

Chapter 5. Lattices, closure operators, and Galois connections.

5.1. Semilattices and lattices. Many of the partially ordered sets P we have seen have a further valuable property: that for any two elements of P , there is a least element \geq both of them, and a greatest element \leq both of them, i.e., a *least upper bound* and a *greatest lower bound* for the pair. In this section we shall study partially ordered sets with this property. To get a better understanding of the subject, let us start by looking separately at the properties of having least upper bounds and of having greatest lower bounds.

Recall that an element x is said to be *idempotent* with respect to a binary operation $*$ if $x*x = x$. The binary operation $*$ itself is often called idempotent if $x*x = x$ holds for all x .

Lemma 5.1.1. *Suppose X is a partially ordered set in which every two elements $x, y \in X$ have a least upper bound; that is, such that there exists a least element which majorizes both x and y . Then if we write this least upper bound as $x \vee y$, and regard \vee as a binary operation on X , this operation will satisfy the identities*

$$\begin{aligned} (\forall x) \quad x \vee x &= x && \text{(idempotence),} \\ (\forall x, y) \quad x \vee y &= y \vee x && \text{(commutativity),} \\ (\forall x, y, z) \quad (x \vee y) \vee z &= x \vee (y \vee z) && \text{(associativity).} \end{aligned}$$

Conversely, given a set X with a binary operation \vee satisfying the above three identities, there is a unique partial order relation \leq on X for which \vee is the least upper bound operation. This relation \leq may be recovered from the operation \vee in two ways: It can be constructed as

$$\{(x, x \vee y) \mid x, y \in X\},$$

or characterized as the set of elements satisfying an equation:

$$\{(x, y) \mid y = x \vee y\}. \quad \square$$

Exercise 5.1:1. Prove the non-obvious part of the above lemma, namely that every idempotent commutative associative binary operation on a set arises from a partial ordering with least upper bounds. Why is this partial ordering unique?

Hence we make

Definition 5.1.2. *An upper semilattice means a pair $S = (|S|, \vee)$, where $|S|$ is a set, and \vee (pronounced “join”) is an idempotent commutative associative binary operation on $|S|$. Informally, the term “upper semilattice” will also be used for the equivalent structure of a partially ordered set in which every pair of elements has a least upper bound.*

Given an upper semilattice $(|S|, \vee)$, we shall consider $|S|$ as partially ordered by the unique ordering which makes \vee the least upper bound operation (characterized in two equivalent ways in the above lemma). The set $|S|$ with this partial ordering is sometimes called the “underlying partially ordered set” of the upper semilattice S .

The join of a finite nonempty family of elements x_i ($i \in I$) in an upper semilattice (which by the associativity and commutativity of the join operation \vee makes sense without specification of an order or bracketing for the elements, and which is easily seen to give the least upper bound of

$\{x_i\}$ in the natural partial ordering) is denoted $\bigvee_{i \in I} x_i$.

The danger of confusion inherent in the symmetry of the partial order concept is now ready to rear its head! Observe that in a partially ordered set in which every pair of elements x, y has a *greatest lower bound* $x \wedge y$, the operation \wedge will also be idempotent, commutative and associative (it is simply the operation \vee for the opposite partially ordered set), though the partial ordering is recovered from it in the opposite way, by defining $x \leq y$ if and only if x can be written $y \wedge z$, equivalently, if and only if $x = x \wedge y$. We have no choice but to make a formally identical definition for the opposite concept (first half of the first sentence below):

Definition 5.1.3. A lower semilattice means a pair $S = (|S|, \wedge)$, where $|S|$ is a set and \wedge (pronounced “meet”) is an idempotent commutative associative binary operation on $|S|$; or informally, the equivalent structure of a partially ordered set in which every pair of elements has a greatest lower bound. If $(|S|, \wedge)$ is such a pair, regarded as a lower semilattice, then $|S|$ will be considered partially ordered in the unique way which makes \wedge the greatest lower bound operation.

The notation for the meet of a finite nonempty family of elements is $\bigwedge_{i \in I} x_i$.

A partially ordered set (X, \leq) in which every pair of elements x and y has both a least upper bound $x \vee y$ and a greatest lower bound $x \wedge y$ is clearly determined – indeed, redundantly determined – by the 3-tuple $L = (X, \vee, \wedge)$. We see that a 3-tuple consisting of a set, an upper semilattice operation, and a lower semilattice operation arises in this way if and only if these operations are compatible, in the sense that the unique partial ordering for which \vee is the least-upper-bound operation coincides with the unique partial ordering for which \wedge is greatest-lower-bound operation.

Is there a nice formulation for this compatibility condition? The statement that for any two elements x and y , the element y can be written $x \vee z$ for some z if and only if the element x can be written $y \wedge w$ for some w would do, but it is awkward. If, instead of using as above the descriptions of how to *construct* all pairs (x, y) with $x \leq y$ with the help of the operations \vee and \wedge , we use the formulas that characterize them as solution-sets of equations, we get the condition that for all elements x and y , $y = x \vee y \Leftrightarrow x \wedge y = x$. But the best expression for our condition – one that does not use any “can be written”s or “ \Leftrightarrow ”s – is obtained by playing off one description of \vee against the other description of \wedge . This is the fourth pair of equations in

Definition 5.1.4. A lattice will mean a 3-tuple $L = (|L|, \vee, \wedge)$ satisfying the following identities for all $x, y, z \in |L|$:

$$\begin{array}{lll}
 x \vee x = x & x \wedge x = x & (\text{idempotence}), \\
 x \vee y = y \vee x & x \wedge y = y \wedge x & (\text{commutativity}), \\
 (x \vee y) \vee z = x \vee (y \vee z) & (x \wedge y) \wedge z = x \wedge (y \wedge z) & (\text{associativity}), \\
 x \wedge (x \vee y) = x & x \vee (x \wedge y) = x & (\text{compatibility}),
 \end{array}$$

in other words, such that $(|L|, \vee)$ is an upper semilattice, $(|L|, \wedge)$ is a lower semilattice, and these two semilattices have the same natural partial ordering. Informally, the term will also be used for the equivalent structure of a partially ordered set in which every pair of elements has a least upper bound and a greatest lower bound.

