finite lattices, every element is the meet of a set of meet irreducible elements; more generally, as we saw in Theorem 2.7, the same holds for lattices with the ascending chain condition. a version of this statement holds in algebraic lattices as well. Let $\mathbf{L}$ be a complete lattice and $a \in L$. The element a is called **strictly meet irreducible** iff $a = \bigwedge X$ implies that $a \in X$ for every subset X of L. **Strictly join irreducible** elements are defined dually.

THEOREM 2.19. *In an algebraic lattice, every element is the mete of a set of strictly meet irreducible elements.*

Proof. Let $a \in L$ and let $b = \bigwedge X$ where

$$X = \{d : a \leq d \text{ and } d \text{ is strictly meet irreducible}\}.$$

It is clear that $a \leq b$. Since $\mathbf{L}$ is algebraic, in order to prove that $b \leq a$, all we need to do is show that $c \leq a$ for all compact $c \leq b$. For the sake of contradiction, suppose c is compact and $c \leq b$ but $c \not\leq a$. Let $F = \{y : a \leq y \text{ but } c \not\leq y\}$. Plainly $a \in F$, so F is not empty. Since c is compact, the join of any chain in F is again a member of F. Hence every chain in F has an upper bound in F. By Zorn's Lemma, let m be maximal in F. Now m is strictly meet irreducible, since $m < d$ implies $m \vee c \leq d$ by the maximality of m. Thus $m \in X$, and so $b \leq m$, contrary to the choice of $m \in F$, since $c \leq b$. ■

This theorem is a lattice-theoretic rendition of Birkhoff's Subdirect Representation Theorem (Theorem 4.1), which is one of the theorems we will find most useful.

Exercises 2.20

1. **a.** Let F be a closed set system on A. Define C_F on the power set of A by $C_F(X) = \bigcap\{K : X \subseteq K \text{ and } K \in F\}$. Prove that C_F is a closure operator on A.

b. Let C be a closure operator on A. Define $F_C = \{C(X) : X \subseteq A\}$. Prove that F_C is a closed set system on A.

c. For C_F defined as in (a), prove that $F = \{C_F(X) : X \subseteq A\}$.

d. For F_C defined as in (b), prove that $C(X) = \bigcap\{K : X \subseteq K \text{ and } K \in F_C\}$ for all $X \subseteq A$.

2. **a.** Prove that if $\mathbf{L}$ is a lattice in which every set has a least upper bound, then $\mathbf{L}$ is complete.

b. Prove that if $\mathbf{L}$ is a lattice in which every chain has a least upper bound then $\mathbf{L}$ is complete.

c. Prove that if $\mathbf{L}$ is a lattice in which every well ordered chain has a least upper bound, then $\mathbf{L}$ is complete.

3. Prove Theorem 2.14: A closure operator is algebraic iff the union of any collection of closed sets that is directed upward by $\subseteq$ is closed itself.

4. Give an example of a complete lattice that cannot be embedded into an algebraic lattice in such a way that arbitrary (infinite) joins and meets are preserved.

5. (A. Tarski) Prove that if $\mathbf{L}$ is a complete lattice and $f : \mathbf{L} \to \mathbf{L}$ is an isotone map, then $f(a) = a$ for some $a \in L$. (Such an element a is called a **fixed point** of f.)

6. Let C be an algebraic closure operator on A. We say that $X \subseteq A$ is **C-independent** iff $x \notin C(X - \{x\})$ for all $x \in X$.

a. Prove the following are equivalent:

i. For every subset $X \subseteq A$ and $u, v \in A$, if $u \in C(X \cup \{v\})$ and $u \notin C(X)$, then $v \in C(X \cup \{u\})$.

ii. For every subset $X \subseteq A$ and $u \in A$, if X is C-independent and $u \notin C(X)$, then $X \cup \{u\}$ is C-independent.

iii. For every $X \subseteq A$, if Y is a maximal C-independent subset of X, then $C(X) = C(Y)$.

iv. For every Y and X with $Y \subseteq X \subseteq A$, if Y is C-independent, then there is a C-independent set Z with $Y \subseteq Z \subseteq X$ and $C(X) = C(Z)$.

b. Suppose that one of the equivalent conditions of (a) is fulfilled. Prove that X and Y are C-independent and $C(X) = C(Y)$, then $|X| = |Y|$.

It is possible to associate with an arbitrary binary relation two closely connected closure operators. Perhaps for this reason, closure operators and complete lattices are quite commonly met in mathematics and quite frequently useful. Here is how we make this association.

Let A and B be any two classes and let R be a binary relation from A to B (that is, $R \subseteq A \times B$). We are going to define two functions, one from the power set of A into the power set of B and then from the power set of B into the power set of A. These functions are called **polarities** of R. Let $X \subseteq A$ and $U \subseteq B$. By definition we take

$$X^{\rightarrow} = \{b : xRb \text{ for all } x \in X\}$$
$$U^{\leftarrow} = \{a : aRu \text{ for all } u \in U\}.$$

$X^{\rightarrow}$ is read "X polar," and $U^{\leftarrow}$ is read "U polar." The fundamental properties of polarities are gathered in the next theorem. The proof of this theorem is left as an exercise.

THEOREM 2.21. *Let A and B be classes and $R \subseteq A \times B$. Let $^{\rightarrow}$ and $^{\leftarrow}$ be the polarities of R. Then Then*

i. *$X \subseteq X^{\rightarrow\leftarrow}$ and $U \subseteq U^{\leftarrow\rightarrow}$ for all $X \subseteq A$ and all $U \subseteq B$.*

ii. *If $Y \subseteq X \subseteq A$, then $X^{\rightarrow} \subseteq Y^{\rightarrow}$.*

ii'. *If $V \subseteq U \subseteq B$, then $U^{\leftarrow} \subseteq V^{\leftarrow}$.*

iii. *$X^{\rightarrow} = X^{\rightarrow\leftarrow\rightarrow}$ and $U^{\leftarrow} = U^{\leftarrow\rightarrow\leftarrow}$ for all $X \subseteq A$ and $U \subseteq B$.*

iv. *$^{\rightarrow\leftarrow}$ is a closure operator on A whose closed sets are exactly the polars of the subsets of B.*

iv'. *$^{\leftarrow\rightarrow}$ is a closure operator on B, whose closed sets are exactly the polars of the subsets of A.*

v. *The lattice of closed subsets of A is isomorphic with the dual of the lattice of closed subsets of B by the map induced by $^{\rightarrow}$. $^{\leftarrow}$ induces the inverse isomorphism.*

The polarities are said to establish a **Galois connection** between the two closed set systems described in this theorem. Part of the usefulness of such Galois connections resides in the possibility of drawing conclusions concerning one of the closed set systems on the basis of information about the other system. Galois connections also offer a mean of analyzing the underlying relations R. We close this section with a list of examples of Galois connections.

EXAMPLE 2.22. **i.** Let $q(x)$ be a polynomial with rational coefficients and let A be the splitting field of $q(x)$. Let B be the group of automorphisms of A. Define R by

$$aRg \text{ iff } g(a) = a.$$

The resulting closed subsets of A are the subfields of A, and the resulting closed subsets of B are the subgroups of B. This is essentially the connection brought to light by Galois in his investigation of the roots of polynomials.

ii. Let A be the n-dimensional affine space over $\mathbb{C}^n$ and let B be the ring $\mathbb{C}[x_0, x_1, \ldots, x_{n-1}]$. Define R by

$$\bar{v}Rp(\bar{x}) \text{ iff } p(\bar{v}) = 0 \text{ in } \mathbb{C}.$$

The resulting closed subsets of $\mathbb{C}^n$ are known as affine algebraic varieties, and the resulting closed subsets of the polynomial ring are the nilradical ideals. This latter statement is a formulation of Hilbert's Nullstellensatz. This Galois connection is a starting point for the development of algebraic geometry.

iii. Let A be a principal ideal domain and let B an A-module. Define R by

$$aRb \text{ iff } ab = 0.$$

The closed subsets of A turn out to be certain ideals of A called annihilators, and the closed subsets of B are certain submodules. This Galois connection is a key to understanding the structure theory of finitely generated modules over principal ideal domains. This theory in turn comprehends

the Fundamental Theorem of Finitely Generated Abelian Groups and the theorems concerning the existence of uniqueness of the Jordan and rational canonical forms of matrices.

iv. Let U be a set, let Q be a finitary operation on U, and let S be a finitary relation on U. We say that S is **closed under** Q (and that Q **preserves** S) iff S is a subuniverse of $\langle U, Q\rangle^n$, where n is the rank of S. The Galois connection established in the ration R defined by

$$S\ R\ Q \text{ iff } S \text{ is closed under } Q$$

is of considerable importance to us. The resulting closed sets of operations are known as **clones**. Clones will be more carefully introduced in Chapter 4. A chapter in Volume 2 elaborates the theory of clones over finite sets U. On the other side of the duality, notice that if $\langle U, F\rangle$ is an algebra, then the unary relations in $F^{\leftarrow}$ that belong to $F^{\leftarrow}$ are precisely the congruence relations of the algebra.

v. Fix a similarity type. Let A be the class of all algebras of this type and let B be the set of all equations that can be expressed using variables and the operation symbols of the type. Define R by

$$\mathbf{C}\ R\ p \approx q \text{ iff } p \approx q \text{ is true in } \mathbf{C}.$$

While the precise definitions of the concepts of equations and truth are deferred to §4.11, our intent here should be clear. (Associativity and commutativity are both expressed by equations; associativity is true in the multiplicative semigroup of 2×2 matrices over the reals, but commutativity is not.) The close sets of algebras turns out to be exactly the varieties, and the closed sets of equations are the sets closed with respect to logical consequence. This Galois connection is central for our subject. Its fundamental properties are among the chief concerns of Chapter 4 and will be fully developed in later volumes.

vi. Let $\mathbf{L}$ be a lattice and take $A = L = B$. Let R be the binary relation $\leq$ on L. The complete lattice of closed sets of the form $X^{\leftarrow}$ where $X \subseteq L$ is called the **Dedekind-MacNeille completion** of $\mathbf{L}$. The Dedekind-MacNeille completion of the ordered set of rational numbers is (isomorphic to) the ordered set of real numbers. The map that assigns $\{a\}^{\leftarrow}$ to a for every $a \in L$ turns out to be an embedding of $\mathbf{L}$ into its Dedekind-MacNeille completion, justifying the word "completion." This map also preserves whatever infinite joins and meets exist in $\mathbf{L}$.

Exercises 2.23

1. Provide a proof for Theorem 2.21.

2. Prove that the Dedekind-MacNeille completion of the rationals with their usual order is isomorphic to the reals with their usual order.

3. Prove that the contention in Example 2.22(6) that the map described embeds $\mathbf{L}$ into its Dedekind-MacNeille completion and that this embedding preserves whatever joins and meets exist in $\mathbf{L}$.

2.3 Modular Lattices: The Rudiments

The study of congruence lattices is central in the development of our subject. Such lattices must be algebraic, but they need have no other property. It turns out, however, that most of the intensively investigated kinds of algebras, such as groups, rings, modules, Boolean algebras, and lattices themselves, always have congruence lattices with the following property:

For any elements $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$, if $\mathbf{c} \leq \mathbf{a}$, then $\mathbf{a} \wedge (\mathbf{b} \vee \mathbf{c}) = (\mathbf{a} \wedge \mathbf{b}) \vee \mathbf{c}$.

This statement is called the **modular law**, and lattices for which it holds are called **modular lattices**. The significance of the modular law, and indeed of lattices generally, was first realized by Richard Dedekind. In this section, we present the rudiments of the theory of modular lattices.

THEOREM 2.24. *(Dedekind [1900]). The congruence lattice of any group is modular.*

Proof. Let $\mathbf{G} = \langle G, \cdot, ^{-1}, e \rangle$ be a group and let $p(x,y,z)$ denote the group theoretic expression

$$x\left(y^{-1}z\right).$$

For all $a, b, c \in G$, the following equalities hold:

$$p(a,b,b) = a$$
$$p(a,a,b) = b.$$

These two equalities allow us to express the join in $\mathbf{Con}\,\mathbf{G}$ in terms of the composition of relations:

$$\phi \vee \psi = \phi \circ \psi \text{ for any } \phi, \psi \in \text{Con } \mathbf{G}.$$

Indeed, $\phi \cup \psi \subseteq \phi \vee \psi$ is clear. To see that $\phi \vee \psi \subseteq \phi \circ \psi$, it is only necessary to prove that $\phi \circ \psi$ is an equivalence relation. The reflexivity of $\phi \circ \psi$ followed directly from the reflexivity of ϕ and ψ. To see the transitivity, suppose $a\ \phi \circ \psi\ b$ and $b\ \phi \circ \psi\ c$. Pick u and v in G so that $a\ \phi\ u\ \psi\ b$ and $b\ \phi\ v\ \psi\ c$. Observe that $a = p(a,b,b)\ \phi\ p(u,b,v)$ since $a\ \phi\ u$ and $b\ \phi\ v$; since $u\ \psi\ b$ and $v\ \psi\ c$, $p(u,b,v)\ \psi\ p(b,b,c) = c$. Written more briefly,

$$a\ \phi\ p(u,b,v)\ \psi\ c$$

or

$$a\ \phi \circ \psi\ c$$

and so $\phi \circ \psi$ is transitive. But now notice that $\phi \circ \psi \subseteq \phi \circ \psi \circ \phi \circ \psi \subseteq \phi \circ \psi$, where the last inclusion is just the transitivity of $\phi \circ \psi$. The symmetry of $\phi \circ \psi$ follows easily from the symmetry of ϕ and ψ and from the inclusion $\psi \circ \phi \subseteq \phi \circ \psi$. In this way we have verified that $\phi \vee \psi = \phi \circ \psi$.

Now we can easily show that **Con G** is modular. Let ϕ, ψ and θ be congruence relations on **G** so that $\phi \leq \theta$. We will deduce

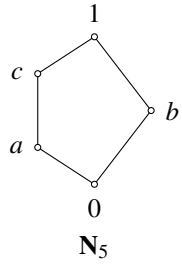
$$(\phi \vee \psi) \wedge \theta \leq \phi \vee (\psi \wedge \theta)$$

or what is the same in this context in view of our reasoning above:

$$(\phi \circ \psi) \cap \theta \subseteq \phi \circ (\psi \cap \theta).$$

So let a and b be elements of G such that $a\ (\phi \circ \psi) \cap \theta\ b$. Hence $a\ \phi \circ \psi\ b$ and $a\ \theta\ b$. Pick $c \in G$ so that $a\ \phi\ c$ and $c\ \psi\ b$. Since $\phi \subseteq \theta$, we obtain $a\ \theta\ c$. So $c\ \theta\ b$, since θ is symmetric and transitive. Thus $c\ (\psi \cap \theta)\ b$, and since $a\ \phi\ c$, we can conclude that $a\ \phi \circ (\psi \cap \theta)\ b$. So **Con G** is modular (consult Exercise 2.2(6) for the apparently missing reverse inclusion). ■

This proof applies to a much wider class of algebras than groups. Indeed, the only property of groups used in this proof was the existence of an expression $p(x,y,z)$, for which two particular equations were satisfies. Any algebra that allows the construction of such an expression $p(x,y,z)$ from its basic operations will have a modular congruence lattice. For these algebras, an even stronger property holds: The join of congruences coincides with the composition of relations. This theme will be taken up again in §4.7 and will be explored in some depth in later volumes.



$\mathbf{N}_5$

Figure 2.5:

Not every lattice is modular. Consider Figure 2.5. Notice that $a \leq c$, but $a \vee (b \wedge c) = a \vee 0 = a$, whereas $(a \vee b) \wedge c = 1 \wedge c = c$. So $\mathbf{N}_5$ is not modular.

There are many statements equivalent to the modular law. Some are included in the next theorem, but others can be found in the next set of exercises.

THEOREM 2.25. *(Dedekind [1900]). For any lattice* **L** *the following statements are equivalent:*

i. **L** *is modular.*

ii. *For any $a,b,c \in L$, if $c \leq a$, then $a \wedge (b \vee c) \leq (a \wedge b) \vee c$.*

iii. $((a \wedge c) \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$ *for all $a,b,c \in L$.*

iv. *For any $a,b,c \in L$, if $a \leq c$, $a \wedge b = c \wedge b$, and $a \vee b = c \vee b$, then $a = c$.*

v. **L** *has no sublattice isomorphic to* $\mathbf{N}_5$.

Proof. (i) $\Leftrightarrow$ (ii) In every lattice, if $c \leq a$, then $(a \wedge b) \vee c \leq a \wedge (b \vee c)$. So the equivalence of (i) and (ii) is clear.

(i) $\Leftrightarrow$ (iii) Since $a \wedge c \leq c$ is true in every lattice, the equation displayed in (iii) must hold in every modular lattice. Conversely, suppose the equation in (iii) holds in L and let a,b, and c be elements of L such that $a \leq c$. Then $a = a \wedge c$, so

$$(a \vee b) \wedge c = ((a \wedge c) \vee b) \wedge c = (a \wedge c) \vee (b \wedge c) = a \vee (b \wedge c).$$

(i) $\Rightarrow$ (iv) According to the following equations, every modular lattice satisfies (iv):

$$\begin{aligned} a &= a \vee a \wedge b && \text{by absorption,} \\ &= a \vee (c \wedge b) && \text{since } a \wedge b = c \wedge b, \\ &= a \vee (b \wedge c) \\ &= (a \vee b) \wedge c && \text{by modularity,} \\ &= (c \vee b) \wedge c && \text{since } a \vee b = c \vee b, \\ &= c. \end{aligned}$$

(iv) $\Leftrightarrow$ (v) Evidently, lattices that satisfy (iv) cannot have sublattices isomorphic to $\mathbf{N}_5$.

(v) $\Leftrightarrow$ (i) We argue the contrapositive. Suppose **L** is not modular. Pick a,b, and c from L so that $a \leq c$ but $a \vee (b \wedge c \neq (a \vee b) \wedge c$. The elements of L in Figure 2.6 constitute a sublattice of **L** isomorphic to $\mathbf{N}_5$:

It is necessary to prove that these elements are actually distinct and that the joins and meets work as indicated in the diagram. First, observe that

$$b \wedge c < a \vee (b \wedge c) < (a \vee b) \wedge c < a \vee b,$$

where the middle $<$ follows since $a \leq c$, and all the inequalities must be strict since $a \vee (b \wedge c) \neq (a \vee b) \wedge c$. (Collapsing the strict inequality at either end collapses the whole chain.) Second, observe that

$$b \wedge c < b < a \vee b,$$

where the strictness once more follows from $a \vee (b \wedge c) \neq (a \vee b) \wedge c$. Also, note that $a \vee (b \wedge c) \not\leq b$ and that $b \not\leq (a \vee b) \wedge c$. Thus, the Hasse diagram drawn above is correct. Finally, to see that the joins and meets are correct, just observe that

$$(a \vee (b \wedge c)) \vee b = a \vee ((b \wedge c) \vee b) = a \vee b$$

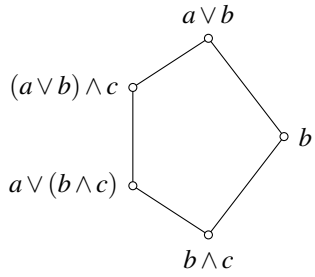


Figure 2.6:

and

$$((a \vee b) \wedge c) \wedge b = ((a \vee b) \wedge b) \wedge c = b \wedge c.$$

■

This theorem reveals several things about the class of all modular lattices. Part (v) allows us to determine the modularity of a lattice by referring to its Hasse diagram, at least if the lattice is relatively small. Actually, this works best in proving a lattice nonmodular through the discovery of a copy of $\mathbf{N}_5$. The task of discovering a copy of $\mathbf{N}_5$ might be exhausting, but once it is in hand, the task is basically finished. Of course, one must be careful to verify that the joins and meets are correct, since the diagram may present only the order structure clearly. As a means for concluding that a lattice is modular, part (v) may not be very helpful, since all possible five-element subsets must be examined. Part (iii) guarantees that subalgebras, homomorphic images, and direct products of modular lattices are once more modular. Modular lattices can be characterized in another very useful manner. manner.