Given a lattice $(|L|, \vee, \wedge)$, we shall consider $|L|$ partially ordered by the unique partial

ordering (characterizable in four equivalent ways) which makes its join operation the least upper bound and its meet operation the greatest lower bound. The set $|L|$ with this partial ordering is sometimes called the “underlying partially ordered set of L ”.

Examples: If S is a set, then the power set $\mathbf{P}(S)$ (the set of all subsets of S), partially ordered by the relation of inclusion, has least upper bounds and greatest lower bounds, given by operations of union and intersection of sets; hence $(\mathbf{P}(S), \cup, \cap)$ is a lattice. Since the definition of Boolean algebra was modeled on the structure of the power set of a set, every Boolean algebra $(|B|, \cup, \cap, {}^c, 0, 1)$ gives a lattice $(|B|, \cup, \cap)$ on dropping the last three operations; and since we know that Boolean rings are equivalent to Boolean algebras, every Boolean ring $(|B|, +, \cdot, -, 0, 1)$ becomes a lattice under the operations $x \vee y = x + y + xy$ and $x \wedge y = xy$.

Every totally ordered set – for instance, the real numbers – is a lattice, since the larger and the smaller of two elements will respectively be their least upper bound and greatest lower bound. The set of real-valued functions on any set X may be ordered by writing $f \leq g$ if $f(x) \leq g(x)$ for all x , and this set is a lattice under *pointwise* maximum and minimum.

Under the partial ordering by divisibility, the set of positive integers has least upper bounds and greatest lower bounds, called “least common multiples” and “greatest common divisors”. Note that if we represent a positive integer by its prime factorization, and consider such a factorization as a function associating to each prime a nonnegative integer, then least common multiples and greatest common divisors reduce to pointwise maxima and minima of these functions.

Given a group G , if we order the set of subgroups of G by inclusion, then we see that for any two subgroups H and K , there is a largest subgroup contained in both, gotten by intersecting their underlying sets, and a smallest subgroup containing both, the subgroup *generated* by the union of their underlying sets. So the set of subgroups of G forms a lattice, called the *subgroup lattice* of G . This observation goes over word-for-word with “group” replaced by “monoid”, “ring”, “vector space”, etc..

Some writers use “ring-theoretic” notation for lattices, writing $x + y$ for $x \vee y$, and xy for $x \wedge y$. Note, however, that a nontrivial lattice is never a ring (by idempotence, its join operation cannot be a group structure). We will not use such notation here.

Although one can easily draw pictures of partially ordered sets and semilattices which are not lattices, it takes a bit of thought to find naturally occurring examples. The next exercise notes a couple of these.

Exercise 5.1:2. (i) If G is an infinite group, show that within the lattice of subgroups of G , the finitely generated subgroups form an upper semilattice under the induced order, but not necessarily a lower semilattice, and the finite subgroups form a lower semilattice but not necessarily an upper semilattice. (For partial credit you can verify the positive assertions; for full credit you must find examples establishing the negative assertions as well.)

(ii) Let us partially order the set of *polynomial* functions on the unit interval $[0, 1]$ by pointwise comparison ($f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in [0, 1]$). Show that this partially ordered set is neither an upper nor a lower semilattice.

Exercise 5.1:3. Give an example of a 3-tuple $(|L|, \vee, \wedge)$ which satisfies all the identities defining a lattice except for *one* of the two compatibility identities. If possible, give a systematic way of constructing such examples. Can you determine for which upper semilattices $(|L|, \vee)$ there will exist operations \wedge such that $(|L|, \vee, \wedge)$ satisfies all the lattice identities except the specified one? (The answer will depend on which identity you leave out; you can try to solve the problem for one or both cases.)

Exercise 5.1:4. Show that the two compatibility identities in Definition 5.1.4 together imply the two idempotence identities.

Exercise 5.1:5. Show that an element of a lattice is a *maximal* element if and only if it is a *greatest* element. Is this true in every upper semilattice? In every lower semilattice?

A *homomorphism* of lattices, upper semilattices, or lower semilattices means a map of their underlying sets which respects the lattice or semilattice operations. If L_1 and L_2 are lattices, one can speak loosely of an “upper semilattice homomorphism $L_1 \rightarrow L_2$ ”, meaning a map of underlying sets which respects joins but not necessarily meets; this is really a homomorphism $(L_1)_\vee \rightarrow (L_2)_\vee$, where $(L_i)_\vee$ denotes the upper semilattice $(|L_i|, \vee)$ gotten by forgetting the operation \wedge ; one may similarly speak of “lower semilattice homomorphisms” of lattices. Note that if $f: |L_1| \rightarrow |L_2|$ is a lattice homomorphism, or an upper semilattice homomorphism, or a lower semilattice homomorphism, it will be an isotone map with respect to the natural order-relations on $|L_1|$ and $|L_2|$, but in general, an isotone map f need not be a homomorphism of any of these sorts.

A *sublattice* of a lattice L is a lattice whose underlying set is a subset of $|L|$ and whose operations are the restrictions to this set of the operations of L . A *subsemilattice* of an upper or lower semilattice is defined similarly, and one can speak loosely of an upper or lower subsemilattice of a lattice L , meaning a subsemilattice of L_\vee or L_\wedge .

Exercise 5.1:6. (i) Give an example of a subset S of the underlying set of a lattice L such that every pair of elements of S has a least upper bound and a greatest lower bound in S under the induced ordering, but such that S is not the underlying set of either an upper or a lower subsemilattice of L .

(ii) Give an example of an upper semilattice homomorphism between lattices that is not a lattice homomorphism.

(iii) Give an example of a bijective isotone map between lattices which is not an upper or lower semilattice homomorphism.

(iv) Show that a bijection between lattices is a lattice isomorphism if either (a) it is an upper (or lower) semilattice homomorphism, or (b) it and its inverse are both isotone.

Exercise 5.1:7. Let k be a field. If V is a k -vector space, then the *cosets of subspaces* of V , together with the empty set, are called the *affine subspaces* of V .

(i) Show that the affine subspaces of a vector space form a lattice.

(ii) Suppose we map the set of affine subspaces of the vector space k^n into the set of *vector subspaces* of k^{n+1} by sending each affine subspace $A \subseteq k^n$ to the vector subspace $s(A) \subseteq k^{n+1}$ spanned by $\{(1, x_0, \dots, x_{n-1}) \mid (x_0, \dots, x_{n-1}) \in A\}$. Show that this map s is one-to-one. One may ask whether s respects meets and/or joins; show that it respects one of these, and respects the other in “most but not all” cases, in a sense you should make precise.

(The study of the affine subspaces of k^n is called *n-dimensional affine geometry*. By the above observations, the geometry of the vector subspaces of k^{n+1} may be regarded as a slight extension of *n-dimensional affine geometry*; this is called *n-dimensional projective geometry*. In view of the relation with affine geometry, a 1-dimensional subspace of k^{n+1} is called a “point” of projective *n*-space, a 2-dimensional subspace, or more precisely, the set of “points” it contains, is called a “line”, etc..)

The methods introduced in Chapters 2 and 3 can clearly be used to establish the existence of *free* lattices and semilattices, and of lattices and semilattices presented by *generators and relations*. As in the case of semigroups, a “relation” means a statement equating two terms formed from the given generators using the available operations – in this case, the lattice or semilattice operations.

Exercise 5.1:8. (i) If P is a partially ordered set, show that there exist universal examples of an upper semilattice, a lower semilattice, and a lattice, with isotone maps of P into their underlying partially ordered sets, and that these may be constructed as semilattices or lattices presented by appropriate generators and relations.

(ii) Show likewise that given any upper or lower semilattice S , there is a universal example of a lattice with an upper, respectively lower semilattice homomorphism of S into it.

(iii) If the S of point (ii) above “is a lattice” (has both least upper bounds and greatest lower bounds), will this universal semilattice homomorphism be an isomorphism? If the P of point (i) above “is a lattice” will the universal isotone maps of that question be isomorphisms of partially ordered sets?

(iv) Show that the universal maps of (i) and (ii) above are in general not surjective, and investigate whether each of them is in general one-to-one.

Exercise 5.1:9. Determine a normal form or other description for the free upper semilattice on a set X . Show that it will be finite if X is finite.

There exists something like a normal form theorem for free lattices [3, §VI.8], but it is much less trivial than the result for semilattices referred to in the above exercise, and we will not develop it here. However, the next exercise develops a couple of facts about free lattices.

Exercise 5.1:10. (i) Determine the structures of the free lattices on 0, 1, and 2 generators.

(ii) Show for some positive integer n that the free lattice on n generators is infinite. (One approach: In the lattice of affine subsets of the plane \mathbb{R}^2 (Exercise 5.1:7), consider the sublattice generated by the five lines $x = 0$, $x = 1$, $x = 2$, $y = 0$, $y = 1$.)

Exercise 5.1:11. (i) Recall (cf. discussion preceding Exercise 4.1:5) that a set map $X \rightarrow Y$ induces maps $\mathbf{P}(X) \rightarrow \mathbf{P}(Y)$ and $\mathbf{P}(Y) \rightarrow \mathbf{P}(X)$. Show that one of these is always, and the other is not always a lattice homomorphism.

(ii) If L is (a) a lattice, respectively (b) an upper semilattice, (c) a lower semilattice or (d) a partially ordered set, show that there exists a universal example of a set X together with

(a) a lattice homomorphism $L \rightarrow (\mathbf{P}(X), \cup, \cap)$,

(b) an upper semilattice homomorphism $L \rightarrow (\mathbf{P}(X), \cup)$,

(c) a lower semilattice homomorphism $L \rightarrow (\mathbf{P}(X), \cap)$, or

(d) an isotone map $L \rightarrow (\mathbf{P}(X), \subseteq)$ (unless you did this case in Exercise 4.1:5).

In each case, first formulate the proper universal properties. These should be based on the construction of part (i) that *does* give lattice homomorphisms. In each case, describe the set X as explicitly as you can.

(iii) In the context of part (i), the map between $\mathbf{P}(X)$ and $\mathbf{P}(Y)$ that does not generally give a lattice homomorphism will nevertheless preserve some of the types of structure named in part (ii). If L is an arbitrary structure of one of those sorts, see whether you can find an example of a set X and a map $|L| \rightarrow \mathbf{P}(X)$ respecting that structure, and universal with respect to induced maps in the indicated direction.

(iv) For which of the constructions that you obtained in parts (ii) and/or (iii) can you show the universal map $|L| \rightarrow \mathbf{P}(X)$ one-to-one? In the case(s) where you cannot prove it is, can you find an example in which it is not one-to-one?

In Exercise 4.6:3, we saw that any partially ordered set without maximal elements has two disjoint cofinal subsets. Let us examine what similar results hold for lattices.

Exercise 5.1:12. Let L be a lattice without greatest element.

(i) If L is *countable*, show that it contains a cofinal chain, that this chain will have two disjoint cofinal subchains, and that these will be disjoint cofinal sublattices of L .

(ii) Show that in general, L may not have a cofinal chain.

- (iii) Must L have two disjoint cofinal sublattices? (I don't know the answer.)
- (iv) Show that L will always contain two disjoint upper subsemilattices, each cofinal in L .

Here is another open question, of a related sort.

Exercise 5.1:13. (i) (Open question, David Wasserman.) If L is a lattice with more than one element, must L have two proper sublattices L_1 and L_2 whose union generates L ?

Parts (ii) and (iv) below, which are fairly easy, give some perspective on this question; parts (iii) and (v) are digressions suggested by (ii) and (iv).

(ii) Show that if A is a group, monoid, ring or lattice which is finitely generated but cannot be generated by a single element, then A is generated by the union of two proper subgroups, subrings, etc.. (You can give one proof that covers all these cases. In the context of Chapter 8, this could be stated as a result about general algebras.)

(iii) Determine precisely which finitely generated groups are not generated by the union of any two proper subgroups.

(iv) Let p be a prime and $\mathbb{Z}[p^{-1}]$ the subring of \mathbb{Q} generated by p^{-1} , and let $\mathbb{Z}[p^{-1}]_{\text{add}}$ denote its underlying additive group. Show that the abelian group $\mathbb{Z}[p^{-1}]_{\text{add}}/\mathbb{Z}_{\text{add}}$ is non-finitely-generated, and cannot be generated by two proper subgroups. (It happens that every finitely generated subgroup of this group also has the property of not being generated by two proper subgroups, but I don't know whether this is significant.)