DEFINITION 2.26. Let $\mathbf{L}$ be a lattice and let $a \in L$. Let ϕ_a and ψ_a be the maps from L into L described by

$$\phi_a(x) = x \wedge a \text{ for all } x \in L$$
$$\psi_a(x) = x \vee a \text{ for all } x \in L.$$

Now let $a, a', b, b' \in L$ such that $a \leq b$ and $a' \leq b'$. The interval $I[a,b]$ **transposes up to** $I[a',b']$ iff $b' = b \vee a'$ and $a = b \wedge a'$. Dually, $I[a,b]$ **transposes down to** $I[a',b']$ iff $b = a \vee b'$ and $a' = a \wedge b'$. We use $I[a,b] \nearrow I[a',b']$ to mean $I[a,b]$ transposes up to $I[a',b']$. $I[a,b] \searrow I[a',b']$ is used to mean $I[a,b]$ transposes down to $I[a',b']$.

We call $I[a,b]$ and $I[a',b']$ **transposes** if either of these relations hold, and we call the appropriate map (either $\psi_{a'}$ or $\phi_{b'}$) between the intervals a **perspectivity**

map. Finally, we say that $I[a,b]$ and $I[a',b']$ are **projective** iff there is a finite sequence

$$I[a,b] = I[c_0,d_0], I[c_1,d_1], \ldots, I[c_n,d_n] = I[a',b']$$

such that $I[c_i,d_i]$ and $I[c_{i+1},d_{i+1}]$ are transposes for $i < n$. The map that results from composing the perspectivity maps associated with this sequence of transposes is called a **projectivity map**.

Note that, in general, intervals can be projective by way of many sequences of transposed intervals. The basic facts about perspectivity maps were realized by R. Dedekind[].

THEOREM 2.27 (Dedekind's Transposition Principle)**.** *Let* $\mathbf{M}$ *be a modular lattice and let a and b be elements of M. The map* ϕ_a *induces an isomorphism from* $\mathbf{I}[b, a \vee b]$ *onto* $\mathbf{I}[a \wedge b, a]$*, and* ψ_b *induces the inverse isomorphism. Moreover, the image under either of these maps of a subinterval is a transpose of that subinterval.*

Proof. In view of modularity, for all $x \in I[b, a \vee b]$

$$\psi_b(\phi_a(x)) = (x \wedge a) \vee b = x \wedge (a \vee b) = x$$

and for all $x \in I[a \wedge b, a]$

$$\phi_a(\psi_b(x)) = (x \vee b) \wedge a = x \vee (b \wedge a) = x.$$

Hence, $\psi_b \circ \phi - a$ induces the identity function on $I[b, a \vee b]$, and $\phi_a \circ \psi_b$ induces the identity function on $I[a \wedge b, a]$. Consequently, ϕ_a induces a one-to-one function from $I[b, a \vee b]$ onto $I[a \wedge b, a]$, and ψ_b induces its inverse. To conclude that these functions are isomorphisms, it is only necessary to note that ϕ_a is isotone, since $x \leq y$ implies $x \wedge a \leq y \wedge a$ is true in every lattice.

To see that subintervals are mapped onto transposes, pick x and y so that $b \leq x \leq y \leq a \vee b$. ϕ_a induces a one-to-one map from $I[x,y]$ onto $I[x \wedge a, y \wedge a]$. To see that these two intervals are transposes, let $y' = y \wedge a$. We need to verify that $x \wedge a = x \wedge y'$ and that $y = x \wedge y'$. This is straightforward:

$$x \wedge y' = x \wedge (y \wedge a) = (x \wedge y) \wedge a = x \wedge a$$

and

$$y \leq (a \vee b) \wedge y \leq (a \vee x) \wedge y \leq (x \vee a) \wedge y \quad \begin{array}{l} \text{modularity} \\ \downarrow \\ = x \vee (a \wedge y) = x \vee y' \\ = (x \vee a) \wedge y \leq y \end{array}$$

■

Applied to the congruence lattice of a group, Dedekind's Transposition Principle is another abstraction of one of the familiar "Isomorphism" theorems. In fact,

this is the first of several results that were first established about normal subgroups of a group but which are related to general results for modular lattices. Other group theoretic results that have led to theorems for modular lattices are the Jordan-Hölder Theorem (see Theorem 2.37) and the Krull-Schmidt Theorem (see the Direct Join Decomposition Theorem).

COROLLARY 2.28. *Projective intervals in a modular lattice are isomorphic.*

COROLLARY 2.29. *Let* $\mathbf{L}$ *be a modular lattice and* $a,b,c \in L$ *with* $a \neq b$.

i. *If a and b both cover c, then* $a \vee b$ *covers both a and b.*

ii. *If c covers both a and b, then a and b both cover* $a \wedge b$.

THEOREM 2.30. *The following statements are equivalent for any lattice* $\mathbf{L}$*:*

i. $\mathbf{L}$ *is modular.*

ii. ϕ_a *and* ψ_b *induce inverse isomorphisms between* $I[b, a \vee b]$ *and* $I[a \wedge b, a]$ *for all a and b in L.*

Proof. That (i) implies (ii) is just part of Dedekind's Transposition Principle. For the converse, suppose $\mathbf{L}$ is not modular. Inside $\mathbf{L}$, find a copy of $\mathbf{N}_5$ and label it as in Figure 2.7. Since $\phi_a(c) = a \wedge b = \phi_a(b)$, we see that ϕ_a is not one-to-one. ■

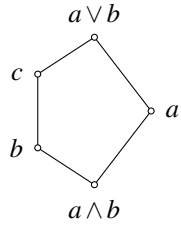


Figure 2.7:

The proof of the next theorem illustrates another use of Dedekind's Transposition Principle.

THEOREM 2.31. *(Birkhoff [1948].) Let* $\mathbf{L}$ *be a modular lattice and let a and b be members of L. Set* $L_0 = I[a \wedge b, a]$ *and* $L_1 = I[a \wedge b, b]$ *and let* $\mathbf{L}_2$ *be the sublattice of* $\mathbf{L}$ *generated by* $L_0 \cup L_1$*. Then* $\mathbf{L}_2 \cong \mathbf{L}_0 \times \mathbf{L}_1$.

Proof. We define the described isomorphism $F : L_0 \times L_1 \to L_2$ for all $x \in I[a \wedge b, a]$ and $y \in I[a \wedge b, b]$ by:

$$F(x,y) = x \vee y.$$

Before showing that F is an isomorphism onto L_2, we note the following simple facts:

i. $I[a\wedge b,a] \nearrow I[b,a\vee b]$ and $I[a\wedge b,b] \nearrow\!\nearrow I[a,a\vee]b$.

ii. If $a\wedge b \le x \le a$, then $x = \phi_a(\psi_b(x)) == a\wedge(x\vee b)$.

iii. If $a\wedge b \le y \le b$, then $y = \phi_b(\psi_a(y)) = b\wedge(y\vee a)$.

iv. If $a\wedge b \le x \le a$, and $a\wedge b \le y \le b$, then

$$\begin{aligned} x\vee y &= \phi_a(\psi_b(x))\vee y = y\vee(a\wedge(x\vee b)) \\ &= (y\vee a)\wedge(x\vee b) = \psi_a(y)\wedge\psi_b(x). \end{aligned}$$

Thus $F(x,y) = x\vee y = \psi_b(x)\wedge\psi_a(y)$ for all $\langle x,y\rangle \in L_0\times L_1$.

CLAIM 0: F is a homomorphism.
Pick $\langle x,y\rangle$ and $\langle x',y'\rangle \in L_0\times L_1$.

$$\begin{aligned} F\left(\langle x,y\rangle\vee\langle x',y'\rangle\right) &= F(x\vee x', y\vee y') \\ &= (x\vee x')\vee(y\vee y') \\ &= F(x,y)\vee F(x',y'). \end{aligned}$$

Thus F preserves joins.

$$\begin{aligned} F\left(\langle x,y\rangle\wedge\langle x',y'\rangle\right) &= F(x\wedge x', y\wedge y') \\ &= \psi_b(x\wedge x')\wedge\psi_a(y\wedge y') \\ &= \psi_b(x)\wedge\psi_b(x')\wedge\psi_a(y)\wedge\psi_a(y') \\ &= \psi_b(x)\wedge\psi_a(y)\wedge\psi_b(x')\wedge\psi_a(y') \\ &= F(x,y)\wedge F(x',y'). \end{aligned}$$

So F preserves meets.
CLAIM 1: F is one-to-one.
Notice that x can be recovered from $x\vee y$ as follows:

$$(x\vee y)\wedge a = ((x\vee b)\wedge(y\vee a)) = (x\vee b)\wedge a = x\vee(b\wedge a) = x.$$

In a similar way, y can be recovered. Hence, $F(x,y) = F(x',y')$ implies $x = x'$ and $y = y'$.
CLAIM 2: F is onto L_2.
Just note that the image of $L_0\times L_1$ under F is a lattice, and it is comprised of all joins $x\vee y$ where $x\in[a\wedge b,a]$ and $y\in[a\wedge b,b]$. ■

The lattice $\mathbf{M}_3$ diagrammed in Figure 2.8 illustrates that, even for finite modular lattices, there may be several distinct ways to represent an element as the join of join irreducible elements. But the next theorem, due to Kurosh [] and Ore [], asserts some uniqueness in such representations.

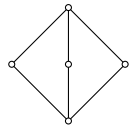


Figure 2.8:

DEFINITION 2.32. Let **L** be a lattice and let M be a finite subset of L. M is called **join irredundant** iff for all proper subsets N of M,

$$\bigvee N < \bigvee M.$$

Meet irredundance is the dual notion.

THEOREM 2.33. *[The Kurosh-Ore Theorem] Let* **L** *be a modular lattice and $a \in L$. Suppose that*

$$a = a_1 \vee a_2 \vee \cdots \vee a_n = b_1 \vee b_2 \vee \cdots \vee b_m$$

where $\{a_i : 0 < i \leq n\}$ and $\{b_j : 0 < j \leq m\}$ are join irredundant sets of join irreducibles and the a_i's are distinct, as are the b_j's. Then $n = m$, and, after renumbering,

$$\begin{aligned} a &= b_1 \vee a_2 \vee a_3 \vee \cdots \vee a_n \\ &= b_1 \vee b_2 \vee a_3 \vee \cdots \vee a_n \\ &\vdots \\ &= b_1 \vee b_2 \vee \cdots \vee b_{n-1} \vee a_n. \end{aligned}$$

Proof. We will first establish that $a = b_j \vee a_2 \vee a_3 \vee \cdots \vee a_n$ for some j. Let $c = a_2 \vee a_3 \vee \cdots \vee a_n$. Then $I[c,a]$ transposes down to $I[c \wedge a_1, a_1]$, so these intervals are isomorphic by Dedekind's Transposition Principle. Since a_1 is join irreducible, a must also be join irreducible in $I[c,a]$. But clearly

$$a = (b_1 \vee c) \vee (b_2 \vee c) \vee \cdots \vee (b_m \vee c),$$

so $a = b_j \vee c$ for some j, as promised.

Suppose, for the moment, that $n < m$. Continuing the above process will ultimately yield a as a join of only n of the b_j's, in contradiction to irredundancy. By symmetry, $m < n$ is also contradictory, so we have $n = m$. The desired equalities now follow easily by iterating the above method. ■

We conclude this section with a simple result concerning complementation in modular lattices. Let **L** be a bounded lattice with largest element 1 and smallest element 0. Let $a \in L$. The element $c \in L$ is called a **complement** of a iff $a \wedge c = 0$ and $a \vee c = 1$. **L** is said to be a **complemented lattice** iff every element of L has a complement; **L** is called **relatively complemented** provided every interval $I[a,b]$ in **L**, when constructed as a sublattice, is a complemented lattice. The lattice of all subspaces of a finite dimensional vector space is easily seen to be a complemented lattice. lattice.

THEOREM 2.34. *Every complemented modular lattice is relatively complemented.*

Proof. Let **M** be a completed modular lattice and let $a \leq x \leq b$ hold in **M**. Let y be a complement of x in **M**. So $x \vee y = 1$ and $x \wedge y \leq 0$. But just notice:

$$\begin{aligned} a &= 0 \vee a = (0 \wedge b) \vee a = ((x \wedge y \wedge b) \vee a = (x \wedge (y \wedge b)) \vee a \\ &= x \wedge ((y \wedge b) \vee a) \end{aligned}$$

and dually,

$$\begin{aligned} b &= 1 \wedge b = (1 \vee a) \wedge b = ((x \vee y) \vee a) \wedge b = (x \vee (y \vee a)) \wedge b \\ &= x \vee ((y \vee a) \wedge b). \end{aligned}$$

But since $(y \vee a) \wedge b = (a \vee y) \wedge b = a \vee (y \wedge b) = (y \wedge b) \vee a$ and since

$$a \leq (y \wedge b) \vee a \leq b$$

we conclude that $(y \wedge b) \vee a$ is a complement of x in $I[a,b]$. ■

Exercises 2.35

1. Let **L** be a finite lattice. Prove that **L** is modular iff $I[a \wedge b, a] \cong I[b, a \vee b]$ for all $a, b \in L$.

2. Construct a lattice that is not modular such that $I[a \wedge b, a] \cong I[b, a \vee b]$ for all $a, b \in L$.

3. Prove that in a modular lattice no element can have two distinct complements that are comparable to each other.

4. A lattice is said to satisfy the **upper covering property** (or said to be **semimodular**) iff given a, b, and c, if $a \prec b$, then either $a \vee c = b \vee c$ or $a \vee c \prec b \vee c$. The **lower covering property** (or **lower semimodularity**) is the dual notion.

   a. Prove that every modular lattice has both the upper and lower covering property.

   b. Construct a nonmodular lattice with both the upper and lower covering property.

   c. Let **L** be a lattice. Prove that **L** is semimodular iff for all $a, b \in L$, if $a \wedge b \prec a$, then $b \prec a \vee b$.

5. Prove that the join irreducible elements of a complemented modular lattice are exactly the atoms of the lattice.

2.4 Modular Lattices wit the Finite Chain Condition

Some very fruitful directions in algebra were opened by the observation that infinite algebras satisfying various "finiteness conditions" were amenable to an almost combinatorial analysis. Often these "finiteness conditions" amount to restrictions on ascending or descending chains in the lattice of congruence relations on the algebras. Noetherian and Artinian rings are specified by just such conditions. Moreover, among vector spaces, the finite dimensional ones are exactly those that have congruence lattices in which every chain is finite. The structure theorems for a very broad range of algebras, which emerge in later chapters (especially Chapter 5), flow from some of the principal theorems concerning modular lattices in which every chain is finite. The length of a finite chain with $n+1$ elements is n.

DEFINITION 2.36. A lattice $\mathbf{L}$ satisfies the **finite chain condition** iff every chain in $\mathbf{L}$ is finite.

Every lattice with the finite chain condition has a greatest element and a least element. According to the Hausdorff Maximality Principle, every lattice has a maximal chain. Hence, lattices with the finite chain condition must have finite maximal chains. A finite maximal chain is one in which each "link" is a covering and in which the top and bottom elements are the 1 and 0 of the lattice. Even so, it is not possible to bound the lengths of the chains in such a lattice, as Figure 2.9 reveals. This sort of pathology does nt happen among modular lattices, as the next theorem confirms. Perhaps this was first realized in the context of congruence lattices of finite groups (the familiar Jordan-Hölder Theorem), and it may have been one of the clues that led Dedekind to formulate the concept of modularity. Ore [] has also been credited with the following result. An alternative proof based on one of the most familiar proofs of the group result is sketched in the exercises.

THEOREM 2.37. *(Dedekind [1900], Birkhoff [1933].) Let* $\mathbf{L}$ *be a modular lattice and let* $a < b$ *in* $\mathbf{L}$*. If there is a finite maximal chain from a to b, then every chain from a to b is finite, and all the maximal ones have the same length. If*

$$a = a_0 \prec a_1 \prec a_2 \prec \cdots \prec a_n = b$$

and

$$a = c_0 \prec c_1 \prec c_2 \prec \cdots \prec c_n = b,$$

then the intervals $I[a_i, a_{i+1}]$ *and* $I[c_j, c_{j+1}]$ *can be matched in such a way that matching intervals are projective.*

Proof. Let us call two finite chains **equivalent** iff they have the same length and have the property described in the final sentence of the theorem (applied to arbitrary chains, not just maximal chains). This specifies an equivalence relation on finite chains between two elements of L.

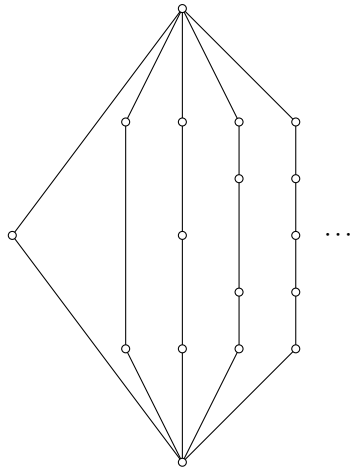


Figure 2.9:

We argue by induction on the length of a finite maximal chain from a to b. Let $a = a_0 \prec a_1 \prec \cdots \prec a_n = b$ be a finite maximal chain from a to b.
INITIAL STEP: $n = 1$. The definition of covering leaves nothing to prove.
INDUCTIVE STEP: $n > 1$. Our induction hypothesis is that the conclusions of the theorem hold for any two elements linked by a maximal chain of length less than n. Let C be a maximal chain from a to b. If $a_1 \in C$, then $C - \{a_0\}$ is a maximal chain from a_1 to b, and the induction hypothesis applied to a_1 and b yields all the desired conclusions. In the remaining case, pick $c \in C$ such that c and a_1 are incomparable. Hence $a_1 \wedge c = a_0$. Let $d = a_1 \vee c$. Thus

$$I[a_0, c] \nearrow I[a_1, d] \text{ and } I[a_0, a_1] \nearrow I[c, d].$$

Let $C_0 = \{c' \in C : c' < c\}$ and $C_1 = \{c' \in C : c \leq c'\}$. Let D_0 be the image of C_0 under the first perspectivity map and let D_1 be any maximal chain from d to b. (Such a chain must exist according to the Hausdorff Maximality Principle applied to $I[d, b]$.) Figure 2.4 suggests an arrangement of these chains. By Dedekind's Transposition Principle, $D_0 \cup \{d\}$ is a maximal chain from a_1 to d, since $C_0 \cup \{c\}$ is a maximal chain from a_0 to c. Thus $D_0 \cup D_1$ is a maximal chain from a_1 to b.