(v) Are the groups of parts (iii) and (iv) above the only ones that are not generated by two proper subgroups?

I could not end an introduction to lattices without showing you the concepts introduced in the next two exercises, though this brief mention, and the results developed in the two subsequent exercises, will hardly do them justice. I will refer in these exercises to the following two 5-element lattices:



(M_5 is sometimes called M_3 .)

Exercise 5.1:14. (i) Show that the following conditions on a lattice L are equivalent:

- (a) For all $x, y, z \in |L|$ with $x \leq z$, one has $x \vee (y \wedge z) = (x \vee y) \wedge z$.
- (b) L has no sublattice isomorphic to N_5 (shown above).
- (c) For every pair of elements $x, y \in |L|$, the intervals $[x \wedge y, y]$ and $[x, x \vee y]$ are isomorphic, the map in one direction being given by $z \mapsto x \vee z$, in the other direction by $z \mapsto z \wedge y$.

(ii) Show that condition (a) is equivalent to an *identity*, i.e., a statement that a certain equation in n variables and the lattice-operations holds for all n -tuples of elements of L . (Condition (a) as stated fails to be an identity, because it refers only to 3-tuples satisfying $x \leq z$.)

(iii) Show that the lattice of subgroups of an abelian group satisfies the above equivalent conditions.

Deduce that the lattice of submodules of a module over a ring will satisfy the same conditions. For this reason, a lattice satisfying these conditions is called *modular*.

(iv) Determine, as far as you can, whether each of the following lattices is in general modular: the lattice of all subsets of a set; the lattice of all subgroups of a group; the lattice of all normal subgroups of a group; the lattice of all ideals of a ring; the lattice of all subrings of a ring; the lattice of all subrings of a Boolean ring; the lattice of elements of a Boolean ring under the operations $x \vee y = x + y + xy$ and $x \wedge y = xy$; the lattice of all sublattices of a lattice; the lattice of all closed subsets of a topological space; the lattices associated with n -dimensional affine

geometry and with n -dimensional projective geometry (Exercise 5.1:7 above).

Exercise 5.1:15. (i) Show that the following conditions on a lattice L are equivalent:

- (a) For all $x, y, z \in |L|$, one has $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$.
- (a*) For all $x, y, z \in |L|$, one has $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.
- (b) L has no sublattice isomorphic either to M_5 or to N_5 .

Note that if one thinks of \vee as “addition” and \wedge as “multiplication”, then (a*) has the form of the distributive law of ring theory. (Condition (a) is also a distributive law, though this identity does not hold in any nonzero ring.) Hence lattices satisfying the above equivalent conditions are called *distributive*.

- (ii) Show that the lattice of subsets of a set is distributive.
- (iii) Determine, as far as you can, whether lattices of each of the remaining sorts listed in parts (iii) and (iv) of the preceding exercise are always distributive.
- (iv) Show that every finitely generated distributive lattice is finite.

Exercise 5.1:16. Let V be a vector space over a field k , let S_1, \dots, S_n be subspaces of V , and within the lattice of all subspaces of V , let L denote the sublattice generated by S_1, \dots, S_n .

- (i) Show that if V has a basis B such that each S_i is spanned by a subset of B , then L is distributive, as defined in the preceding exercise

Below we will prove the converse of (i); so for the remainder of this exercise, we assume L is distributive.

To prove the existence of a basis as in the hypothesis of (i), it will suffice to prove that V contains a direct sum of subspaces, with the property that each S_i is the sum of some subfamily thereof; so this is what we will aim for. (You’ll give the details of why this yields the desired result in the last step.)

You may assume the last result of the preceding exercise, that every finitely generated distributive lattice is finite.

- (ii) Let $T = S_1 + \dots + S_n$, the largest element of L . Assuming L has elements other than T , let W be maximal among these. Show that there is a *least* element $U \in L$ not contained in W .
- (iii) Let E be a subspace of V such that $U = (U \cap W) \oplus E$. (Why does one exist?) Show that every member of L is either contained in W , or is the direct sum of E with a member of L contained in W .
- (iv) Writing L' for the sublattice of L consisting of members of L contained in W , show that the lattice of subspaces of V generated by $\{S_1, \dots, S_n, E\}$ is isomorphic to $L' \times \{0, E\}$.
- (v) Conclude by induction that there exists a family of subspaces $E_1, \dots, E_r \subseteq V$ such that every member of L , and hence in particular, each of S_1, \dots, S_n , is the direct sum of a subset of this family. Deduce that V has a basis B such that each S_i is spanned by a subset of B .

Exercise 5.1:17. Let us show that the result of the preceding exercise fails for infinite families $(S_i)_{i \in I}$. Our example will be a chain of subspaces, so

- (i) Verify that every chain in a lattice is a distributive sublattice.

Now let k be a field, and V the k -vector-space of all k -valued functions on the nonnegative integers. You may assume the standard result that V is uncountable-dimensional. For each nonnegative integer n , let $S_n = \{f \in V \mid f(i) = 0 \text{ for } i < n\}$.

- (ii) Show that V does not have a basis B such that each S_i is spanned by a subset of B . (One way to start: Verify that for each n , S_{n+1} has codimension 1 in S_n , and that the intersection of these subspaces is $\{0\}$, but that $\dim(V)$ is uncountable.)

The preceding exercise suggests

Exercise 5.1:18. Can you find necessary and sufficient conditions on a lattice L for it to be true that for every homomorphism f of L into the lattice of subspaces of a vector space V , there exists a basis B of V such that every subspace $f(x)$ ($x \in |L|$) is spanned by a subset of B ?

We remark that the analog of Exercise 2.3:2, with the finite lattice N_5 in place of the finite group S_3 , is worked out for $n = 3$ in [124].