According to the induction hypothesis, $D_0 \cup D_1$ and $a_1 \prec a_2 \prec \cdots \prec b$ are equivalent. By the Dedekind Transposition Principle, $C_0 \cup D_1 \cup \{c\}$ and

$$a = a_0 \prec a_1 \prec \cdots \prec a_n = b$$

are equivalent. The induction hypothesis also yields $\{c\} \cup D_1$ equivalent to C_1, so C is equivalent to $C_0 \cup D_1 \cup \{c\}$. Therefore, C is equivalent to

$$a = a_0 \prec a_1 \prec \cdots \prec a_n = b.$$

■

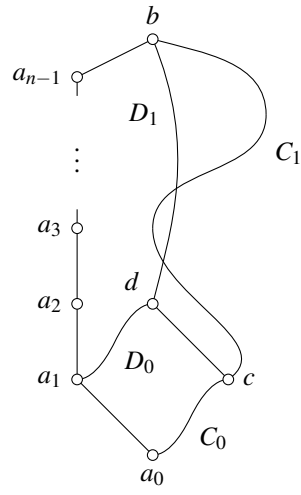


Figure 2.10:

DEFINITION 2.38. A lattice **L** is said to be of **finite height** iff there is a finite upper bound to the length of the chains in **L**. The least such upper bound is called the **height of L**. A lattice **L** is said to be **sectionally of finite height** iff L has a least element 0, and for every $a \in L$, the interval $I[0,a]$ is of finite height. In this case the height of $I[0,a]$ will be denoted by $h(a)$ and called the **height of** a.

Every lattice of finite height satisfies the finite chain condition. From Theorem 2.37, it follows that every modular lattice with the finite chain condition is a lattice of finite height. In modular lattices sectionally of finite height, the height function is very well behaved, as the next theorem reveals. The proof is left as an exercise.

THEOREM 2.39. *Let* **L** *be a modular lattice sectionally of finite height. Then for all* $a, b \in L$,

i. $h(0) = 0$.

ii. *If* $a < b$, *then* $h(a) < h(b)$.

iii. $h(a) + h(b) = h(a \vee b) + h(a \wedge b)$.

■

THEOREM 2.40. *Every bounded modular lattice in which* 1 *is the join of a finite set of atoms is a relatively complemented modular lattice of finite height.*

Proof. Let **L** be a bounded modular lattice in which 1 is the join of a finite set A of atoms. We first argue that **L** is of finite height. In view of Theorem 2.37 and the fact that every chain can be extended to a maximal chain, we need only find one finite maximal chain from 0 to 1. We construct this chain recursively as follows. Let $c_0 = 0$. If $c_i \neq 1$, then pick $a_i \in A$ such that $a_i \not\leq c_i$ and let $c_{i+1} = c_i \vee a_i$.

Since A is finite, this construction stops after finitely many steps and produces a chain $0 = c_0 < c_1 < \cdots < c_n = 1$. But observe that $c_i \wedge a_i = 0$, since a_i is an atom not less than c_i. Consequently, $I[0, a_i]$ transposes up to $I[c_i, c_{i+1}]$. By Dedekind's Transposition Principle, we conclude that $c_i \prec c_{i+1}$ for all $i < n$, so there is a finite maximal chain from 0 to 1.

According to Theorem 2.34, we need only show that **L** is complemented. The construction we just used actually produces complements. Indeed, let $x \in L$ and proceed as follows. Let $d_0 = x$. If $d_i \neq 1$, then pick $a_i \in A$ such that $a_i \not\leq d_i$ and let $d_{i+1} = d_i \vee a_i$.

Since A is finite, this construction stops after finitely many steps, producing let us say $a_0, a_1, \ldots, a_n$. Take $y = a_0 \vee a_1 \vee \cdots \vee a_n$. By the construction we have $x \vee y = 1$. By Theorem 2.39, $x \wedge y = 0$ iff $h(x \wedge y) = 0$. First observe that for each $i < n$,

$$h(x \vee a_0 \vee \cdots \vee a_i \vee a_{i+1}) = h(x \vee a_0 \vee \cdots \vee a_i) + h(a_{i+1}).$$

Consequently,

$$h(x \vee y) = h(x) + h(a_0) + \cdots + h(a_n),$$

and, by the same reasoning applied to the sequence of a_i's, we have

$$h(y) = h(a_0) + \cdots + h(a_n).$$

Therefore $h(x \vee y) = h(x) + h(y)$. By Theorem 2.39,

$$h(x \wedge y) = h(x) + h(y) - h(x \vee y) = 0.$$

■

The three conditions listed in the Theorem 2.39 above are familiar properties of the dimension function applied to subspaces of a finite dimensional vector space.

DEFINITION 2.41. Let **L** be a lattice with least element 0. A function $d : L \to \omega$ with the following properties

i. $d(0) = 0$.

ii. If $a < b$, then $d(a) < d(b)$.

iii. $d(a) + d(b) = d(a \vee b) + d(a \wedge b)$.

is called a **dimensional function** on **L**. A lattice is said to be **finite dimensional** iff it is bounded and there is a dimension function on it.

The height function might be regarded as the natural dimension function on a modular lattice that is sectionally of finite height. This is the case, for example, with the lattice of finite dimensional subspaces of a vector space. The height function h puts $h(a) = 1$ for every atom a in the lattice, a condition we did not include in the definition of dimension function but which will be of use in Chapter 4.

THEOREM 2.42. *For any lattice* $\mathbf{L}$ *the following are equivalent:*

i. $\mathbf{L}$ *is finite dimensional.*

ii. $\mathbf{L}$ *is a modular lattice with the finite chain condition.*

iii. $\mathbf{L}$ *is a modular lattice of finite height.*

Proof. As we remarked after Theorem 2.37, parts (ii) and (iii) are equivalent. Part (i) follows from (iii) by Theorem 2.39, since lattices of finite height are bounded.

To prove that (i) implies (ii), let $\mathbf{L}$ be a bounded lattice with dimension function d. By properties (i) and (ii) of the dimension function, no chain in $\mathbf{L}$ can have length greater than $d(1)$. Therefore, $\mathbf{L}$ has the finite chain condition. Finally, to verify the modularity of $\mathbf{L}$, pick $a,b,c \in L$ with $a \leq c$. Since $\mathbf{L}$ is a lattice, we know that $a \vee (b \wedge c) \leq (a \vee b) \wedge c$. In view of property (ii) of d, equality will hold iff $d(a \vee (b \wedge c)) = d((a \vee b) \wedge c)$. But observe by property (iii) of d:

$$\begin{aligned} d(a \vee (b \wedge c)) &= d(a) + d(b \wedge c) - d(a \wedge b \wedge c) \\ &= d(a) - d(a \wedge b) + d(b \wedge c) \\ &= d(a \vee b) - d(b) + d(b \wedge c) \\ &= d(a \vee b) - d(b \vee c) + d(c) \\ &= d(a \vee b) - d(a \vee b \vee c) + d(c) \\ &= d(a \vee b) + d((a \vee b) \wedge c) - d(a \vee b) \\ &= d((a \vee b) \wedge c). \end{aligned}$$

■

The Kurosh-Ore Theorem is a step toward a unique join decomposition theorem for modular lattices, but it falls short. It is not difficult to devise finite modular lattice in which there are elements that can be written as joins of many different finite sets of join irreducibles. But it is possible, for modular lattices of finite height, to obtain a stronger decomposition theorem , the Direct Join Decomposition Theorem. This stronger result concerns not arbitrary joins of finite sets, but rather joins of sets of a rather restricted kind that we will call directly join independent sets. Moreover, the uniqueness obtained is really "uniqueness up to direct join isotopy," where direct join isotopy is a certain equivalence relation between the elements of the lattice. Just as Theorem 2.37 was inspired by the Jordan-Hölder Theorem from group theory, the Direct Join Decomposition Theorem can be traced to some well-known theorems in group theory. In

Kronecker [] it was shown that any finite Abelian group can be written as a direct product of directly indecomposable Abelian groups in an essentially unique way. This result has been extended in various ways. Perhaps the best known goes under the name of the Krull-Schmidt Theorem, which asserts that every group whose normal subgroup lattice satisfies the finite chain condition can be decomposed into a direct product of directly indecomposable groups in essentially only one way (see also Wedderburn [].) It was Ore [] who realized that this unique factorization property could really be traced to a purely lattice-theoretic property o the congruence lattice of the group. It has turned out to be possible to use the resulting Direct Join Decomposition Theorem to obtain unique direct factorization results for much wider classes of algebras than just finite groups. This point is taken up again in Chapter 4 and more extensively in Chapter 5, which is devoted to unique direct factorizations. Our development of the Direct Join Decomposition Theorem relies heavily on Jónsson [].

DEFINITION 2.43. Let $\mathbf{L}$ be a lattice with least element 0. A subset $M \subseteq L$ is **directly join independent** iff whenever N is a finite subset of M and $a \in M - N$, then $a \wedge \bigvee N = 0$. An element $a \in L$ is called **directly join irreducible** iff $0 < a$ and $a = b \vee c$ with $\{b,c\}$ directly join independent and $b \neq c$, then $b = 0$ or $c = 0$. $\mathrm{IND}(\mathbf{L})$ denotes the collection of directly join independent subsets of $\mathbf{L}$.

For notational convenience, we introduce a partial operation $\oplus$, referred to as **direct join**, on the lattice $\mathbf{L}$. $a \oplus b$ is defined whenever $a \wedge b = 0$, and it takes the value $a \vee b$ in that case. Thus, $a \oplus b = c$ is equivalent to the assertion of the following three conditions:

i. $\{a,b\}$ is directly join independent,

ii. $a \vee b = c$, and

iii. $a \neq b$ or $a = 0 = b$.

DEFINITION 2.44. Let $\mathbf{L}$ be a lattice with least element 0 and let $a,b \in L$. The elements a and b are **directly join isotopic in one step** iff there is $c \in L$ such that $a \oplus c = b \oplus c$. a and b are said to be **directly join isotopic** iff there is a finite sequence $d_0, d_1, \ldots, d_n$ of elements of L such that $a = d_0$, $d_n = b$, and d_i is directly join isotopic with d_{i+1} in one step, for each $i = 0, 1, \ldots, n-1$.

Notice that if a and b are directly join isotopic in a lattice $\mathbf{L}$, then $I[0,a]$ and $I[0,b]$ are projective intervals in $\mathbf{L}$; the associated projectivity map will be referred to as a **join isotopy map**. One-step direct join isotopy is not a transitive relation, even in finite modular lattices; see Exercises 2.49(10).

The notions dual to direct join independence, direct join irreducibility, and direct join isotopy apply to lattices with a greatest element and are referred to, respectively, as **direct meet independence**, **direct meet irreducibility**, and **direct meet isotopy**. Actually, these dual notions 9and variants of them) are the ones used in Chapters 4 and 5 to obtain unique direct factorization results for algebras. But it has now been traditional in lattice theory to approach this mate-

rial from the direct join viewpoint. For the remainder of this section, we will be concerned exclusively with the selfdual class of modular lattices of finite height.

To illustrate these concepts, let $\mathbf{L}$ be the lattice of all subsets of some set X. It is easy to see that a collection of subsets of X will be directly join independent iff it is a collection of pairwise disjoint sets. The only directly join irreducibles are the singleton sets, and two sets are directly join isotopic iff they are equal. Taking $\mathbf{L}$ to be the lattice of subspaces of the three-dimensional real vector space, it is not very hard to classify the directly join independent subsets of L. Accomplishing the same task for the four-dimensional real vector space is more time-consuming but may provide a better intuitive feel for these notions.

In the setting of finite dimensional lattices, direct join independent takes on an especially simple form.

THEOREM 2.45. *Let* $\mathbf{L}$ *be a fintie dimensional lattice with dimension function* d *and let* $M \subseteq L$. *M is directly join independent iff M is finite and* $d\left(\bigvee M\right) = \sum_{a\in M} d(a)$.

Proof. First suppose that M is directly join independent and let N be any finite subset of M. We will prove by induction on $|N|$ that

$$d\left(\bigvee N\right) = \sum_{a\in N} d(a).$$

This will entail that M is finite, since $d(a) \geq 1$ provided that $a \neq 0$ and since $d(1)$ is an upper bound on $d\left(\bigvee N\right)$ for all finite N. In the initial step of the induction, N is empty and the conclusion is immediate. For the inductive step, let $a \in N$ and set $N' = N - \{a\}$. Then

$$\begin{aligned}
d\left(\bigvee N\right) &= d\left(a \vee \bigvee N'\right) \\
&= d(a) + d\left(\bigvee N'\right) - d\left(a \wedge \bigvee N'\right) \\
&= d(a) + d\left(\bigvee N'\right) - d(0) \qquad \text{since } M \text{ is directly join independent} \\
&= d(a) + d\left(\bigvee N'\right) \\
&= d(a) + \sum_{a'\in N'} d(a') \qquad \text{by the induction hypothesis} \\
&= \sum_{b\in N} d(b).
\end{aligned}$$

For the converse, we need the following extension of the dimension formula occurring as (iii) in Definition 2.41: For any n distinct elements $a_0, a_1, \ldots, a_{n-1}$ of L

$$\begin{aligned}
d(a_0 \vee a_1 \vee \cdots \vee a_{n-1}) = d(a_0) + d(a_1) + \cdots + d(a_{n-1}) \\
- [d(a_0 \wedge (a_1 \vee \cdots \vee a_{n-1})) \\
+ d(a_1 \wedge (a_2 \vee \cdots \vee a_{n-1})) + \cdots \\
+ d(a_{n-2} \wedge a_{n-1})].
\end{aligned}$$

This formula can be established by induction. Now observe that the formula above depends on the order in which the a_i's have been indexed. Plainly, we have one such formula for each of the $n!$ ways of indexing available.

Suppose that M is a set with n elements such that

$$d\left(\bigvee M\right) = \sum_{b \in M} d(b).$$

Let N be any subset of M, say with k elements, and pick $c \in N$, setting $N' = N - \{c\}$. We must argue that $c \wedge \bigvee N' = 0$. Now let $M = \{a_0, a_1, \ldots, a_{n-1}\}$ so that $c = a_{n-k}$ and $N' = \{a_{n-k+1}, \ldots, a_{n-1}\}$. According to the extended dimension formula above and the condition just imposed on $d(\bigvee M)$, we conclude that

$$d(a_0 \wedge (a_1 \vee \cdots \vee a_{n-1})) + \cdots + d\left(c \wedge \bigvee N'\right) + \cdots + d(a_{n-2} \wedge a_{n-1}) = 0.$$

Since d only produces non-negative values, we conclude that all the terms of this sum are 0. In particular, $d(c \wedge \bigvee N') = 0$. But this implies that $c \wedge \bigvee N' = 0$, as desired. Hence, M is directly join independent. ■

Before turning to the Direct Join Decomposition Theorem, we gather in the next theorem the fundamental properties of directly join independent sets in modular lattices of finite height that we shall use. Most of these properties follow very easily from the definitions and Theorem 2.45. However, the ten properties listed are more than useful tools. in fact, they constitute all the conditions on the family $\mathrm{IND}(\mathbf{L})$ necessary to establish the Direct Join Decomposition Theorem for the finite dimensional lattice $\mathbf{L}$. As a consequence, any family I of subsets of L that has all the properties attributed below to $\mathrm{IND}(\mathbf{L})$ will give rise to a variant of the Direct Join Decomposition Theorem. We could have introduced an abstract concept of "join independence family," using the ten properties below as a definition, and then established a more general theorem. Observe that the notion of direct join irreducibility depends on the notion of direct join independence. Direct join isotopy and the direct join operation are also derivative notions. To obtain a variant of the Direct Jon Decomposition Theorem for a "join independence family" I, these notions must both be modified by referring them to I in place of $\mathrm{IND}(\mathbf{L})$. The specific notion of direct join independence introduced above would then be one example of a "join independence family." In §5.3, we will invoke the Direct Join Decomposition Theorem for a slightly different notion of join independence. That notion and the one defined in 2.43 are the only kinds of "join independence families" in this volume.

THEOREM 2.46. *Let* $\mathbf{L}$ *be a modular lattice of finite height.*

i. *If* $N \subseteq M \in IND(\mathbf{L})$, *then* $N \in IND(\mathbf{L})$.

ii. *If* $M \in IND(\mathbf{L})$, *then* $M \cup \{0\} \in IND(\mathbf{L})$.

iii. *If* $a \oplus b \in M \in IND(\mathbf{L})$, *then* $(M - \{a \oplus b\}) \cup \{a, b\} \in IND(\mathbf{L})$.

iv. *If* $a, b \in M \in IND(\mathbf{L})$ *and* $a \neq b$, *then* $(M - \{a, b\}) \cup \{a \oplus b\} \in IND(\mathbf{L})$.

v. *If $M \in IND(\mathbf{L})$ and $f : M \to L$ such that $f(x) \leq x$ for all $x \in M$, then $\{f(x) : x \in M\} \in IND(\mathbf{L})$.*

vi. *If $a \oplus a' = b \oplus b' = a \vee b' = a' \vee b$, then $a \oplus b' = a' \oplus b = a \oplus a'$.*

vii. *If $\{a, a'\} \in IND(\mathbf{L})$ with $a \neq a'$ and $b < a$, then $b \oplus a' < a \oplus a'$.*

viii. *If $b \leq a \oplus a'$, $b \not\leq a$, and $\{a \wedge (a' \vee b), b\} \in IND(\mathbf{L})$, then $\{a, b\} \in IND(\mathbf{L})$.*

ix. *If $a = a \oplus b$, then $b = 0$.*

x. *If $a \oplus b$ is directly join isotopic with c, then there are a' and b' such that $c = a' \oplus b'$, and a is directly join isotopic with a' and b is directly join isotopic with b'.*

Proof. **i.** This is completely straightforward.

ii. This is also immediate.

iii. Suppose that $a \oplus b$, $a_1, \ldots, a_{n-1}$ is a list of all the distinct elements of M. To see that $(M - \{a \oplus b\}) \cup \{a, b\}$ is directly join independent, we invoke Theorem 2.45.

$$\begin{aligned} d(a \vee b \vee a_1 \vee a_2 \vee \cdots \vee a_{n-1}) &= d\left((a \oplus b) \vee a_1 \vee a_2 \vee \cdots \vee a_{n-1}\right) \\ &= d(a \oplus b) + d(a_1) + d(a_2) + \cdots + d(a_{n-1}) \\ &= d(a) + d(b) + d(a_1) + d(a_2) + \cdots + d(a_{n-1}). \end{aligned}$$

iv. The argument just given for (iii) can be easily rearranged to prove this part.

v. This follows easily from the definition of direct join independence.

vi. We assume that none of a, a', b, b' is 0, since otherwise the desired result is immediate from the definition of direct join independence. Hence $a \neq a'$ and $b \neq b'$. The hypotheses now give the following dimension equations:

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

In turn, these equations yield

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

From these equations we obtain $d(a \wedge b') + d(a' \wedge b) = 0$. Therefore both $d(a \wedge b') = 0$ and $d(a' \wedge b) = 0$. Hence $a \wedge b' = 0 = a' \wedge b$. Thus both $\{a, b'\}$ and $\{a', b\}$ are directly join independent sets of cardinality two and so $a \oplus b' = a' \oplus b = a \oplus a'$ as desired.

vii. According to (v), $\{b, a'\}$ is directly join independent and $d(b) < d(a)$ since $b < a$. So

$$\begin{aligned} d(b \oplus a') &= d(b) + d(a') \\ &< d(a) + d(a') \\ &= d(a \oplus a') \end{aligned}$$

Thus $b \oplus a' \leq a \oplus a'$ and $b \oplus a' \neq a \oplus a'$.

viii. Just notice that $a \wedge b = a \wedge b \wedge b \leq a \wedge (a' \vee b) \wedge b = 0$. (Here the hypothesis that $b \leq a \oplus a'$ is unnecessary.)

ix. Since $d(a) = d(a) + d(b)$, we conclude that $d(b) = 0$ and so $b = 0$.

x. It suffices to prove that when $a \oplus b$ is directly join isotopic to c in one step. So pick d with $a \oplus b \neq d \neq c$ so that

$$(a \oplus b) \oplus d = c \oplus d.$$

Thus, $I[0, a \vee b]$ and $I[0, c]$ are projective, and the isotopy map that takes $x \in I[0, a \vee b]$ to $(x \vee d) \wedge c$ is a lattice isomorphism by the Dedekind Transposition Principle. Now let $a' = (a \vee d) \wedge c$ and $b' = (b \vee d) \wedge c$. $a' \wedge b' = 0$, since $a \wedge b = 0$, and $a' \vee b' = c$, since $a \vee b = a \vee b$, by the isomorphism. Hence $a' \oplus b' = c$. To see that a and a' are directly join isotopic, just observe:

$$\begin{aligned} a' \oplus d &= a' \vee d \\ &= ((a \vee d) \wedge c) \vee d \\ &= (a \vee d) \wedge (c \vee d) \text{ by modularity} \\ &= (a \vee d) \wedge (c \oplus d) \\ &= (a \vee d) \wedge ((a \vee b) \oplus d) \\ &= (a \vee d) \wedge (a \vee b \vee d) \\ &= a \vee d \\ &= a \oplus d. \end{aligned}$$

■

The properties attributed to IND($\mathbf{L}$) and to the partial operation of direct join by the first four parts of Theorem 2.46 make the direct join easy to manipulate. For example, they entail that direct join is associative in a strong sense. Parentheses can be rearranged without the worry of whether the operations are defined (a concern when dealing with partial operations). 0's can be inserted and deleted without trouble. Also note that direct join is commutative, as a consequence of the definition itself. In the proofs below, we have mostly neglected to point out such uses of Theorem 2.46. The reader should note that $a \oplus a$ is only defined when $a = 0$.

The next lemma is the key to our proof of the Direct Join Decomposition Theorem. This lemma is taken from Jónsson [].

LEMMA. *Let* **L** *be a modular lattice of finite height and let* $a,a',b,b',d \in L$ *such that*

$$a \oplus a' \oplus d = b \oplus b' \oplus d$$

Then there are c *and* c' *with* $c \leq b$ *and* $c' \leq b'$ *such that*

$$a \oplus a' \oplus d = c \oplus c' \oplus a' \oplus d.$$

Proof. Let $e = a \oplus a' \oplus d = b \oplus b' \oplus d$. We view the lemma as an assertion about 6-tuples $(e,d,a,a',b,b'$ of elements of L. The following three cases are exhaustive and mutually exclusive:

i. $e = a \vee b' \vee d = a' \vee b \vee d$.

ii. $a \vee b' \vee d < e$.

iii. $a \vee b' \vee d = e$ but $a' \vee b \vee d < e$.

CASE I: $e = a \vee b' \vee d = a' \vee b \vee d$.

In view of Theorem 2.46 (vi) (and with the help of parts (i), (ii), (iii) as well), the lemma is false. A counterexample to the lemma is a 6-tuple (e,d,a,a',b,b') such that

$$e = a \oplus a' \oplus d \text{ and } e = b \oplus b' \oplus d,$$

but no choice of c and c' will fulfill the lemma. Since **L** is finite dimensional, it has the finite chain condition (Theorem 2.42), so every nonempty subset of L has both minimal and maximal members (Theorem 2.6). Fix e so that it is minimal among all first entries of counterexamples. Next, fix d so that it is maximal among all second entries of counterexamples with first entry e. Of course, there may exist 6-tuples with first entry e and second entry d that are not counterexamples. In fact, we have already observed that 6-tuples falling into Case I cannot be counterexamples. We will prove the lemma by showing that the same applies to the remaining cases.

CASE II: $a \vee b' \vee d < e$.

Let $e_1 = a \vee b' \vee d$. So $e_1 < e$; this means that e_1 is not the first entry of any counterexample. Let

$$a_1 = a \qquad a'_1 = a' \wedge e_1$$
$$b_1 = b \wedge e_1 \qquad b'_1 = b'.$$

By Theorem 2.46(v), $\{a_1,a'_1,d\}$ and $\{b_1,b'_1,d\}$ are both directly join independent. Moreover, modularity yields

$$\begin{aligned} a_1 \vee a'_1 \vee d &= a \vee (a' \wedge e_1) \vee d = (a \vee d) \vee (a' \wedge e_1) \\ &= (a \vee d \vee a') \wedge e_1 = e \wedge e_1 = e_1. \\ b_1 \vee b'_1 \vee d &= (b \wedge e_1) \vee (b' \vee d) = (b' \vee d) \vee (b \wedge e_1) \\ &= (b \vee d \vee b') \wedge e_1 = e \wedge e_1 = e_1. \end{aligned}$$

Thus, $(e_1, d, a_1, a_1', b_1, b_1')$ is a 6-tuple fulfilling the hypotheses of the lemma. It is not a counter example. So pick $c \leq b_1$ and $c' \leq b_1'$ such that $e_1 = c \oplus c' \oplus a_1' \oplus d$.

Now $c \vee c' \vee a' \vee d = c \vee c' \vee a_1' \vee a' \vee d = e_1 \vee a' = a \vee b' \vee d \vee a' = e$. All that remains for Case II is to show that $\{c, c', a', d\}$ is directly join independent. But observe that by Theorem 2.46 (iv),

$$\{a_1', c \oplus c' \oplus d\} \quad \text{is directly join independent}$$

and

$$\{a \oplus d, a'\} \quad \text{is directly join independent.}$$

Another way to write the first of these two sets is

$$\left\{a \wedge (a \vee b' \vee d), c \oplus c' \oplus d\right\}.$$

With the help of Theorem 2.46 (v), we deduce that

$$\left\{a' \wedge \left((a \vee d) \vee (c \oplus c' \oplus d)\right), c \oplus c' \oplus d\right\} \text{ is directly join independent.}$$

Now, Theorem 2.46 (viii) entails that

$$\{a', c \oplus c' \oplus d\} \text{ is directly join independent,}$$

since $c \oplus c' \oplus d \leq e_1 < e = a' \oplus (a \vee d)$. By Theorem 2.46 (i) and (iii), we conclude that $\{c, c', a', d\}$ is directly join independent, as desired. So no 6-tuple beginning with e that falls into Case II is a counterexample to the lemma. Moreover, in Case II, for our fixed e, we can conclude that

$$c < b$$

for otherwise $c = b$, and since $c \leq b_1 < b$, we obtain $b = b_1 = b \wedge e_1$. In turn, this implies that $b \leq e_1$ and so $e_1 = e_1 \vee b = a \vee b' \vee d \vee b = e$, contradicting that $e_1 < e$. Case II is settled.

The following claim, which is easily established using modularity and Theorem 2.46 (v), is used several times in Case III and also in the proof of the Direct Join Decomposition Theorem.

CLAIM: Let $x, y, z \in L$ and define $x^\sharp$ to be $x \wedge z$. If $y \leq x \leq y \oplus z$, then $x = x^\sharp \oplus y$.

■

CASE III: $a \vee b' \vee d = e$ but $a' \vee b \vee d < e$.

Interchanging a with b and a' with b', we obtain the 6-tuple (e, d, b, b', a, a') for our fixed e and d. This 6-tuple fails falls into Case II. Since we have already verified Case II, we pick $c_1 < a$ and $c_1' \leq a'$ so that

$$e = c_1 \oplus c_1' \oplus b' \oplus d.$$

By Theorem 2.46 (iv), $\{c_1', c_1 \oplus b \oplus d\}$ is directly join independent. Now we invoke the claim, with a' as x, c_1' as y, $c_1 \oplus b' \oplus d$ as z, and $a^\sharp = a' \wedge (c_1 \vee b' \vee d)$. Hence $a' = a^\sharp \oplus c_1'$. Further, we have

$$\begin{array}{rcll} e & = & a \oplus a^\sharp \oplus (c_1' \oplus d) & \\ e & = & c_1 \oplus b' \oplus (c_1' \oplus d) & (\star) \end{array}$$

according to Theorem 2.46 (iii) and (iv).
SUBCASE IIIA: $0 < c_1'$.

In this subcase, Theorem 2.46 (vii) yields $d < c_1' \oplus d$. Thus the lemma holds at the 6-tuple $(e, c_1' \oplus d, a, a^\sharp, c_1, b')$, in view of $(\star)$. So pick $c_2 \leq c_1$ and $c_2' \leq b'$ so that

$$e = c_2 \oplus c_2' \oplus a^\sharp \oplus (c_1' \oplus d).$$

Observe that

$$\begin{aligned} e &= c_2 \oplus c_2' \oplus (a^\sharp \oplus c_1') \oplus d \\ &= c_2 \oplus c_2' \oplus a' \oplus d \\ e &= c_2 \oplus a' \oplus (c_2' \oplus d). \end{aligned} \tag{$\star\star$}$$

We apply the claim again, this time taking b' as x, c_2' as y, and $c_2 \oplus a' \oplus d$ as z. Thus for $b^\sharp = b' \wedge (c_2 \vee a' \vee d)$, the claim gives us

$$b' = b^\sharp \oplus c_2'.$$

Hence

$$e = b \oplus b^\sharp \oplus (c_2' \oplus d). \tag{$\star\star\star$}$$

But $c_2' > 0$, for otherwise $e = c_2 \oplus a' \oplus d < a \oplus a' \oplus d = e$, where the strict inequality comes from $c_2 \leq c_1 < a$ (by Theorem 2.46 (vii)). Hence $d < c_2' \oplus d$, again by Theorem 2.46 (vii). In view of $(\star\star($ and $(\star\star\star)$, the 6-tuple

$$(e, c_2' \oplus d, c_2, a', b, b^\sharp)$$

fulfills the hypotheses of the lemma. Since $d < c_2' \oplus d$, the lemma holds at this 6-tuple. So pick $c_3 \leq b$ and $c_3' \leq b^\sharp$ so that

$$e = c_3 \oplus c_3' \oplus a' \oplus (c_2' \oplus d).$$

Note that by Theorem 2.46 (iii) and (iv), we have

$$e = c_3 \oplus (c_3' \oplus c_2') \oplus a' \oplus d.$$

Since $c_3 \leq b$ and $c_3' \oplus c_2' \leq b^\sharp \vee b' \leq b' \vee b' = b'$, this subcase is settled.
SUBCASE IIIB: $c_1' = 0$.

In this case, we have that $c_1 < a$ and

$$\begin{aligned} e &= a \oplus a' \oplus d \text{ and} \\ e &= b' \oplus c_2 \oplus d. \end{aligned}$$

In the event that $a \vee c_2 \vee d = e$, w get $e = a \vee d$. So Theorem 2.46 (v) and (vii) yield $a' = 0$. Then choosing $c = b$ and $c' = b'$ demonstrates the lemma. For hte remainder of cases, we take $a \vee c_2 \vee d < e$. Thus the 6-tuple (e, d, a, a', b', c_1) falls into Case II. So pick $c_2 \leq c_1$ and $c_2' \leq b'$ such that

$$e = c_2 \oplus a' \oplus (c_2' \oplus d). \tag{$\star\star$}$$

Now the same reasoning, word for word, used to complete Subcase IIIA, from the point labeled $(\star\star)$ in that case, can be used to complete this subcase. ■

THEOREM 2.47 (The Direct Join Decomposition Theorem)**.** *Let* **L** *be a modular lattice of finite height. Every element of* **L** *is the join of a finite directly join independent set of directly join irreducible elements. If M and N are finite directly join independent sets of directly join irreducible elements of L such that* $\bigvee M$ *and* $\bigvee N$ *are directly join isotopic, then there is a one-to-one function f from M onto N such that x is directly join isotopic with* $f(x)$ *for each* $x \in M$.

Proof. There are two parts to the theorem: the existence of direct join decompositions and their uniqueness up to direct join isotopy. The existence follows by an easy induction on the dimension of elements. We omit the details except to say that Theorem 2.46 (ix) has a role to play. The uniqueness is established by induction on the smaller of $|M|$ and $|N|$. Without loss of generality, suppose $|N| \leq |M|$.

INITIAL STEP: $|N| = 0$.

In this case, N is empty, so $\bigvee N = 0$. It follows from the definition of direct join isotopy and Theorem 2.46 (ix) that $\bigvee M = 0$. Thus M is empty, as desired, or $M = \{0\}$. The last alternative is excluded because 0 is not directly join irreducible.

INDUCTIVE STEP:

M is nonempty, since $|N| \leq |M|$ and N is nonempty. It follows from Theorem 2.46 (ix) that only 0 can be directly join isotopic to 0. From this and the obvious inductive extension of Theorem 2.46 (x) to arbitrary finite direct joins, pick g to be a one-to-one function from M into $I[0, \bigvee N]$ such that x is directly join isotopic with $g(x)$ for all $x \in M$, and $\bigvee \{g(x) : x \in M\} = \bigvee N$.

Pick $a \in \{g(x) : x \in M\}$ and $b' \in N$. Let $a' = \bigvee (\{g(x) : x \in M\} - \{a\})$ and $b = \bigvee (N - \{b'\})$.

Thus $a \oplus a' = b \oplus b'$. Letting $d = 0$ in the lemma, pick c and c' such that $a \oplus a' = c \oplus c' \oplus a'$, $c \leq b$, and $c' \leq b'$. Hence a and $c \oplus c'$ are directly join isotopic. From Theorem 2.46 (ix) and (x), it follows that any element directly join isotopic to a directly join irreducible element must itself be directly join irreducible. Now, $c \oplus c'$ is directly join isotopic to an element of M (by way of a), so $c \oplus c'$ is directly join irreducible. Thus either $c = 0$ or $c' = 0$>

CASE I: $c = 0$ and a is directly join isotopic with c'.

Using $c' \leq b' \leq c' \oplus a'$, pick $c^\sharp$ so that $b' = c' \oplus c^\sharp$. (This can be done using the claim that was isolated during the proof of the lemma.) Since b' is directly join irreducible and $c' \neq 0$, it follows that $b' = c'$. Therefore

$$a \oplus a' = b \oplus b' = c \oplus c' \oplus a' = b' \oplus a'.$$

In particular,

$$b \oplus b' \approx a' \oplus b'$$

and so a' and b are directly join isotopic. But

$$a' = \bigvee (\{g(x) : x \in M\} - \{a\})$$
$$b = \bigvee (N - \{b\}).$$

According to the induction hypothesis, pick a one-to-one function h' from $\{g(x) : x \in M\} - \{a\}$ onto $N - \{b'\}$ such that y is directly join isotopic with $h'(y)$ for all $y \in$

$\{g(x) : x \in M\} - \{a\}$. Extend h' to a function $h : \{g(x) : x \in M\} \to N$ by setting $h(a) = b'$. Then $h \circ g$ is a one-to-one function from M onto N such that $h(g(x))$ is directly join isotopic with x for all $x \in M$.
CASE II: $c' = 0$ and a is directly join isotopic with c.

Again using the claim established in the proof of the lemma and $c \leq b \leq c \oplus a'$, pick $c^\sharp$ with $b = c \oplus c^\sharp$. By the existence part of this theorem, let C be a finite directly join independent set of directly join irreducible elements such that $c^\sharp = \bigvee C$. So

$$\bigvee (C \cup \{c\}) = b = \bigvee (N - \{b'\}).$$

By the induction hypothesis, let h' be a one-to-one function from $C \cup \{c\}$ onto $N - \{b'\}$ such that $h'(y)$ is directly join isotopic with y for all $y \in C \cup \{c\}$. Extend h' to $h : C \cup \{c\} \cup \{b'\} \to N$ by setting $h(b') = b'$. Now

$$c \oplus c^\sharp \oplus b' = b \oplus b' = c \oplus c' \oplus a = c \oplus a'$$

and in particular $c^\sharp \oplus b'$ is directly join isotopic with a'. But

$$c^\sharp \oplus b' = \bigvee (C \cup \{b'\}) \text{ and}$$
$$a' = \bigvee (\{g(x) : x \in M\} - \{a\}).$$

The induction hypothesis supplies a one-to-one map f' from $\{g(x) : x \in M\} - \{a\}$ onto $C \cup \{b'\}$ such that y and $f'(y)$ are directly join isotopic, for all $y \in \{g(x) : x \in M\} - \{a\}$. Extend f' to f from $\{g(x) : x \in M\}$ onto $C \cup \{b'\} \cup \{c\}$ by setting $f(a) = c$. The desired one-to-one function from M onto N is $h \circ f \circ g$.

■

Ore's formulation of a direct join decomposition theorem for modular lattices of finite height can now be drawn as a corollary.

COROLLARY 2.48. *(Ore [1935], [1936].) Let* **L** *be a modular lattice of finite height. Every element of L is the join of a finite directly join independent set of directly join irreducible elements. If M and N are finite directly join independent sets of directly join irreducible elements such that* $\bigvee M$ *and* $\bigvee N$*, then there is a one-to-one function from M onto N such that* $I[0,x]$ *and* $I[0, f(x)]$ *are projective for all* $x \in M$.

Exercises 2.49

1. **a.** Prove that every element of a relatively complemented lattice of finite height is the join of finitely many atoms.
   b. Prove htat if **L** is a bounded modular lattice and 1 is the join of finitely many atoms, then every element of L is the join of finitely many atoms.