5.2. 0, 1, and completeness. We began this chapter with the observation that many natural examples of partially ordered sets have the property that every *pair* of elements has a least upper bound and a greatest lower bound. But most of these examples in fact have the stronger property that such bounds exist for every *set* of elements. E.g., in the lattice of subgroups of a group, one can take the intersection of, or the subgroup generated by the union of an arbitrary set of subgroups. The property that every subset $\{x_i \mid i \in I\}$ has a least upper bound (denoted $\bigvee_I x_i$) and a greatest lower bound (denoted $\bigwedge_I x_i$) leads to the class of nonempty *complete* lattices, which we shall consider in this section.

Note that in an ordinary lattice, because every *pair* of elements x, y has a least upper bound $x \vee y$, it follows that for every positive integer n , every family of n elements x_0, \dots, x_{n-1} has a least upper bound, namely $\bigvee x_i = x_0 \vee \dots \vee x_{n-1}$. Hence, to get least upper bounds for *all* families, we need to bring in the additional cases of *infinite* families, and the *empty* family.

Now every element of a lattice is an upper bound of the empty family, so a *least upper bound* for the empty family means a least *element* in the lattice. Such an element is often written 0 , or when there is a possibility of ambiguity, 0_L . Likewise, a greatest lower bound for the empty family means a greatest element, commonly written 1 or 1_L .

It is not hard to see that the two conditions that a partially ordered set have pairwise least upper bounds and that it have a least element (a least upper bound of the empty family) are independent: either, neither, or both may hold. On the other hand, existence of pairwise joins and existence of infinite joins (joins indexed by infinite families, with repetition allowed just as in the case of pairwise joins) are not independent; the latter condition implies the former. However, we may ask whether the property “existence of infinite joins” can somehow be decomposed into the conjunction of existence of pairwise joins, and some natural condition which *is* independent thereof. The next result shows that it can, and more generally, that for any cardinal α , the condition “there exist joins of all families of cardinality $< \alpha$ ” can be so decomposed.

Lemma 5.2.1. *Let P be a partially ordered set, and α an infinite cardinal. Then the following conditions are equivalent:*

- (i) *Every nonempty subset of P with $< \alpha$ elements has a least upper bound in P .*
- (ii) *Every pair of elements of P has a least upper bound, and every nonempty chain in P with $< \alpha$ elements has a least upper bound.*

The dual statements concerning greatest lower bounds are likewise equivalent to one another.

Proof. (i) \Rightarrow (ii) is clear.

Conversely, assuming (ii) let us take any set X of $< \alpha$ elements of P , and index it by an ordinal $\beta < \alpha$: $X = \{x_\varepsilon \mid \varepsilon < \beta\}$. We shall prove inductively that for $0 < \gamma \leq \beta$, there exists a least upper bound $\bigvee_{\varepsilon < \gamma} x_\varepsilon$. Because we have not assumed a least upper bound for the empty set, this need not be true for $\gamma = 0$, so we start the induction by observing that for $\gamma = 1$, the set $\{x_\varepsilon \mid \varepsilon < 1\} = \{x_0\}$ has least upper bound x_0 . Now let $\gamma > 1$ and assume our result is true for all positive $\delta < \gamma$. If γ is a successor ordinal, $\gamma = \delta + 1$, then we apply the existence of pairwise

least upper bounds in P and see that $(\bigvee_{\varepsilon < \delta} x_\varepsilon) \vee x_\delta$ will give the desired least upper bound $\bigvee_{\varepsilon < \gamma} x_\varepsilon$. On the other hand, if γ is a limit ordinal, then the elements $\bigvee_{\varepsilon < \delta} x_\varepsilon$ where δ ranges over all nonzero members of γ will form a chain of $< \alpha$ elements in P , which by (ii) has a least upper bound, and this is the desired element $\bigvee_{\varepsilon < \gamma} x_\varepsilon$. So by induction, $\bigvee_{\varepsilon < \beta} x_\varepsilon$ exists, proving (i).

The final statement follows by duality. \square

Definition 5.2.2. *Let α be a cardinal. Then a lattice or an upper semilattice L in which every nonempty set of $< \alpha$ elements has a least upper bound will be called $< \alpha$ -upper semicomplete. A lattice or a lower semilattice satisfying the dual condition is said to be $< \alpha$ -lower semicomplete. A lattice satisfying both conditions will be called $< \alpha$ -complete.*

When these conditions hold for all cardinals α , one calls L upper semicomplete, respectively lower semicomplete, respectively complete.

One can similarly speak of a lattice or semilattice as being *upper or lower $\leq \alpha$ -semicomplete* if all subsets of cardinality $\leq \alpha$ have least upper bounds, respectively greatest lower bounds. Note, however, that upper or lower $\leq \alpha$ -semicompleteness is equivalent to upper or lower $< \alpha'$ -semicompleteness respectively, where α' is the successor of the cardinal α . Since not every cardinal is a successor, the class of conditions named by the “ $<$ ” terms properly contains the class of conditions named by the “ \leq ” terms.

Consequently, in the interest of brevity, many authors write “ α -semicomplete” and “ α -complete” for what we are calling “ $< \alpha$ -semicomplete” and “ $< \alpha$ -complete”. I prefer to use a more transparent terminology, however.

We have observed, in effect, that every lattice is $< \aleph_0$ -complete; so the first case of interest among the above completeness conditions is that of $\leq \aleph_0$ -completeness, equivalently, $< \aleph_1$ -completeness. This property is commonly called *countable completeness*, even by authors who in their systematic notation would write it as \aleph_1 -completeness. Countable upper and lower semicompleteness are defined similarly. We will not, however, use these terms in these notes, since the cases of greatest interest to us in this section, and the only cases we will be concerned with after this section, are the full completeness conditions.

Note that in a partially ordered set (e.g., a lattice) with *ascending chain condition*, all nonempty chains have least upper bounds – since they in fact have greatest elements. Likewise in a partially ordered set with *descending chain condition*, all chains have greatest lower bounds.

Exercise 5.2:1. Suppose β and γ are infinite cardinals, and X a set of cardinality \geq both β and γ . Let $L = \{S \subseteq X \mid \text{card}(S) < \beta \text{ or } \text{card}(X-S) < \gamma\}$. Verify that L is a lattice, and investigate for what cardinals α this lattice is upper, respectively lower $< \alpha$ -semicomplete.