2. Prove Theorem 2.39– i.e., prove that the height function is a dimension function.

3. The idea of this exercise is to reprove Theorem 2.37, following the lines of hte Zassenhaus-Schreier approach to the Jordan-Hölder Theorem of group theory. Thus, we need an analog of the Zassnehaus "Butterfly" Lemma and of the Schreier Refinement Theorem. Let $\mathbf{L}$ be a modular lattice Recall from the proof of Theorem 2.37 that two chains $a_0 < a_1 < \cdots < a_{n-1}$ and $b_0 < b_1 < \cdots < b_{m-1}$ are **equivalent** provided that $n = m$ and the intervals $I[a_i, a_{i+1}]$ and $I[b_j, b_{j+1}]$ can be matched in such a way that matching intervals are projective. Finally, we say that the chain C' is a **refinement** of the chain C in $\mathbf{L}$ iff $C \subseteq C'$.

- **a.** Suppose that $a_0 \leq a_1$ and $b_0 \leq b_1$ in $\mathbf{L}$. Prove that $I[a_0 \vee (a_1 \wedge b_0), a_0 \vee (a_1 \wedge b_1)]$ and $I[b_0 \vee (b_1 \wedge a_0), b_0 \vee (b_1 \wedge a_1)]$ are projective intervals. [Pictorial hint: Draw a diagram of a lattice meeting all the requirements of this statement. There should be some resemblance to a butterfly in this diagram.]
- **b.** Let $a \leq b$ in $\mathbf{L}$. Using (a), prove that any two chains from a to b in $\mathbf{L}$ have equivalent refinements.
- **c.** Deduce Theorem 2.37 from (a) and (b).

4. Prove that if $\mathbf{L}$ is a semimodular lattice of finite height, then any two maximal chains in $\mathbf{L}$ have the same length (i.e., Theorem 2.37 can be established, in part, for semimodular lattices).

5. Let $\mathbf{L}$ be a semimodular lattice with a least element 0. Let a and b be atoms of $\mathbf{L}$ and $c \in L$. Prove that if $c < a \vee c \leq b \vee c$, then $a \vee c = b \vee c$.

6. Let $\mathbf{L}$ be a modular lattice with least element 0 and M be any set of elements of $L - \{0\}$. Prove that M is directly join independent iff $(\bigvee N) \wedge (\bigvee P) = \bigvee (N \cap P)$, for all finite $N, P \subseteq M$.

7. Let $\mathbf{L}$ be a modular lattice with least element 0 and let $a_0, a_1, \ldots, a_n$ be $n+1$ distinct elements of $L - \{0\}$. Prove that $\{a_0, a_1, \ldots, a_n\}$ is directly join independent iff

$$(a_0 \vee a_1 \vee \cdots \vee a_i) \wedge a_{i+1} = 0$$

for all $i < n$.

8. Let $\mathbf{L}$ be the congruence lattice of the three-dimensional vector space over the real numbers. Describe the directly join independent subsets of L and the directly join irreducible elements of L. Do the same for the four-dimensional vector spaces over the real numbers.

***9.** In this exercise, we sketch an alternative approach to Corollary 2.48 that essentially follows Ore's original path to the result. Let $\mathbf{L}$ be a modular lattice of finite height and suppose that

$$a_0 \oplus a_1 \oplus \cdots \oplus a_{n-1} = a.$$

For each $i < n$, let $\bar{a}_i = a_0 \vee \cdots \vee a_{i-1} \vee a_{i+1} \vee \cdots \vee a_{n-1}$.

a. Prove that $\{\bar{a}_i : i < n\}$ is a direclty meet independent set with n elements and that 0 is the meet of this set, if $a = 1$.

b. Let $b \in L$. Prove $\bigvee((b \vee \bar{a}_i) \wedge a_i) = (\bigvee(b \vee \bar{a}_i)) \wedge (\bigvee a_i)$. Now suppose further that $a = b_0 \oplus \cdots \oplus b_{m-1}$ and that all the a_i's and b_j's are directly join irreducible. The definition of the $\bar{b}_j$ is similar to that for $\bar{a}_i$. The goal is to prove that $m = n$ and that there is a permutation f of $\{0, 1, \ldots, n-1\}$ such that

$$a = a_i \oplus \bar{b}_{f(i)} = b_{f(0)} \oplus \cdots \oplus b_{f(i)} \oplus a_{i+1} \cdots \oplus a_{n-1}$$

for all $i < n$. This is accomplished by induction on the dimension $d(a)$ fo the element a. Check the initial step and then do the next steps to esbalish the inductive step of the arugment.

c. Fix $i < n$. Prove that there is $k < m$ such that $a_i \vee \bar{b}_k < a$, then there is $j \neq k$ such that

$$a = b_j \oplus \bar{a}_i = a_i \oplus \bar{b}_j.$$

[Here is a hint: For each $r < m$, define $c_r = (a_i \vee \bar{b}_r) \wedge b_r$. Argue that the c_r's are directly join independent and let their direct join be e. Note that $e = a_i \oplus (e \wedge \bar{a}_i)$, as in the claim used in the lemma for the Direct Join Decomposition Theorem (use (b)). Prove that $e < a$, and use the induction hypothesis on the two direct join decompositions of e. Now, using the dimension function and Theorem 2.46, finish this part (c).]

d. Now dispense with the hypothesis in (c) that $a_i \vee \bar{b}_k < a$ for some k. in view of the interchangeability of the a_i's and b_j's, this amounts to eliminating the possibility that $a_i \vee \bar{b}_k = a$ and $b_k \vee \bar{a}_i < a$ for all $k < m$.

e. Now complete the inductive proof of the assertion made before part (c) above, and deduce it from Corollary 2.48.

10. (R. Freese) Prove that the lattice in Figure 2.11 is modular and that a and c are directly join isotopic but not in one step.

2.5 Distributive Lattices

Just as modularity emerges as a fundamental property of the congruence lattice of groups and algebras closely connected with groups, it turns out that the congruence lattice of lattices themselves and of algebras closely connected with them have a stronger property:

$$\mathbf{a} \wedge (\mathbf{b} \vee \mathbf{c}) = (\mathbf{a} \wedge \mathbf{b}) \vee (\mathbf{a} \wedge \mathbf{c}) \textbf{ for all } \mathbf{a}, \mathbf{b}, \textbf{ and } \mathbf{c}.$$

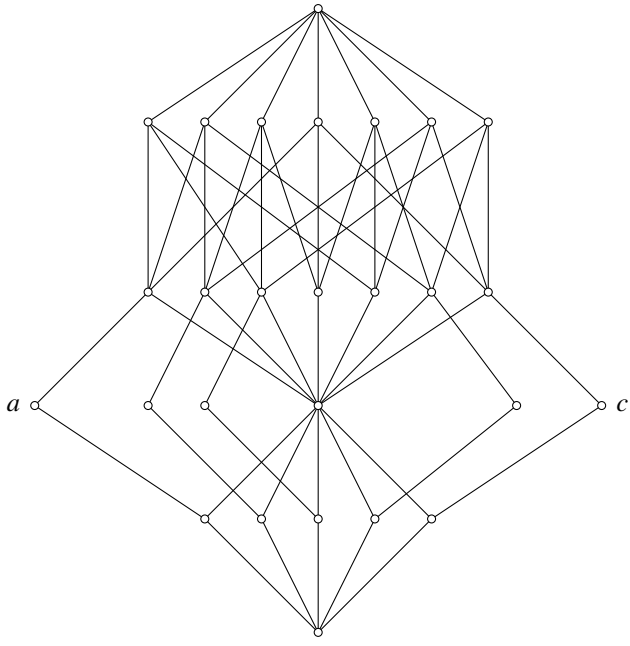


Figure 2.11:

This statement is called the **distributive law**, and lattices for which it holds are called **distributive lattices**. The earliest lattices to be investigated were distributive lattices; the obvious analog with the distributive law for multiplication and addition made this statement appealing. Indeed, some early writers in lattice theory considered all lattices to be distributive. Every chain is easily seen to be distributive, as is the lattice of all subsets of any given set. It is also clear that every distributive lattice is modular, so all conclusions of the last two sections apply to to distributive lattices. We shall see here that most of these results hold in a much sharper form.

THEOREM 2.50. *(Funayama and Nakayama [1942].) The congruence lattice of any lattice is distributive.*

Proof. Let **L** be a lattice. First observe that $(\phi \wedge \psi) \vee (\phi \wedge \theta) \leq \phi \wedge (\psi \vee \theta)$ holds in **Con L**, since it holds in every lattice. We will establish the reverse inclusion. Our argument shares some features of the proof that the congruence lattice of a group must be modular (Theorem 2.24). Let $M(x,y,z)$ be the lattice theoretic expression

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x).$$

Straightforward lattice arguments show that for all $a, b \in L$

$$M(a,a,b) = a$$
$$M(a,b,a) = a$$
$$M(b,a,a) = a.$$

Now suppose that $a(\phi \wedge (\psi \wedge \theta))b$, where $\phi, \psi, \theta \in \mathrm{Con}\ \mathbf{L}$. This means that $a\phi b$ and $a(\psi \vee \theta)b$. According to Theorem 1.24 (ii), $\psi \vee \theta$ is the smallest equivalence relation of L that includes both ψ and θ. This equivalence relation is obviously $\psi \cup \psi \circ \theta \cup \psi \circ \theta \circ \psi \cup \psi \circ \theta \circ \psi \circ \theta \cup \cdots$. Thus there must be a finite sequence $c_0, c_1, c_2, \ldots, c_n$ such that $a = c_0$, $b = c_n$, and

$$\begin{aligned} &c_i \psi c_{i+1} && \text{if } i \text{ is even and } i < n \\ &c_i \theta c_{i+1} && \text{if } i \text{ is odd and } i < n. \end{aligned}$$

Notice that for all $i \leq n$, $a = M(a,a,c_i)\phi M(a,b,c_i)$. This implies that

$$\begin{aligned} &M(a,b,c_i)(\phi \wedge \psi)M(a,b,c_{i+1}) && \text{if } i \text{ is even and } i < n \\ &M(a,b,c_i)(\phi \wedge \theta)M(a,b,c_{i+1}) && \text{if } i \text{ is odd and } i < n. \end{aligned}$$

Since $a = M(a,b,a) = M(a,b,c_0)$ and $b = M(a,b,b) = M(a,b,c_n)$, we conclude that $a((\phi \wedge \psi) \vee (\phi \wedge \theta))b$, as desired. Therefore, $\mathbf{Con}\ \mathbf{L}$ is distributive. ■

This proof applies to a wider class of algebras than lattices. In fact, it applies to any algebra for which there is a term $M(x,y,z)$ that can be built up from the basic operations and variables so that the three equations mentioned in the proof are satisfied. This line of investigation will be taken up in §4.12 and pursued in greater depth in our second volume.

Although every distributive lattice is necessarily modular, there are modular lattices that fail to be distributive. The smallest such lattice is $\mathbf{M}_3$, diagramed in Figure 2.12. In $\mathbf{M}_3$ we have $a \wedge b = 0$ and $a \wedge c = 0$ but $a \wedge (b \vee c) = a \wedge 1 = a$.

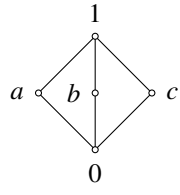


Figure 2.12:

There are many statements equivalent to the distributive law. Some are contained in the next theorem, while others can be found in the exercises below. For the history of this theorem, consult pages 133–134 of Birkhoff [].

THEOREM 2.51. *For any lattice* $\mathbf{L}$ *the following statements are equivalent:*

i. **L** *is distributive.*

ii. $a \wedge (b \vee c) \leq (a \wedge b) \vee (a \wedge c)$ *for all* $a, b, c \in L$.

iii. $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ *for all* $a, b, c \in L$.

iv. $(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) = (a \vee b) \wedge (b \vee c) \wedge (c \vee a)$ *for all* $a, b, c \in L$.

v. For all $a, b, c \in L$, *if* $a \wedge c = b \wedge c$ *and* $a \vee c = b \vee c$, *then* $a = b$.

vi. **L** *has no sublattice isomorphic with either* $\mathbf{N}_5$ *or* $\mathbf{M}_3$.

Proof.

i. $\Leftrightarrow$ **ii.** This follows easily, since $(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee c)$ in all lattices.

i. $\Leftrightarrow$ **iv.** First, assume that **L** is distributive. Obtain (iv) as follows:

$$\begin{aligned}
&(a \vee b) \wedge (b \vee c) \wedge (c \vee a) \\
&\quad = (((a \vee b) \wedge b) \vee ((a \vee b) \wedge c)) \wedge (c \vee a) \\
&\quad = (b \vee ((a \wedge c) \vee (b \wedge c))) \wedge (c \vee a) \\
&\quad = (b \vee (a \wedge c)) \wedge (c \vee a) \\
&\quad = (b \wedge (c \vee a)) \vee ((a \wedge c) \wedge (c \vee a)) \\
&\quad = ((b \wedge c) \vee (b \wedge a)) \vee (a \wedge c) \\
&\quad = (a \wedge b) \vee (b \wedge c) \vee (c \wedge a).
\end{aligned}$$

Now assume that (iv) holds. Then **L** is modular, since if $a \leq c$, then

$$\begin{aligned}
(a \vee b) \wedge c &= (a \vee b) \wedge ((b \vee c) \wedge c) \\
&= (a \vee b) \wedge (b \vee c) \wedge (c \vee a) \\
&= (a \wedge b) \vee (b \wedge c) \vee (c \wedge a) \\
&= (a \wedge b) \vee (b \wedge c) \vee a \\
&= ((a \wedge b) \vee a) \vee (b \wedge c) \\
&= a \vee (b \wedge c).
\end{aligned}$$

Using the modularity of **L**, the distributive law can be deduced as follows:

$$\begin{aligned}
&a \wedge (b \vee c) \\
&\quad = (a \wedge (a \vee b)) \wedge (b \vee c) \\
&\quad = ((a \wedge (c \vee a)) \wedge (a \vee b)) \wedge (b \vee c) \\
&\quad = a \wedge (a \vee b) \wedge (b \vee c) \wedge (c \vee a) \\
&\quad = (a \wedge ((a \wedge b) \vee (b \wedge c)) \vee (c \wedge a)) \qquad \text{by modularity} \\
&\quad = (a \wedge (b \wedge c)) \vee (a \wedge b) \vee (c \wedge a) \qquad \text{by modularity} \\
&\quad = (a \wedge b) \vee (a \wedge c).
\end{aligned}$$

i. $\Leftrightarrow$ iii. A direct proof of this is left as an exercise. However, observe that (iii) is the dual of (i). On the other hand, (iv) is its own dual. Since (i) $\Leftrightarrow$ (iv), we know that the class of distributive lattices is selfdual. So (i) and (iii) are equivalent.

i. $\Rightarrow$ v. Suppose that $a \wedge c = b \wedge c$ and $a \vee c = b \vee c$. Then

$$\begin{aligned} a = a \wedge (a \vee c) &= a \wedge (b \vee c) \\ &= (a \wedge b) \vee (a \wedge c) \\ &= (a \wedge b) \vee (b \wedge c) \\ &= b \wedge (a \vee c) = b \wedge (b \vee c) = b. \end{aligned}$$

v. $\Rightarrow$ vi. It is easy to find violations of (v) in both $\mathbf{N}_5$ and $\mathbf{M}_3$. Thus any lattice satisfying (v) cannot have a sublattice isomorphic to either of these lattices.

vi. $\Rightarrow$ iv. Since $\mathbf{N}_5$ is excluded as a sublattice, we know that $\mathbf{L}$ is modular. We argue the contrapositive. So suppose $\mathbf{L}$ fails to satisfy (iv). Since $(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) \leq (a \vee b) \wedge (b \vee c) \wedge (c \vee a)$ is true in every lattice, we can pick elements $a, b, c \in L$ such that

$$(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) < (a \vee b) \wedge (b \vee c) \wedge (c \vee a).$$

Let $d = (a \wedge b) \vee (b \wedge c) \vee (c \wedge a)$ and $u = (a \vee b) \wedge (b \vee c) \wedge (c \vee a)$ and define

$$\begin{aligned} a' &= (d \vee a) \wedge u \\ b' &= (d \vee b) \wedge u \\ c' &= (d \vee c) \wedge u. \end{aligned}$$

We contend that Figure 2.13 is a sublattice of $\mathbf{L}$. That is, these five elements are distinct and

$$\begin{aligned} a' \vee b' = b' \vee c' = c' \vee a' = u \\ a' \wedge b' = b' \wedge c' = c' \wedge a' = d. \end{aligned}$$

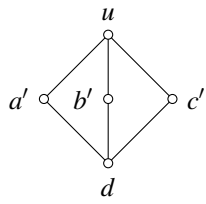


Figure 2.13:

Actually, verifying these equalities suffices, because they imply that the five elements are distinct, in view of $d < u$. Moreover, from modularity we have

$$\begin{aligned} a' &= d \vee (a \wedge u) \\ b' &= d \vee (b \wedge u) \\ c' &= d \vee (c \wedge u). \end{aligned}$$

Because of all this symmetry, it suffices to establish, say

$$a' \wedge c' = d.$$

Reason as follows:

$$\begin{aligned} & a' \wedge c' \\ &= ((d \vee a) \wedge u) \wedge (d \vee c) \wedge u) \\ &= (d \vee a) \wedge (d \vee c) \wedge u \\ &= ((a \wedge b) \vee (b \wedge c) \vee (c \wedge a) \vee a) \wedge (d \vee c) \wedge u \\ &= ((b \wedge c) \vee a) \wedge (a \vee c) \wedge u \\ &= ((b \wedge c) \vee a) \wedge ((a \wedge b \vee c) \wedge u \\ &= ((b \wedge c) \vee a) \wedge ((a \wedge b) \vee c) \wedge (a \vee b) \wedge (b \vee c) \wedge (c \vee a) \\ &= ((b \wedge c) \vee a) \wedge ((a \wedge b) \vee c) \\ &= (b \wedge c) \vee (a \wedge ((a \wedge b) \vee c)) \qquad \text{by modularity} \\ &= (b \wedge c) \vee (((a \wedge b) \vee c) \wedge a) \\ &= (b \wedge c) \vee ((a \wedge b) \vee (c \wedge a)) \qquad \text{by modularity} \\ &= d. \end{aligned}$$

■

The import of this theorem is like that of the analogous Theorem 2.25 concerning modular lattices. Since the class of distributive lattices is specified by a set of equations, it is a variety. Either (iii) or (iv) above tells us that the dual of a distributive lattice is again distributive. Part (vi) characterizes the class of distributive lattices by forbidding certain sublattices, offering some prospect for using a Hasse diagram to check whether a lattice is distributive.

Let $\mathbf{L}$ be a lattice, $a \in L$, and recall the maps $\phi_a(x) = a \wedge x$ and $\psi_a(x) = a \vee x$. Notice that $\phi_a : L \to I(a]$ and $\psi_a : L \to I[a)$. These maps are always isotone, but even in the modular lattice $\mathbf{M}_3$ they fail to be a homomorphism. In distributive lattices, the situation is nicer.

THEOREM 2.52. *The following statements are equivalent for any lattice* $\mathbf{L}$*:*

i. $\mathbf{L}$ *is distributive.*

ii. *For any* $a \in L$*, both* ϕ_a *and* ψ_a *are homomorphisms.*

In modular lattices, projective intervals are isomorphic by a composition of a sequence of such maps associated to a sequence of transposed intervals. In general, these sequences can be arbitrarily long, with no way to shorten them. In distributive lattices, the situation is nicer.

THEOREM 2.53. *Let* $\mathbf{L}$ *be a distributive lattice. If* $I[a,b]$ *and* $I[c,d]$ *are projective in* $\mathbf{L}$*, then either these two intervals are transposes or there are intervals* $I[u,v]$ *and* $I[u',v']$ *such that*

$$I[a,b] \nearrow I[u,v] \searrow I[c,d]$$

and

$$I[a,b] \searrow I[u',v'] \nearrow I[c,d].$$

Proof. First, observe that in any lattice, if $I[a_0,b_0]$ transposes up to $I[a_1,b_1]$, which transposes up to $I[a_2,b_2]$, then $I[a_0,b_0]$ transposes up to $I[a_2,b_2]$ and that a similar phenomenon happens for transposing down. The conclusion of the theorem will follow if we can show how to "reverse the kinks" in a chain of perspectivity maps. More precisely, we want to show that if $I[a_0,b_0] \nearrow I[a_1,b_1] \searrow I[a_2,b_2]$, then there are a_3 and b_3 such that $I[a_0,b_0] \searrow I[a_3,b_3] \nearrow I[a_2,b_2]$. Just define

$$a_3 = a_0 \wedge a_2$$

and

$$b_3 = b_0 \wedge b_2$$

From symmetry considerations, it is enough to demonstrate that

$$a_0 \wedge b_3 = a_3$$

and

$$a_0 \vee b_3 = b_0.$$

From the first equation, we have:

$$\begin{aligned} a_0 \wedge b_3 &= a_0 \wedge b_0 \wedge b_2 \\ &= a_0 \wedge b_2 \\ &= b_0 \wedge a_1 \wedge b_2 \\ &= b_0 \wedge a_2 \\ &= a_0 \wedge b_2 \wedge b_0 \wedge a_2 \\ &= a_0 \wedge a_2 \\ &= a_3. \end{aligned}$$

For the second equation, we have:

$$\begin{aligned} a_0 \vee b_3 &= a_0 \vee (b_0 \wedge b_2) \\ &= b_0 \wedge (a_0 \vee b_2) \quad \text{by modularity} \\ &= b_0 \wedge ((b_0 \wedge a_1) \vee b_2) \\ &= (b_0 \wedge a_1) \vee (b_0 \wedge b_2) \quad \text{by modularity} \\ &= b_0 \wedge (a_1 \vee b_2) \quad \text{by distributivity} \\ &= b_0 \wedge b_1 \\ &= b_0. \end{aligned}$$

■

Another useful property that holds for distributive lattices but not for modular lattices in general is presented in the next lemma. Recall the notions of join prime and meet prime from Definition 2.4.

LEMMA 2.54. *In any distributive lattice, an element is join irreducible iff it is join prime; it is meet irreducible iff it is meet prime.*

Proof. We concern ourselves only with the "join" aspects of the theorem. The "meet" statement will follow, since the class of distributive lattices is selfdual. In any lattice, join prime elements are always join irreducible. So let a be a join irreducible element in the distributive lattice $\mathbf{L}$. To see that a is join prime, suppose that $a \leq b \vee c$. So $a = a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$, by distributivity. Because a is join irreducible, either $a = a \wedge b$ or $a = a \wedge c$. Hence, either $a \leq b$ or $a \leq c$. ■

THEOREM 2.55. *Let* $\mathbf{L}$ *be a distributive lattice and let* $N, M \subseteq L$ *where both* N *and* M *are finite join irredundant sets of join irreducible elements. If* $\bigvee N = \bigvee M$*, then* $N = M$.

Proof. Let $a \in N$. Then a is join irreducible, and so, by Lemma 2.54, a is join prime. Now $a \leq \bigvee N = \bigvee M$. Hence we can pick $b \in M$ with $a \leq b$, since a is join prime. Likewise, we can pick $c \in N$ such that $b \leq c$. Therefore $a \leq c$. Since N is join irredundant, we obtain $a = c$. Hence, $a = b \in M$. Consequently, $N \subseteq M$. The reverse inclusion is obtained by a similar argument. ■

THEOREM 2.56. *Let* $\mathbf{L}$ *be a distributive lattice with the descending chain condition. For every element* $a \in L$*, there is a unique join irredundant finite set* N *of join irreducible elements such that* $a = \bigvee N$.

Proof. The existence of the set N is guaranteed by Theorem 2.7, and uniqueness is just a restatement of Theorem 2.55. ■

Theorem 2.56 has a much stronger conclusion than the Kurosh-Ore Theorem (2.33), which holds more generally for all modular lattices. This simple

result has some use in algebraic geometry, and we will use it in the investigation of subdirect representations of algebras with distributive congruence lattices.

As shown by $\mathbf{M}_3$, elements of modular lattices can have several complements. This cannot happen in distributive lattices. In view of Theorem 2.51(v), an element of a distributive lattice can have at most one complement relative to any bounded interval. Thus complements and relative complements in distributive lattices are unique whenever they exist. According to Dilworth [], there are nonmodular lattices in which every element has a unique complement. The construction is very elaborate and is not included here. On the other hand, every lattice in which relative complements are unique must be distributive, by Theorem 2.51. See Exercise 2.63(7) regarding uniquely complemented modular lattices. The complemented elements in a distributive lattice can be used to decompose the lattice.

THEOREM 2.57. *Let* $\mathbf{L}$ *be a bounded distributive lattice and let* $a, a^* \in L$ *where* a *and* a^* *are complements of each other.* $\mathbf{L} \cong \mathbf{I}(\mathbf{a}] \times \mathbf{I}[\mathbf{a})$.

Proof. Define $f : L \to I(a] \times I[a)$ by $f(x) = (x \wedge a, x \vee a)$ for all $x \in L$. This f is the desired isomorphism. Given distributivity, the demonstration that f is a homomorphism presents no difficulty, so we omit it. To see that f is one-to-one, suppose $f(c) = f(b)$. This means that $c \wedge a = b \wedge a$ and $c \vee a = b \vee a$. But then, by Theorem 2.51 (v), $c = b$. Finally, f is onto $I(a] \times I[a)$: suppose $c \leq a \leq b$. Just observe that

$$((b \wedge a^*) \vee c) \vee a = b$$

and

$$((b \wedge a^*) \vee c) \wedge a = c.$$

So $f((b \wedge a^*) \vee c) = (c, b)$. ■

The converse of this theorem holds in the following sense. Suppose that $\mathbf{L} = \mathbf{L_0} \times \mathbf{L_1}$ where $\mathbf{L}_0$ and $\mathbf{L}_1$ are bounded lattices. Then the elements $\langle 1,0\rangle$ and $\langle 0,1\rangle$ are complements of each other in $\mathbf{L}$ and $\mathbf{L}_0 \cong \mathbf{I}(\langle 1,0\rangle]$ while $\mathbf{L}_1 \cong \mathbf{I}[\langle 1,0\rangle)$. Distributivity plays no role here. Even in the proof of the theorem, the full power of distributivity is not needed to obtain the decomposition of $\mathbf{L}$ into the direct product of other lattices. Later we will see how to decompose relatively complemented lattices of finite length.

A complemented distributive lattice is called a **Boolean lattice**. In Chapter 1, we defined Boolean algebras in such a way that complementation was a basic unary operation. Thus the relation between Boolean lattices and Boolean algebras is like that between groups treated as algebras with one operation–the group multiplication–and groups as we have introduced them in Chapter 1. Homomorphic images of Boolean lattices are again Boolean lattices and direct products of systems of Boolean lattices are also Boolean lattices. However, subalgebras of Boolean lattices are not generally Boolean lattices. For example, the only chains that are Boolean are those with one or two elements, but long chains are quite

common in most Boolean lattices. Theorem 2.57 can obviously be applied to finite Boolean lattices to obtain the following result:

COROLLARY 2.58. *Every finite Boolean lattice is isomorphic to a direct power of the two-element chain.*

Another way to formulate this corollary is the following: *Every finite Boolean lattice is isomorphic to the lattice of all subsets of some set, where the join of the two subsets is just their union and the meet of two subsets is just their intersection.*

This reformulation of the corollary is a simple consequence of the connection between sets and characteristic functions. Indeed, the members of the kth direct power of the two-element chain with elements 0 and 1 can be regarded as the characteristic functions defined on a k-element set. The correlation of subsets with their characteristic functions is an isomorphism from the lattice of subsets onto the direct power.

Neither the corollary nor its reformulation hold for arbitrary finite distributive lattices or for arbitrary infinite Boolean lattices in place of finite Boolean lattices. But it is possible to accommodate these lattices by giving up only a little of the power of the conclusion. It turns out that every finite distributive lattice is isomorphic, in a rather special way, to a sublattice of a direct power of the two-element chain (or, in the language of the reformulation, to a sublattice of the lattice of all subsets of some set, consisting of certain kinds of subsets.)

Let $\mathbf{J} = \langle J, \leq \rangle$ be any ordered set. An **order ideal** of $\mathbf{J}$ is just a subset of J that is closed downward. That is, $I \subseteq J$ is an order ideal of $\mathbf{J}$ iff for all a and b in J, if $a \in I$ and $b \leq a$, then $b \in I$. Let $\mathbf{L}$ be a lattice and let $J(\mathbf{L})$ be the set of all nonzero join irreducible elements of $\mathbf{L}$. The ordered set obtained by restricting the ordering of $\mathbf{L}$ to $J(\mathbf{L})$ is denoted by $\mathbf{J}(\mathbf{L})$, and the set of order ideals of $\mathbf{J}(\mathbf{L})$ is denoted by Ord $\mathbf{J}(\mathbf{L})$. Evidently, $\langle \text{Ord } \mathbf{J}(\mathbf{L}), \cap, \cup \rangle$ is a distributive lattice. **Ord** $\mathbf{J}(\mathbf{L})$ will denote this lattice. Also let $\text{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$ stand for the set of all isotone maps from the ordered set $\mathbf{J}^\partial$ into the two element chain $\mathbf{C}_2$. (Recall that the superscript $^\partial$ indicates the operation of forming the dual of an ordered set.) Evidently $\mathbf{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$ is a sublattice of the direct power of $\mathbf{C}_2^J$ of the two-element chain. The connection between **Ord** $\mathbf{J}(\mathbf{L})$ and $\mathbf{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$, where $\mathbf{J} = \mathbf{J}(\mathbf{L})$, is that the characteristic functions on the order ideals are exactly those isotone maps, and this correlation establishes an isomorphism.

The next theorem is due to Birkhoff [].

THEOREM 2.59 (The Representation Theorem for Finite Distributive Lattices)**.** *Let* $\mathbf{L}$ *be a finite distributive lattice and let* $\mathbf{J}$ *be the ordered set of nonzero join irreducible elements of* $\mathbf{L}$*. Then* $\mathbf{L} \cong \mathbf{Ord}\ \mathbf{J}(\mathbf{L}) \cong \mathbf{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$*. Moreover, each projection function on* $\mathbf{C}_2^J$ *maps* $\text{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$ *onto* C_2*.*

Proof. Our proof focuses on isotone maps rather than order ideals, so that the last sentence of the theorem can be easily handled. In view of the remarks preceding the theorem, the whole proof could be reformulated in terms of order ideals.

For each $a \in L$, define $f_a : J \to \{0,1\}$ by $f_a(b) = 1$ if $b \leq a$ and $f_a(b) = 0$ if $b \not\leq a$ for all $b \in J$. For each $a \in L$, f_a is easily seen to be isotone from $\mathbf{J}^\partial$ into $\mathbf{C_2}$. Define $F : L \to \mathrm{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$ by

$$F(a) = f_a \text{ for all } a \in L.$$

CLAIM 0: F is a homomorphism.

Let $a, b \in L$ and $c \in J$. Notice

$$c \leq a \vee b \text{ iff } c \leq a \text{ or } c \leq b,$$

because c is join prime, and

$$c \leq a \wedge b \text{ iff } c \leq a \text{ and } c \leq b.$$

With the help of these two observations, it is straightforward to argue that F is a homomorphism.

CLAIM 1: F is one-to-one.

Suppose $f_a = f_b$. Then the set of join irreducibles below a is the same as the set of join irreducibles below b. Since every element of a finite lattice is the join of the set of join irreducibles below it, we conclude that $a = b$.

CLAIM 2: F is onto $\mathrm{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$.

Let $h : J \to \{0,1\}$ be isotone (for $\mathbf{J}^\partial$!). So H is order reversing for the order $\leq$ on $\mathbf{L}$. Let $a = \bigvee\{c : h(c) = 1\}$. To see that $h = f_a$, observe that for $d \in J$, $d \leq a$ iff $h(d) = 1$, because d is join prime and h is order reversing.

CLAIM 3: Let $a \in J$ and let ρ_a be the associated projection function. Then ρ_a maps $\mathrm{Iso}(\mathbf{J}^\partial, \mathbf{C_2})$ onto C_2.

Notice that, for $b \in L$, $\rho_a(h) = \rho_a(f_b) = f_b(a) = 1$ or 0, depending on whether $a \leq b$. Since $a > 0$, we conclude that $\rho_a(f_a) = 1$ while $\rho_a(f_0) = 0$. Hence ρ_a is onto C_2. ■

The last sentence of the theorem is an assertion that the given embedding of $\mathbf{L}$ into the direct power of the two-element chain links $\mathbf{L}$ closely to each factor of the power. In the language to be introduced in Chapter 4, this sentence asserts that every finite distributive lattice is a subdirect power of the two-element chain.

Next, we extend this result to infinite distributive lattices. Let $\mathbf{L}$ be a lattice and I be an ideal of $\mathbf{L}$. I is said to be a **prime ideal** provided $a \wedge b \in I$ implies $a \in I$ or $b \in I$ for all $a, b \in L$. This is equivalent to saying that I is a meet prime element of Idl $\mathbf{L}$. In the proof above (join) primeness played a crucial role. We will replace it with primeness of ideals. In place of the fact that every element in a finite distributive lattice is the join of join irreducible elements, we will need to know that in any distributive lattice, every ideal is the intersection of prime ideals.

What we need is gathered in the next theorem, due to M. H. Stone [] and A. Tarski [].

THEOREM 2.60. *[The Prime Ideal Theorem for Distributive Lattices] Let* $\mathbf{L}$ *be a distributive lattice.*

i. *If I is an ideal of $\mathbf{L}$ and F is a filter of $\mathbf{L}$ such that I and F are disjoint, then there is a prime ideal J of $\mathbf{L}$ such that $I \subseteq J$ with J disjoint from F.*

ii. *Every ideal of $\mathbf{L}$ is the intersection of the prime ideals that include it.*

Proof. To prove (i), we let H be the collection of ideals containing I but disjoint from F. By Zorn's Lemma, pick a maximal member J of H. We need to demonstrate that J is prime. For the sake of obtaining a contradiction, suppose not. Pick $a,b \in L$ such that $a \wedge b \in J$ but neither a nor b belongs to J. By the maximality of J, $(J \vee I(a]) \cap F$ and $(J \vee I(b]) \cap F$ are both nonempty. So pick $c,d \in J$ such that $a \vee c$ and $b \vee d$ belong to F. Since F is a filter, $(a \vee c) \wedge (b \vee d) \in F$. By distributivity, $(a \vee c) \wedge (b \vee d) = (a \wedge b) \vee (a \wedge d) \vee (c \wedge b) \vee (c \wedge d)$, so this element belongs to J as well. This contradicts that J and F are disjoint.

To prove (ii), we let I be any ideal of $\mathbf{L}$. If $I = L$, there is little to prove. So we suppose that I is a proper deal. According to (i), for each $a \notin I$, there is a prime ideal J_a such that $I \subseteq J_a$ and $a \notin J_a$. Clearly, I is the intersection of all these prime ideals. ■

Let $\mathbf{L}$ be a distributive lattice and $P(\mathbf{L})$ be the set of all proper prime ideals of $\mathbf{L}$. $P(\mathbf{L})$ is ordered by set inclusion, so let $\mathbf{P}(\mathbf{L})$ denote this ordered set. The collection of all order ideals of $\mathbf{P}(\mathbf{L})$ is evidently a distributive lattice, with set union taken for the join and set intersection taken for the meet. Call this lattice $\mathbf{Ord\ P}(\mathbf{L})$. (Notice that the elements of Ord $\mathbf{P}(\mathbf{L})$ are collections of prime ideals of $\mathbf{L}$.) Once more, the connection between order ideals and their characteristic functions yields $\mathbf{Ord\ P}(\mathbf{L}) \cong \mathbf{Iso}(\mathbf{P}^\partial, \mathbf{C_2})$, where $\mathbf{P}$ is taken as $\mathbf{P}(\mathbf{L})$. This means that the theorem below, due to G. Birkhoff [] and M.H.Stone [], can be reformulated to give a representation of any distributive lattice $\mathbf{L}$ by sets under the operations of intersection and union.

THEOREM 2.61. *[The Representation Theorem for Distributive Lattices] Let $\mathbf{L}$ be any distributive lattice and let $\mathbf{P}$ be the set of all proper prime ideals of $\mathbf{L}$ ordered by set inclusion. $\mathbf{L}$ can embed into $\mathbf{Iso}(\mathbf{P}^\partial, \mathbf{C_2})$ in such a way that the projection functions, restricted to the image of $\mathbf{L}$, are onto $\mathbf{C}_2$.*

Proof. For each $a \in L$, define $f_a : P \to \{0,1\}$ by $f_a(I) = 0$ iff $a \in I$. It is easy to check that $f_a \in \mathrm{Iso}(\mathbf{P}^\partial, \mathbf{C_2})$ by $F(a) = f_a$ for all $a \in L$.
CLAIM 0: F is a homomorphism.

Let I be a prime ideal and $a,b \in L$. Observe that

$$a \vee b \in I \text{ iff } a \in I \text{ and } b \in I$$

and, because I is prime,

$$a \wedge b \in I \text{ iff } a \in I \text{ or } b \in I.$$

These two observations lead immediately to the conclusion that F is a homomorphism.
CLAIM 1: F is one-to-one.

Suppose $F(a) = F(b)$. So $f_a(I) = f_b(I)$ for all proper prime ideals I. This means that a and b belong to exactly the same prime ideals. By the Prime Ideal Theorem (ii), $I(a] = I(b]$. Therefore $a = b$, and so F is one-to-one.
CLAIM 2: Let I belong to P and let ρ_I be the associated projective function. Then $\rho_I \circ F$ maps L onto C_2.