The upper and lower semicompleteness conditions, when not restricted as to cardinality, have an unexpectedly close relation.

Proposition 5.2.3. *Let L be a partially ordered set. Then the following conditions are equivalent:*

- (i) *Every subset of L has a least upper bound; i.e., L is the underlying partially ordered set of an upper semicomplete upper semilattice with least element.*
- (i*) *Every subset of L has a greatest lower bound; i.e., L is the underlying partially ordered set of a lower semicomplete lower semilattice with greatest element.*

(ii) L is the underlying partially ordered set of a nonempty complete lattice.

Proof. To see the equivalence of the two formulations of (i), recall that a least upper bound for the empty set is a least element, while the existence of least upper bounds for all other subsets is what it means to be an upper semicomplete upper semilattice.

To show (i) \Rightarrow (ii), observe that the existence of a least element shows that L is nonempty, and the upper complete upper semilattice condition gives half the condition to be a complete lattice. It remains to show that any nonempty subset X of L has a greatest lower bound u . In fact, the least *upper* bound of the set of *all lower* bounds for X will be the desired u ; the reader should verify that it has the required property.

Conversely, assuming (ii), we have by definition least upper bounds for all *nonempty* subsets of L . A least upper bound for the empty set is easily seen to be given by the greatest lower bound of all of L . (How is the nonemptiness condition of (ii) used?)

Since (ii) is self-dual and equivalent to (i), it is also equivalent to (i*). \square

Exercise 5.2:2. If T is a topological space, show that the open sets in T , partially ordered by inclusion, form a complete lattice. Describe the meet and join operations (finite and arbitrary) of this lattice. Translate these results into statements about the set of closed subsets of T .

(General topology buffs may find it interesting to show that, on the other hand, the partially ordered set $\{\text{open sets}\} \cup \{\text{closed sets}\}$ is not in general a lattice, nor is the partially ordered set of *locally closed* sets.)

Exercise 5.2:3. Which ordinals, when considered as ordered sets, form complete lattices?

Exercise 5.2:4. (i) Show that every *isotone map* from a nonempty complete lattice into itself has a fixed point.

(ii) Can you prove the same result for a larger class of partially ordered sets?

Exercise 5.2:5. Let L be a complete lattice.

(i) Show that the following conditions are equivalent: (a) L has no chain order-isomorphic to an uncountable cardinal. (b) For every subset $X \subseteq |L|$ there exists a countable subset $Y \subseteq X$ such that $\bigvee Y = \bigvee X$.

(ii) Let a be any element of L . Are the following conditions equivalent? (a) L has no chain order-isomorphic to an uncountable cardinal and having join a . (b) Every subset $X \subseteq L$ with join a contains a countable subset Y also having join a .

When we were motivating the statement of Zorn's Lemma in the preceding chapter, we said that in the typical construction where one calls on it, if one has a chain of partial constructions, then their "union" is generally a partial construction extending them all. This means that the set of partial constructions is a partially ordered set in which every chain has not merely an *upper bound* but a *least upper bound*. This leads to the following question: Suppose we state a "weakened" form of Zorn's Lemma, saying only that partially ordered sets with *this* property have maximal elements – which is virtually all one ever uses. Is this equivalent to the full form of Zorn's Lemma? This is answered in

Exercise 5.2:6. Show, without assuming the Axiom of Choice, that the statement "If P is a nonempty partially ordered set such that all nonempty chains in P have least upper bounds, then P has a maximal element", implies the full form of Zorn's Lemma. (If possible, make your proof self-contained, i.e., avoid using the relation between Zorn's Lemma and the Axiom of Choice etc..)

Our proof in Lemma 5.2.1 that the existence of least upper bounds of *chains* made a lattice

upper semicomplete really only used well-ordered chains, i.e., chains order-isomorphic to ordinals. In fact, one can do still better:

Exercise 5.2:7. Recall from Exercise 4.6:2 that every totally ordered set has a cofinal subset order-isomorphic to a regular cardinal.

- (i) Deduce that for P a partially ordered set and α an infinite cardinal, the following two conditions are equivalent:
 - (a) Every chain in P of cardinality $< \alpha$ has a least upper bound.
 - (b) Every chain in P which is order-isomorphic to a regular cardinal $\beta < \alpha$ has a least upper bound.
- (ii) With the help of the above result, extend Lemma 5.2.1, adding a third equivalent condition.

There are still more ways than those we have seen to decompose the condition of being a complete lattice, as shown in point (ii) of

Exercise 5.2:8. (i) Show that following conditions on a partially ordered set L are equivalent:

- (a) Every nonempty subset of L having an upper bound has a least upper bound.
 - (b) Every nonempty subset of L having a lower bound has a greatest lower bound.
 - (c) L satisfies the *complete interpolation property*: Given two nonempty subsets X, Y of L , such that every element of X is \leq every element of Y , there exists an element $z \in L$ which is \geq every element of X and \leq every element of Y .
- (ii) Show that L is a nonempty complete lattice if and only if it has a greatest and a least element, and satisfies the above equivalent conditions.
 - (iii) Give an example of a partially ordered set which satisfies (a)-(c) above, but is not a lattice.
 - (iv) Give an example of a partially ordered set with greatest and least elements, which has the *finite interpolation property*, i.e., satisfies (c) above for all finite nonempty families X and Y , but which is not a lattice.

This condition-splitting game is carried further in

Exercise 5.2:9. If σ and τ are conditions on sets of elements of partially ordered sets, let us say that a partially ordered set L has the (σ, τ) -*interpolation property* if for any two subsets X and Y of L such that X satisfies σ , Y satisfies τ , and all elements of X are \leq all elements of Y , there exists an element $z \in L$ which is \geq every element of X and \leq every element of Y . Now consider the *nine* conditions on L gotten by taking for σ and τ all combinations of the three properties “is empty”, “is a pair” and “is a chain”.

- (i) Find simple descriptions for as many of these nine conditions as you can; in particular, note cases that are equivalent to conditions we have already named.
- (ii) Show that L is a nonempty complete lattice if and only if it satisfies all nine of these conditions.
- (iii) How close to independent are these nine conditions? To answer this, determine as well as you can which of the $2^9 = 512$ functions from the set of these conditions to the set $\{\text{true}, \text{false}\}$ can be realized by appropriate choices of L . (Remark: A large number of these combinations *can* be realized, so to show this, you will have to produce a large number of examples. I therefore suggest that you consider ways that examples with certain *combinations* of properties can be obtained from examples of the separate properties.)

Exercise 5.2:10. (i) We saw in Exercise 5.1:2(ii) that the set of real polynomial functions on the unit interval $[0,1]$, partially ordered by the relation $(\forall x \in [0,1]) f(x) \leq g(x)$, does not form a lattice. Show, however, that it has the finite interpolation property. (This gives a solution to Exercise 5.2:8(iii), but far from the easiest solution. The difficulty in proving this result arises

from the possibility that some members of X may be tangent to some members of Y .)

(ii) Can you obtain similar results for the partially ordered set of real polynomial functions on a general compact set $K \subseteq \mathbb{R}^n$?

Although we write the least upper bound and greatest lower bound of a set X in a complete lattice as $\bigvee X$ or $\bigvee_{x \in X} x$ and $\bigwedge X$ or $\bigwedge_{x \in X} x$, and call these the meet and join of X , these “meet” and “join” are not *operations* in quite the sense we have been considering so far. An operation is supposed to be a map $S^n \rightarrow S$ for some n . One may allow n to be an infinite cardinal (or other set), but when we consider complete lattices, there is no fixed cardinality to use. This suggests that we should consider each of the symbols \bigvee and \bigwedge to stand for a *system* of operations, of varying finite and infinite arities. But how large is this system? In a given complete lattice L , all meets and joins reduce (by dropping repeated arguments) to meets and joins of families of cardinalities $\leq \text{card}(|L|)$. But if we want to develop a general theory of complete lattices, then meets and joins of families of arbitrary cardinalities will occur, so this “system of operations” will not be a *set* of operations. We shall eventually see that as a consequence of this, though complete lattices are in many ways like algebras, not all of the results that we prove about algebras will be true for them (Exercise 7.10:6(iii)).

Another sort of complication in the study of complete lattices comes from the equivalence of the various conditions in Proposition 5.2.3: Since these lattices can be characterized in terms of different systems of operations, there are many natural kinds of “maps” among them: maps which respect arbitrary meets, maps which respect arbitrary joins, maps which respect both, maps which respect meets of all *nonempty* sets and joins of all *pairs*, etc.. The term “homomorphism of complete lattices” will mean a map respecting meets and joins of all nonempty sets, but the other kinds of maps are also of interest. These distinctions are brought out in:

Exercise 5.2:11. (i) Show that every nonempty complete lattice can be embedded, by a map which respects arbitrary *joins* (including the join of the empty set), in a power set $\mathbf{P}(S)$, for some set S , and likewise may be embedded in a power set by a map which respects arbitrary *meets*.

(ii) On the other hand, show, either using Exercise 5.1:15(ii) or by a direct argument, that the finite lattices M_5 and N_5 considered there cannot be embedded by any *lattice homomorphism*, i.e., any map respecting both *finite* meets and *finite* joins, in a power set $\mathbf{P}(S)$.

An interesting pair of invariants related to point (i) above is examined in

Exercise 5.2:12. (i) For L a nonempty complete lattice and α a cardinal, show that the following conditions are equivalent: (a) L can be embedded, by a map respecting arbitrary *meets*, in the power set $\mathbf{P}(S)$ of a set of cardinality α , (b) There exists a subset $T \subseteq |L|$ of cardinality $\leq \alpha$ such that every element of L is the *join* of a (possibly infinite) subset of T , (c) L can be written as the image, under a map respecting arbitrary joins, of the power set $\mathbf{P}(U)$ of a set of cardinality α .

From condition (b) above we see that for every L there will exist α such that these equivalent conditions hold. Let us call the least cardinal with this property the *upward generating number* of L , because of formulation (b). Dually, we have the concept of *downward generating number*.

(ii) Find a finite lattice L for which these two generating numbers are not equal.

The above exercise concerned complete lattices. On the other hand, if L is a lower complete lower semilattice with no greatest element, we can't map any power set $\mathbf{P}(X)$ onto it by a homomorphism of such semilattices, since a partially ordered set with greatest element can't be taken even by an isotone map onto one without greatest element. As a next best possibility, one

might ask whether one can map onto any such L some lower complete lower semilattice of the form ω^X , since this does not in general have a greatest element, but is nonetheless a full direct product. A nice test case for this idea would be to take for L the lattice of all finite subsets of a set S . The first part of the next exercise shows that for this case, the answer to the above question is yes.

Exercise 5.2:13. (i) Let S be any set, and $\mathbf{P}_{\text{fin}}(S)$ the lower complete lower semilattice of finite subsets of S . Let ω^S denote the lower complete lower semilattice of natural-number-valued functions on S (under pointwise inequality), and $|\omega^S|$ the underlying set of this lattice, so that $\omega^{|\omega^S|}$ is the lower complete lower semilattice of natural-number-valued functions on that set. Show that there exists a surjective homomorphism $\omega^{|\omega^S|} \rightarrow \mathbf{P}_{\text{fin}}(S)$.

Suggestion: For each $s \in S$ let $s^* \in |\omega^{|\omega^S|}|$ denote the function sending each element of ω^S to its value at s . Now map each $f \in |\omega^{|\omega^S|}|$ to the set of those $s \in S$ such that $f \geq s^*$. Show that this set is finite, and this map has the desired properties.

(ii) If L is an arbitrary nonempty lower semicomplete lower semilattice, must there exist a surjective homomorphism $\omega^X \rightarrow L$ of such semilattices for some set X ? If not, can you find necessary and sufficient conditions on L for such a homomorphism to exist?

We began the preceding section by translating the concepts of a partially ordered set in which every pair of elements had a least upper bound and/or a greatest lower bound into those of a set with one or two binary operations satisfying certain identities. One can similarly formalize the concepts of ($<\alpha$ -)complete lattice and semilattice in terms of operations and identities. The details depend on how one deals with the fact that infinitary meets and joins are applied to families of many different cardinalities. The next exercise regards meet and join as operators defined on subsets of $|L|$ rather than on tuples of elements. This avoids the need to set up distinct operations of each infinite arity, but at the cost of creating a structure rather different from the general concept of ‘‘algebra’’ we will develop in Chapter 8.