Just note that $(\rho_I \circ F)(a) = \rho_I(F(a)) = \rho_I(f_a) = f_a(I) = 0$ or 1, depending on whether a belongs to I or not. ■

As with the finite case, we also obtain the result that every distributive lattice is a subdirect power of the two-element chain. The line of reasoning yielding the Representation Theorem for Distributive Lattices can be easily modified for Boolean algebras, yielding the following theorem due to M. H. Stone [].

THEOREM 2.62 (The Representation Theorem for Boolean Algebras)**.** *Every finite Boolean algebra is isomorphic to a direct power of the two-element Boolean algebra. Every Boolean algebra is embeddable into a direct power of the two-element Boolean algebra in such a way that each projection function of the direct power maps the images of the Boolean algebra onto the two-element boolean algebra.*

Of course, another version of this theorem is the statement that every Boolean algebra is isomorphic to a Boolean algebra composed of subsets of a set with the operations of set intersection, set union, and set complementation.

The story of representation of distributive lattices and of Boolean algebras does not end here. In fact, underlying these representation theorems are more powerful results. Notice that, in the proofs above, each distributive lattice was correlated, in a natural way, with an ordered set. (For finite distributive lattices, this set was the set of all nonzero join irreducible elements, whereas for infinite distributive lattices, this was the collection of all proper prime ideals.) In turn, each ordered set was correlated with a distributive lattice–its lattice of order ideals. Unfortunately, for infinite distributive lattices $\mathbf{L}$, we obtain only the conclusion that $\mathbf{L}$ can be embedded into $\mathbf{Ord}\,\mathbf{P}(\mathbf{L})$. In general, $\mathbf{P}(\mathbf{L})$ will not determine $\mathbf{L}$ up to isomorphism. But it turns out that $\mathbf{P}(\mathbf{L})$ can be given a topology, and the resulting topological ordered set does determine $\mathbf{L}$ up to isomorphism. Indeed, not only will there be a natural one-to-one correlation of distributive lattices with certain kinds of ordered topological spaces, but, roughly speaking, this correlation matches lattice homomorphisms with continuous maps (in the reverse direction), lattice ideals with open subsets, and lattice filters with closed subsets. The reader may well imagine that many results about distributive lattices and Boolean algebras can now be deduced using the concrete setting of sets. The deeper correlation with topological ordered sets allows topological tools to be brought to bear on lattice-theoretic problems. These correlations, known as the Stone and Priestley dualities, are taken up in a later volume.

Exercises 2.63

1. Let $\mathbf{L}$ be the lattice of natural numbers with the meet of the two numbers

taken to be their greatest common divisor and their join taken to be their least common multiple. Prove that $\mathbf{L}$ is a distributive lattice.

2. Prove that the complemented elements of any distributive lattice comprise a subuniverse of that lattice.

3. Prove that in any distributive lattice with a 0, direct join isotopy coincides with equality and that a set is directly join independent iff every pair of distinct elements of it meet to 0. Finally, prove that a finite directly join independent set of which 0 is not a member must be join irredundant. Thus the Direct Join Decomposition Theorem for distributive lattices is a corollary of Theorem 2.55.

4. **a.** Let $\mathbf{L}$ be a finite distributive lattice and let $J(\mathbf{L})$ be the set of join irreducible elements of $\mathbf{L}$. Prove that the length of any maximal chain in $\mathbf{L}$ is $|J(\mathbf{L})|$.

b. Prove that in any finite distributive lattice, the number of join irreducible elements and the number of meet irreducible elements is the same.

5. (Nachbin []) Let $\mathbf{L}$ be a distributive lattice. Prove that $\mathbf{L}$ is relatively complemented iff every proper prime ideal of $\mathbf{L}$ is a maximal ideal. [Hint: The Prime Ideal Theorem is useful in establishing both implications.]

***6.** Prove that every finitely generated distributive lattice is finite.

7. Prove that every uniquely complemented modular lattice is distributive. [Hint: Pick $x, y, z \in L$ such that $x \wedge z = y \wedge z$ and $x \vee z = y \vee z$. Pick a complement u of $x \wedge z$ in $I[0, z]$ and then a complement v of $x \vee z$ in $I[u, 1]$. Prove that v is a complement of both x and y.]

8. Prove that in a distributive lattice, no interval can be projective with a proper subinterval of itself.

9. **a.** Construct two countably infinite nonisomorphic Boolean lattices $\mathbf{L}_0$ and $\mathbf{L}_1$ such that $\mathbf{Ord\,P}(\mathbf{L_0}) \cong \mathbf{Ord\,P}(\mathbf{L_1})$.

b. Prove that if $\mathbf{L}_0$ and $\mathbf{L}_1$ are finite distributive lattices such that $\mathbf{J}(\mathbf{L_0}) \cong \mathbf{J}(\mathbf{L_1})$, then $\mathbf{L}_0 \cong \mathbf{L}_1$.

10. Let $\mathbf{L}_0$ and $\mathbf{L}_1$ be finite distributive lattices and let $\mathbf{J}_0$ and $\mathbf{J}_1$ be their respective ordered sets of nonzero join irreducible elements. Let h be a homomorphism from $\mathbf{L}_0$ into $\mathbf{L}_1$ such that h preserves top and bottom elements. Let $J(h)$ be the map with domain J_1 defined by $J(h)(b) = \bigwedge \{x : h(x) \geq b\}$ for all $b \in J_1$.

a. Prove that $J(h)$ is an isotone map from $\mathbf{J}_1$ into $\mathbf{J}_0$.

b. Prove that $J(h)$ is onto $\mathbf{J}_0$ iff h is one-to-one.

2.6 Congruence Relations on Lattices

For every lattice $\mathbf{L}$ the lattice $\mathbf{Con\,L}$ is a distributive algebraic lattice, according to Theorem 2.50 and Corollary 1.23. In this section, we will examine the congruences of lattices in general and investigate $\mathbf{Con\,L}$ for particular kinds of lattices $\mathbf{L}$.

Let $\mathbf{L}$ be a lattice and let $\theta \in \mathbf{Con\,L}$. It is a simple but useful observation that the congruence classes modulo θ are convex sublattices of $\mathbf{L}$. It is also useful to notice that if $a\ \theta\ b$ iff $a \wedge b\ \theta\ a \vee b$ for all $a, b \in L$. These remarks mean that θ is determined by the pairs (a,b) belonging to θ with $a \leq b$.

THEOREM 2.64. *Let* $\mathbf{L}$ *be a lattice and* θ *be a reflexive binary relation on* L. θ *is a congruence relation of* $\mathbf{L}$ *iff for all* $a, b, c \in L$,

i. $a\ \theta\ b$ *iff* $a \wedge b\ \theta\ a \vee b$.

ii. *If* $a \leq b \leq c$, $a\ \theta\ b$, *and* $b\ \theta\ c$, *then* $a\ \theta\ c$.

iii. *If* $a \leq b$ *and* $a\ \theta\ b$, *then* $a \wedge c\ \theta\ b \wedge c$ *and* $a \vee c\ \theta\ b \vee c$.

Proof. All three of these properties are clearly possessed by congruence relations, so we will only deal with the converse. Property (i) and the commutativity of $\wedge$ and $\vee$ yield the symmetry of θ. According to (iii), if $a \leq c \leq b$ and $a\ \theta\ b$, then $a\ \theta\ c$ and $c\ \theta\ b$–a kind of convexness. To see the transitivity of θ, suppose $a\ \theta\ b\ \theta\ c$. It will follow from this convexness that $a\ \theta\ c$, provided we first prove that $(a \wedge b \wedge c)\ \theta\ (a \vee b \vee c)$. Observe that we have the following string of inclusions and that the formulas resulting from replacing $\leq$ by θ in these inclusions are also true:

$$a \wedge b \wedge c \leq (a \vee b) \wedge (b \wedge c) = b \wedge c \leq b \vee c = (a \wedge b) \vee (b \vee c) \leq a \vee b \vee c.$$

Therefore, two applications of (ii) yield the desired conclusion. Hence θ is transitive and, thus, an equivalence relation.

In checking the Substitution Property, we will only deal with $\vee$, leaving the similar argument for $\wedge$ aside. Since transitivity is already verified, we need only show that $a\ \theta\ b$ implies $a \vee c\ \theta\ b \vee c$. Property (i) tells us that $a \wedge b\ \theta\ a \vee b$. From property (iii), we get $(a \wedge b) \vee c\ \theta\ a \vee b \vee c$. But $a \vee c$ and $b \vee c$ belong to $I[(a \wedge b) \vee c, a \vee b \vee c]$. So the convexity yields $a \vee c\ \theta\ b \vee c$. ■

Intuitively, $\mathbf{L}/\theta$ is formed by collapsing the congruence classes, which are always convex, to points. In a lattice, any congruence relation that collapses a given interval may of necessity collapse others. Consider the transposed intervals $I[a \wedge b, a]$ and $I[b, a \vee b]$. If $a\ \theta\ a \wedge b$, then $b = b \vee (a \wedge b)\ \theta\ b \vee a$, so collapsing $I[a \wedge b, a]$ to a point forces $I[b, a \vee b]$ to collapse to a point. More generally, if θ collapses an interval I to a point, then it must also collapse to points all intervals projective with I; it is also clear that the congruence must collapse all intervals projective with subintervals of the original interval. Generally, however, this does not give a satisfactory account of the situation. The reader should easily verify that collapsing any nontrivial interval of $\mathbf{M}_3$ collapses the whole lattice to

a point. To obtain a characterization of congruences in lattices, we modify the notion of projectivity.

DEFINITION 2.65. Let $\mathbf{L}$ be a lattice and let $I[a,b]$ and $I[c,d]$ be intervals in $\mathbf{L}$. We say that $I[a,b]$ **transposes weakly down** into $I[c,d]$ iff $b = a \vee d$ and $c \leq a$. We say that $I[a,b]$ **transposes weakly up** into $I[c,d]$ iff $a = b \wedge c$ and $b \leq d$. We denote these weak transposes, respectively, by $I[a,b] \searrow_w I[c,d]$ and $I[a,b] \nearrow_w I[c,d]$. Finally, we say that $I[a,b]$ is **weakly projective** into $I[c,d]$ iff there is a finite sequence $I[a_0,b_0] = I[a,b]$, $I[a_1,b_1], \ldots, I[a_n,b_n] = I[c,d]$ such that $I[a_i,b_i]$ transposes weakly into $I[a_{i+1},b_{i+1}]$ for all $i < n$.

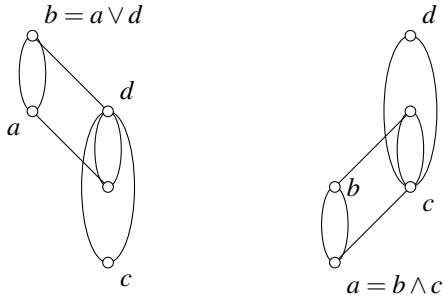


Figure 2.14:

The relation "weakly projective into" between intervals is transitive and reflexive, but it is not symmetric. Recall that $\mathrm{Cg}^{\mathbf{L}}(c,d)$ denotes the principal congruence of $\mathbf{L}$ generated by the pair $\langle c,d \rangle$.

THEOREM 2.66. *(R.P. Dilworth [1950]). Let* $\mathbf{L}$ *be a lattice with* $a \leq b$ *and* $c \leq d$ *in* $\mathbf{L}$. *Then* $\langle a,b \rangle \in \mathrm{Cg}^{\mathbf{L}}(c,d)$ *iff there is a finite sequence*

$$a = e_0 \leq e_1 \leq e_2 \leq \cdots \leq e_n = b$$

such that $I[e_i,e_{i+1}]$ *is weakly projective into* $I[c,d]$ *for all* $i < n$.

Proof. Let θ be the binary relation defined on L by $u \; \theta \; v$ iff there is a finite sequence as described above, where $a = u \wedge v$ and $b = u \vee v$. It is clear from the definition of weak projectivity that

$$\langle c,d \rangle \in \theta \subseteq \mathrm{Cg}^{\mathbf{L}}(c,d).$$

We shall show that $\theta = \mathrm{Cg}^{\mathbf{L}}(c,d)$. It suffices to show that θ is a congruence relation. Since θ is obviously reflexive, we will do this by verifying the three properties listed in Theorem 2.64. Properties (i) and (ii) are immediate. For the

purposes of property (iii), let $a = e_0 \leq \cdots \leq e_n = b$ be as described above and let $f \in L$. Note that $a \wedge f = e_0 \wedge f \leq \cdots \leq e_n \wedge f = b \wedge f$ and that $I[e_i \wedge f, e_{i+1} \wedge f]$ is weakly projective into $I[e_i, e_{i+1}]$ and hence into $I[c,d]$. Joins work similarly. ■

While this characterization of principal congruences in lattices is useful, it can be sharpened in certain classes of lattices. A lattice **L** is said to have the **projectivity property** iff whenever $I[a,b]$ is weakly projective into $I[c,d]$, then $I[a,b]$ is actually projective with a subinterval of $I[c,d]$. For lattices with this property, principal congruences can be characterized as in Theorem 2.66, except that each $I[e_i, e_{i+1}]$ is projective with a subinterval of $I[c,d]$.

EXAMPLE 2.67. A lattice without the projective property.

Proof. Let **L** be the lattice diagrammed in Figure 2.15. The interval $I[b,a]$ transposes weakly down into $I[0,c]$, which transposes up to $I[f,e]$. So $I[b,a]$ is weakly projective into $I[f,e]$. But b is meet irreducible, so it is impossible for $I[b,a]$ to transpose up to any other interval. Since c is the only element that gives a when joined with b, $I[d,c]$ is the only interval to which $I[b,a]$ transposes. Similarly, since c is join irreducible and b is the only element that gives d when met with c, we find that $I[b,a]$ is the only interval to which $I[d,c]$ transposes. Hence $I[b,a]$ is not projective with $I[f,e]$ (or any of its subintervals), even though it is weakly projective into $I[f,e]$. ■

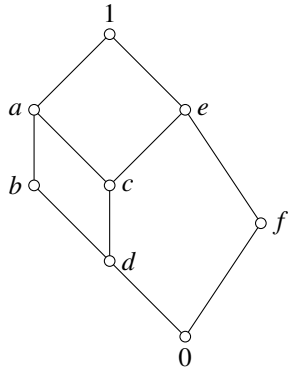


Figure 2.15:

THEOREM 2.68. *Every modular lattice and every relatively complemented lattice has the projectivity property.*

Proof. First let **L** be relatively complemented. Suppose that $I[a',b']$ transposes up to $I[c,d]$ weakly and that $a' \leq a \leq b \leq b'$. We argue that $I[a,b]$ is projective with a subinterval of $I[c,d]$. Let a^* be a complement of a in $I[a',b]$. Now just

note that

$$a' = a \wedge a^*$$
$$b = a \vee a^*$$

so $I[a,b] \searrow I[a',a^*]$ and

$$a' = a^* \wedge c$$

since $a' = b' \wedge c$ and $a' \leq a^* \leq b'$. So $I[a',a^*] \nearrow I[c,a^* \vee c]$. Therefore $I[a,b]$ is projective with $I[c,a^* \vee c]$, a subinterval of $I[c,d]$. The proof can now be completed by a straightforward induction on the length of the chain of weak projectivity.

The case when $\mathbf{L}$ is modular is an easy consequence of Dedekind's Transposition Principle. ■

In distributive lattices, principal congruence relations have an especially simple form.

THEOREM 2.69. *Let* $\mathbf{L}$ *be a distributive lattice and let* $I[c,d]$ *be an interval in* $\mathbf{L}$. $\langle a,b\rangle \in \mathrm{Cg}^{\mathbf{L}}(c,d)$ *iff* $a \wedge c = b \wedge c$ *and* $a \vee d = b \vee d$ *for all* $a,b \in L$.

Proof. Let $\theta = \{\langle a,b\rangle : a \wedge c = b \wedge c \text{ and } a \vee d = b \vee d\}$. First we prove that θ is a congruence relation. θ is clearly an equivalence relation on L. Suppose $a\ \theta\ b$. Then

$$\begin{aligned}(a \vee e) \wedge c &= (a \wedge c) \vee (e \wedge c)\\ &= (b \wedge c) \vee (e \wedge c)\\ &= (b \vee e) \wedge c\end{aligned}$$

and evidently

$$(a \vee e) \vee d = (b \vee e) \vee d.$$

So $(a \vee e)\ \theta\ (b \vee e)$ for all $e \in L$. Meets behave in a similar fashion. Therefore, θ is a congruence relation.

Since $\langle c,d\rangle \in \theta$, all that remains is to prove that $\theta \subseteq \phi$ for every congruence ϕ which collapses c and d. So suppose $a\ \theta\ b$. Thn

$$\begin{aligned}a = a \wedge (a \vee b) = a \wedge (b \vee d) &= (a \wedge b) \vee (a \wedge d)\\ &\ \phi\ (a \wedge b) \vee (a \wedge c)\\ &= (a \wedge b) \vee (b \wedge c)\\ &= b \wedge (a \vee c)\\ &\ \phi\ b \wedge (a \vee d)\\ &= b \wedge (b \vee d) = b\end{aligned}$$

and so $\theta \subseteq \phi$, as desired. ■

The next theorem presents a distinctive property of the congruence lattices of lattices. Roughly speaking, it says that the congruences on the direct product of two lattices can be taken apart into congruences on the factor lattices.

THEOREM 2.70. *If* $\mathbf{L}_0$ *and* $\mathbf{L}_1$ *are lattices, then*

$$\mathbf{Con}\,\mathbf{L}_0 \times \mathbf{L_1} \cong \mathbf{Con}\,\mathbf{L}_0 \times \mathbf{Con}\,\mathbf{L}_1.$$

Proof. Actually, part of this theorem is true for algebras in general:

$$\mathbf{Con}\,\mathbf{L}_0 \times \mathbf{Con}\,\mathbf{L}_1 \text{ is embeddable in } \mathbf{Con}\,\mathbf{L}_0 \times \mathbf{L_1}.$$

Define $h : \text{Con } \mathbf{L}_0 \times \text{Con } \mathbf{L}_1 \to \text{Con } \mathbf{L}_0 \times \mathbf{L_1}$ by

$$\langle a_0, a_1\rangle \; h(\theta_0, \theta_1) \; \langle b_0, b_1\rangle \text{ iff } a_0 \; \theta_0 \; b_0 \text{ and } a_1 \; \theta_1 \; b_1.$$

We leave it to the reader to verify that h is actually an embedding. For algebras in general, this embedding can fail to be onto Con $\mathbf{L}_0 \times L_1$, but for lattices it is always onto. To see this, let $m(x,y,z)$ stand for the expression $(x \vee y) \wedge z$, $p(x,y)$ stand for $x \wedge y$, and $q(x,y)$ stand for $x \vee y$. In any lattice, the following equalities then hold for all x and y:

$$\begin{aligned} m(x,y,y) &= y \\ p(x,y) &= p(y,x) \\ q(x,y) &= q(y,x) \\ m\,(x, p(x,y), q(x,y)) &= x. \end{aligned}$$

To see that h is onto Con $\mathbf{L}_0 \times L_1$, let $\theta \in \text{Con } \mathbf{L}_0 \times L_1$. Define θ_0 on L_0 by

$$a_0 \; \theta_0 \; b_0 \text{ iff } \langle a,c\rangle \; \theta \; \langle b,c\rangle \text{ for some } c \in L_1.$$

CLAIM: $a \; \theta_0 \; b$ iff $\langle a,c\rangle \; \theta \; \langle b,c\rangle$ for all $c \in L_1$.

Suppose $\langle a,d\rangle \; \theta \; \langle b,d\rangle$ and that $c \in L_1$. Now, $\langle p(a,b),c\rangle \; \theta \; \langle p(a,b),c\rangle$ and $\langle q(a,b),c\rangle \; \theta \; \langle q(a,b),c\rangle$. Hence

$$\begin{aligned} \langle a,c\rangle &= \langle m\,(a, p(a,b), q(a,b))\,, m(d,c,c)\rangle \\ &= m\,(\langle a,d\rangle, \langle p(a,b),c\rangle, \langle q(a,b),c\rangle) \\ &\;\theta\; m\,(\langle b,d\rangle, \langle p(a,b),c\rangle, \langle q(a,b),c\rangle) \\ &= \langle m\,(b, p(a,b), q(a,b))\,, m(d,c,c)\rangle \\ &= \langle m\,(b, p(b,a), q(b,a))\,, m(d,c,c)\rangle \\ &= \langle b,c\rangle. \end{aligned}$$

Using the claim, it is easy to prove that $\theta_0 \in \text{Con } \mathbf{L}_0$. In a similar way, we define $\theta \in \text{Con } \mathbf{L}_1$. Now to see that $h(\theta_0, \theta_1) = \theta$, just follow the implications below:

$$\begin{aligned} \langle a_0, a_1\rangle \; h(\theta_0, \theta_1) \; \langle b_0, b_1\rangle &\implies a_0 \; \theta_0 \; b_0 \text{ and } a_1 \; \theta_1 \; b_1 \\ &\implies \langle a_0, c\rangle \; \theta \; \langle b_0, c\rangle \text{ and } \langle d, a_1\rangle \; \theta \; \langle d, b_1\rangle \\ &\qquad \text{for all } d \in L_0 \text{ and all } c \in L_1 \\ &\implies \langle a_0, a_1\rangle \; \theta \; \langle b_0, a_1\rangle \; \theta \; \langle b_0, b_1\rangle \\ &\implies \langle a_0, a_1\rangle \; \theta \; \langle b_0, b_1\rangle. \end{aligned}$$

Hence $h(\theta_0, \theta_1) \subseteq \theta$.

Now suppose $\langle a_0, a_1\rangle \ \theta \ \langle b_0, b_1\rangle$. So

$$\begin{aligned}\langle a_0, a_1\rangle &= m(\langle a_0, a_1\rangle, \langle p(a_0, b_0), a_1\rangle, \langle q(a_0, b_0), a_1\rangle)\\ &\ \theta\ m(\langle b_0, b_1\rangle, \langle p(a_0, b_0), a_1\rangle, \langle q(a_0, b_0), a_1\rangle)\\ &= m(\langle b_0, b_1\rangle, \langle p(b_0, a_0), a_1\rangle, \langle q(b_0, a_0), a_1\rangle)\\ &= \langle m(b_0, p(b_0, a_0), q(b_0, a_0)), m(b_1, a_1, a_1)\rangle\\ &= \langle b_0, a_1\rangle.\end{aligned}$$

Thus $a_0 \ \theta_0 \ b_0$. Similarly, $a_1 \ \theta_1 \ b_1$. Therefore $h(\theta_0, \theta_1) = \theta$. ■

This theorem holds for a wider class of algebras than lattices. In fact, the crucial requirement is the existence of several expressions [$m(x, y, z)$, $p(x, y)$ and $q(x, y)$] built from variables and the fundamental operations for which certain equalities hold in both factor algebras. This theorem supplies us with a strategy for describing **Con L** for certain lattices **L**. In the first step, we try to write **L**, up to isomorphism, as a direct product of lattices that cannot themselves be written as direct products of other lattices. The second step then consists of analyzing **Con L** in the case that **L** cannot be factored further. Let **A** be an algebra. **A** is **directly indecomposable** iff **A** has more than one element and if $\mathbf{A} \cong \mathbf{B} \times \mathbf{C}$ then either **B** has only one element or **C** has only one element.

Carrying out the first step in our strategy is easy, if we impose some kind of finiteness condition on **L**.

THEOREM 2.71. *Every lattice of finite height is isomorphic to a direct product of finitely many directly indecomposable lattices.* ■

This theorem can be proven by a straightforward induction on height. The proof is left in the hands of the reader.

The second step in the strategy requires more work and additional hypotheses to obtain useful conclusions. We call an algebra **A simple** iff Con **A** has exactly two elements. Every simple algebra is directly indecomposable (just consider the kernels of the projective functions), but it is not hard to invent finite modular lattices that are directly indecomposable (by virtue of having a prime number of elements) but fail to be simple.

THEOREM 2.72. *(R.P. Dilworth [1950].) Every directly indecomposable relatively complemented lattice of finite height is simple.*

Proof. Let **L** be a directly indecomposable relatively complemented lattice of finite height. Let θ be a congruence on **L** different from the identity relation. Pick $u < v$ so that $u \ \theta \ v$. Let u^* be a complement of u in $I[0, v]$. Now observe that

$$0 = u \wedge u^* \ \theta \ v \wedge u^* = (u \vee u^*) \wedge u^* = u^* \neq 0.$$

Thus, $\{x : 0 \ \theta \ x \neq 0\}$ is not empty. Since **L** has finite length, in view of Theorem 2.6 pick a to be a maximal element of $\{x : 0 \ \theta \ x\}$. Since this set is an ideal,

it is easy to see that a is the largest element in the ideal. Theorem 2.57 suggests that $\mathbf{L}$ might be decomposable as $I(a] \times I[a)$. That theorem requires that a have a complement and that $\mathbf{L}$ be distributive. We are still able to carry out roughly the same argument, the difficult point being to establish enough "distributivity."
CLAIM 0: $x \,\theta\, y$ iff $(x \wedge y) \vee a' = x \vee y$ for some $a' \leq a$.

Suppose that $x \,\theta\, y$. Let a' be a complement of $x \wedge y$ in $I[0, x \vee y]$. Then $a' \wedge (x \vee y) \,\theta\, a' \wedge (x \wedge y) = 0$. So $a' \leq a$ by the maximality of a. The converse is immediate.
CLAIM 1: $a \vee (x \wedge y) = (a \vee x) \wedge (a \vee y)$ for all $x, y \in L$.

Observe that

$$x \,\theta\, a \vee x$$
$$y \,\theta\, a \vee y$$

and so

$$x \wedge y \,\theta\, (a \vee x) \wedge (a \vee y).$$

By Claim 0, pick $a' \leq a$ so that

$$((x \wedge y) \wedge (a \vee x) \wedge (a \vee y)) \vee a' = (x \wedge y) \vee ((a \vee x) \wedge (a \vee y)).$$

Using the absorption axioms, we obtain

$$(x \wedge y) \vee a' = (a \vee x) \wedge (a \vee y).$$

Finally, using this equation, we obtain

$$a \vee (x \wedge y) \leq (a \vee x) \wedge (a \vee y) = a' \vee (x \wedge y) \leq a \vee (x \wedge y)$$

and therefore $a \vee (x \wedge y) = (a \vee x) \wedge (a \vee y)$. Thus, a possesses some distributivity in $\mathbf{L}$. We need to know that a satisfies the dual of Claim 1 as well. To accomplish this, we will characterize the property in Claim 1 in a selfdual way. Call an element b **distributive** iff $b \vee (x \wedge y) = (b \vee x) \wedge (b \vee y)$ for all $x, y \in L$.
CLAIM 2: b is distributive iff no nontrivial subinterval of $I(b]$ is projective with a nontrivial subinterval of $I[b)$.

Suppose first that b is distributive and that $I[x, y]$ is a subinterval of $I(b]$ and $I[u, v]$ is a subinterval of $I[b)$ such that $I[x, y]$ and $I[u, v]$ are projective. Define $\phi = \{\langle z, w\rangle : z \vee b = w \vee b\}$. Because b is distributive, ϕ is a congruence relation on $\mathbf{L}$. Now $x \,\phi\, y$, so $u \,\phi\, v$. But this means that $u = u \vee b = v \vee b = v$. Therefore, $I[u, v]$ is trivial and since $I[x, y]$ is projective with $I[u, v]$, it follows that $I[x, y]$ is trivial, too.

For the converse, suppose that b is not distributive. Pick x and y so that $b \vee (x \wedge y) < (b \vee x) \wedge (b \vee y)$. Let $\psi = \mathrm{Cg}^{\mathbf{L}}(0, b)$. Then $b \vee (x \wedge y) \,\psi\, (b \vee x) \wedge (b \vee y)$. Use Theorem 2.66 to pick a sequence

$$b \vee (x \wedge y) = e_0 \leq e_1 \leq \cdots \leq e_n = (b \vee x) \wedge (b \vee y)$$

such that $I[e_i, e_{i+1}]$ is weakly projective into $I(b]$ for each $i < n$. Since $b \vee (x \wedge y) < (b \vee x) \wedge (b \vee y)$, choose $j < n$ with $e_j < e_{j+1}$. So $I[e_j, e_{j+1}]$ is a nontrivial subinterval of $I[b)$ that is weakly projective into $I(b]$. By Theorem 2.68, $I[e_j, e_{j+1}]$ is projective with a subinterval of $I(b]$, as desired.

CLAIM 3: $a \wedge (x \vee y) = (a \wedge x) \vee (a \wedge y)$ for all $x, y \in L$.
CLAIM 4: If $a \wedge x = a \wedge y$ and $a \vee x = a \vee y$, then $x = y$.

Observe that

$$\begin{aligned} x \;\theta\; a \vee x \\ x \wedge y \;\theta\; (a \vee x) \wedge y \\ x \wedge \;\theta\; (a \vee y) \wedge y = y. \end{aligned}$$

By Claim 0, pick $a' \leq a$ so that

$$((x \wedge y) \wedge y) \vee a' = (x \wedge y) \vee y = y.$$

Therefore $a' \leq y$ and $a' \leq y \wedge a = x \wedge a \leq x$. This means that $a' \leq x \wedge y$. Consequently, $x \wedge y = y$. By a similar argument, $x \wedge y = x$. Therefore $x = y$.

Now define $h : L \to I[a)$ by $h(x) = (x \wedge a, x \vee a)$. Easy computations using Claims 1 and 3 reveal that h is a homomorphism. Claim 4 is virtually the statement that h is one-to-one. The fact that h is onto $I(a] \times I[a)$ follows as in Theorem 2.57, using Claims 1 and 3 above in place of the distributivity of **L**. Since **L** is directly indecomposable and $a \neq 0$, we deduce that $I[a)$ has only one element. Therefore $a = 1$ and θ is $L \times L$. Hence **L** has just two congruences; it is simple. ■

COROLLARY 2.73. *Every relatively complemented lattice of finite height is isomorphic to a direct product of simple relatively complemented lattices of finite height.*

G. Birkhoff [] and K. Menger [] had earlier established that every complemented finite dimensional lattice is isomorphic to a direct product of simple lattices. The complemented finite dimensional simple modular lattices turned out to be the two-element chain and the subspace lattices of nondegenerate finite dimensional projective geometries. See §4.8 for an account of this important result.

Combined with Theorem 2.70, Corollary 2.73 yields the following corollary.

COROLLARY 2.74. *(R.P. Dilworth [1950].) If* **L** *is a relatively complemented lattice of finite height, then* **Con L** *is a Boolean lattice.*

The corollary can be drawn as well from the following theorem.

THEOREM 2.75. *If* **L** *is a lattice with the finite chain condition and the projectivity property, then* **Con L** *is a Boolean lattice.*

Proof. First suppose $a \prec b$ in **L**. Let $c < d$ and $\langle c, d \rangle \in \mathrm{Cg}^{\mathbf{L}}(a, b)$. According to Theorem 2.66, pick a finite sequence $c = e_0 \leq e_1 \leq \cdots \leq e_n = d$ such that $I[e_i, e_{i+1}]$ is projective with a subinterval of $I[a, b]$. Since $a \prec b$, there are no proper subintervals, and therefore $I[a, b]$ is projective with some subinterval $I[e_i, e_{i+1}]$ of $I[c, d]$. Again by Theorem 2.66, we have $\langle a, b \rangle \in \mathrm{Cg}^{\mathbf{L}}(c, d)$, so

$\mathrm{Cg}^{\mathbf{L}}(a,b)$ is an atom in **Con L**. Since **L** has the finite chain condition, there is a finite maximal chain from 0 to 1; let $0 = d_0 \prec d_1 \prec \cdots \prec d_n = 1$ be one such chain. Apparently,

$$\mathrm{Cg}^{\mathbf{L}}(0,1) = \mathrm{Cg}^{\mathbf{L}}(d_0,d_1) \vee \mathrm{Cg}^{\mathbf{L}}(d_1,d_2) \vee \cdots \vee \mathrm{Cg}^{\mathbf{L}}(d_{n-1},d_n).$$

This means that **Con L** is a bounded distributive lattice in which the top element is the join of finitely many atoms. By Theorem 2.40, **Con L** is relatively complemented. Therefore **Con L** is a Boolean lattice. ■

COROLLARY 2.76. *If* **L** *is a modular lattice of finite height, then* **Con L** *is Boolean.*

Exercises 2.77

1. Prove that if $a \prec b$ in the lattice **L** and $\theta \in$ **Con L**, then either $a/\theta = b/\theta$ or $a/\theta \prec b/\theta$ in $\mathbf{L}/\theta$.

2. Prove that every lattice of finite height is isomorphic to a direct product of finitely many directly indecomposable lattices of finite height. Does a similar assertion hold for lattices satisfying the ascending chain condition?

3. Give an example of a finite directly indecomposable modular lattice that is not simple.

4. Let **L** be a lattice. For each ideal I of **L**, define

$$\Theta(I) = \{\langle a,b\rangle : a \vee c = b \vee c \text{ for some } c \in I\}$$

and for each congruence $\theta \in$ Con **L**, define

$$I(\theta) = \{a : a/\theta \leq b/\theta \text{ in } \mathbf{L}/\theta \text{ for all } b \in L\}.$$

a. Prove that if **L** is distributive and I is an ideal of **L**, then $\Theta(I) \in$ Con **L**.

b. Prove that $I(\theta) =$ Idl **L**, for all $\theta \in$ Con **L**.

c. Prove that the following are equivalent:

i. **L** is distributive.

ii. For each ideal I of **L**, $\Theta(I) \in$ Con **L** and $I(\Theta(I)) = I$.

iii. Every ideal of **L** has the form $I(\theta)$ for some $\theta \in$ Con **L**.

5. Let **L** be a lattice and $a \in L$. a is said to be **standard** iff

$$c \wedge (a \vee b) = (c \wedge a) \vee (c \wedge b) \text{ for all } c,b \in L.$$

The element a is said to be **neutral** iff

$$(a \wedge b) \vee (b \wedge c) \vee (c \wedge a) = (A \vee b) \wedge (b \vee c) \wedge (c \vee a) \text{ for all } c, b \in L.$$

Recall that a is said to be distributive iff

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \text{ for all } b, c \in L.$$

Thus an element is called distributive, standard, or neutral, depending on whether a certain part of the distributive law is true when that element is present.

a. Prove that a is distributive iff $\{\langle b, c\rangle : a \vee b = a \vee c\}$ is a congruence relation on $\mathbf{L}$.

b. Prove that a is standard iff a is distributive and for all $b, c \in L$, if $a \wedge b = a \wedge c$ and $a \vee b = a \vee c$, then $b = c$.

c. Prove that a is neutral iff for any $b, c \in L$ the sublattice generated by $\{a, b, c\}$ is distributive.

d. Prove that every neutral element is standard and that every standard element is distributive.

e. Prove that if $\mathbf{L}$ is modular, then every distributive element is neutral.

f. Prove that the set of neutral elements of $\mathbf{L}$ is the intersection of the maximal distributive sublattices of $\mathbf{L}$.

g. Prove that if $\mathbf{L}$ is a complemented modular lattice then a is neutral iff a has a unique complement.

C H A P T E R T H R E E

Unary and Binary Operations

3.1 Introduction

In the two previous chapters, we introduced the reader to some of the fundamental kinds of algebras with which we will be dealing throughout this work, such as lattices, semilattices, and Boolean algebras. Before turning to the formal presentation of the basics in Chapter 4, we devote this chapter to some other kinds of naturally occurring algebras. Such examples are central and important, because they help provide the diversity that is necessary for the discovery of general results in algebra.

The idea of an **algebra** allows arbitrarily many operations of arbitrary rank, and yet, as we remarked at the beginning of Chapter 1, the surprising fact is that all the classical algebras are built with unary and binary operations, especially binary ones. Moreover, except for **R**-modules, the classical algebras require only one or two binary operations and usually no unary operations. (For each r in the ring **R** of scalars, **R**-modules have one operation f_r on multiplication by r.) The same is true for the lesser-known algebras we will describe in this chapter. Experience has thus shown that almost all of the diversity of the theory of algebras already occurs for one or two binary operations. In this chapter, we provide a respresentative sampling of this diversity and at the same time describe some special kinds of algebras that will be useful for our later work.

Before leaving this introduction, we will briefly address the question, which we mentioned in Chapter 1, of whether there is any mathematical basis for thinking binary operations are special. (We will address this question more systematically when we study the algebraic representation of lattices, monoids, and groups in later volumes.) We may especially ask whether a single binary operation in inherently different for some other collection of operations, such as a single ternary operation or two binary operations, and whether a finite collection of operations is inherently different from an infinite collection.

To understand the relationships between operations of various arities, it helps

to have the notion of a **term operation** of an algebra $\mathbf{A} = \langle A, F_1, F_2, \dots \rangle$, which we now present somewhat informally. (For a more formal presentation, see Definition 4.2 in §4.1.) The set of term operations is the smallest set that contains each F_i, for $0 \leq i < n$, the coordinate projection operation

$$p_i^n(x_0, \dots, x_{n-1}) = x_i$$

and that is closed under composition.

3.2 Unary Algebras

3.3 Semigroups

3.4 Groups, Quasigroups, and Latin Squares

3.5 Representations in End$A$ and Sym$A$

3.6 Categories

C H A P T E R F O U R

Fundamental Algebraic Results

THEOREM 4.1 (The Subdirect Representation Theorem)**.** *Every algebra* **A** *has a subdirect representation with subdirectly irreducible factors that are quotient algebras of* **A**.

C H A P T E R F I V E

Unique Factorization

Bibliography