Exercise 5.2:14. Let $|L|$ be a set, and suppose that \vee and \wedge are operators associating to each nonempty subset $X \subseteq |L|$ an element of $|L|$ which will be denoted $\vee(X)$, respectively $\wedge(X)$. Show that these operators are the meet and join operators of a complete lattice structure on $|L|$ if and only if

(a) For every $x_0 \in |L|$, $\vee(\{x_0\}) = x_0 = \wedge(\{x_0\})$.

(b) For every nonempty set Y of nonempty subsets of $|L|$,

$$\vee(\{\vee(X) \mid X \in Y\}) = \vee(\bigcup_{X \in Y} X) \quad \text{and} \quad \wedge(\{\wedge(X) \mid X \in Y\}) = \wedge(\bigcup_{X \in Y} X).$$

(c) The pairwise operations defined by $a \vee b = \vee(\{a, b\})$ and $a \wedge b = \wedge(\{a, b\})$ satisfy the two ‘‘compatibility’’ identities of Definition 5.1.4.

To motivate the next definition, let us consider the following situation. Suppose L is the complete lattice of all subgroups of a group G , and let $K \in L$ be a *finitely generated* subgroup of G , generated by elements g_1, \dots, g_n . Suppose K is majorized in L by the join of a family of subgroups H_i ($i \in I$), i.e., is contained in the subgroup generated by the H_i . Then each of g_1, \dots, g_n can be expressed by a group-theoretic term in elements of $\bigcup |H_i|$. But any group-theoretic term involves only finitely many elements; hence K will actually be contained in the subgroup generated by *finitely many* of the H_i . The converse also holds: If K is a *non-finitely generated* subgroup of G , then K equals (and hence is contained in) the join of all the cyclic subgroups it contains, but is *not* contained in the join of any finite subfamily thereof.

The property we have just shown to characterize the finitely generated subgroups in the lattice

of all subgroups of G has a certain parallelism with the property defining *compact* subsets in a topological space, namely that if they are covered by a family of open subsets, they are covered by some finite subfamily. Hence one makes the definition

Definition 5.2.4. An element k of a complete lattice (or more generally, of a complete upper semilattice) L is called *compact* if every set of elements of L with $\text{join} \geq k$ has a finite subset with $\text{join} \geq k$.

By the preceding observations, the compact elements of the subgroup lattice of a group are precisely the finitely generated subgroups. We will be able to generalize this observation when we have a general theory of algebraic objects.

We noted in Exercise 5.1:2(i) that the finitely generated subgroups of a group form an upper subsemilattice of the lattice of all subgroups. This suggests

Exercise 5.2:15. Do the compact elements of a complete lattice L always form an upper subsemilattice?

Exercise 5.2:16. Show that a complete lattice L has ascending chain condition if and only if all elements of L are compact.

There seems to be no standard name for an element of a complete lattice having the dual property to compactness; sometimes such elements are called *co-compact*.

We examined in Exercise 5.2:11 the embedding of semilattices and lattices in power sets $\mathbf{P}(S)$ (and found that though there were embeddings that respected meets, and embeddings that respected joins, there were not in general embeddings that respected both). Let us look briefly at another fundamental sort of complete lattice, and the problem of embedding arbitrary lattices therein.

If X is a set, and \approx_0 and \approx_1 are two equivalence relations on X , let us say \approx_1 *extends* \approx_0 if it contains it, as a subset of $X \times X$, and write $\approx_0 \leq \approx_1$ in this situation. Let $\mathbf{E}(X)$ denote the set of equivalence relations on X , partially ordered by this relation \leq . (One could use the reverse of this order, saying that \approx_0 is a *refinement* of \approx_1 when the latter extends the former, and justify considering the refinement to be “bigger” by the fact that it gives “more” equivalence classes. So our choice of the sense to give to our ordering is somewhat arbitrary; but let us stick with the ordering based on set-theoretic inclusion.)

Exercise 5.2:17. (i) Verify that the partially ordered set $\mathbf{E}(X)$ forms a complete lattice. Identify the elements $0_{\mathbf{E}(X)}$ and $1_{\mathbf{E}(X)}$.

(ii) Let L be any nonempty complete lattice, and $f: L \rightarrow \mathbf{E}(X)$ a map respecting arbitrary meets (a complete lower semilattice homomorphism respecting greatest elements). Show that for any $x, y \in X$, there is a *least* $d \in L$ such that $(x, y) \in f(d)$. Calling this element $d(x, y)$, verify that the map $d: X \times X \rightarrow L$ satisfies the following conditions for all $x, y, z \in X$:

$$(a_0) \quad d(x, x) = 0_L,$$

$$(b) \quad d(x, y) = d(y, x),$$

$$(c) \quad d(x, z) \leq d(x, y) \vee d(y, z).$$

(iii) Prove the converse, i.e., that given a nonempty complete lattice L and a set X , any function $d: X \times X \rightarrow L$ satisfying (a₀)-(c) arises as in (ii) from a unique complete lower semilattice-homomorphism $f: L \rightarrow \mathbf{E}(X)$ respecting greatest elements.

In the remaining parts, we assume that $f: L \rightarrow \mathbf{E}(X)$ and $d: X \times X \rightarrow L$ are maps related as in (ii) and (iii).

(iv) Show that the map f respects least elements, i.e., that $f(0_L) = 0_{\mathbf{E}(X)}$, if and only if d

satisfies

- (a) $d(x, y) = 0_L \Leftrightarrow x = y$ (a strengthening of (a₀) above).
- (v) Show that f respects joins of finite nonempty families if and only if d satisfies
 - (d) whenever $d(x, y) \leq p \vee q$ (where $x, y \in X$, $p, q \in |L|$), there exists a finite “path” from x to y in X , i.e., a sequence $x = z_0, z_1, \dots, z_n = y$, such that for each $i < n$, either $d(z_i, z_{i+1}) \leq p$ or $d(z_i, z_{i+1}) \leq q$.

A function d which satisfies (a)-(c) above might be called an “ L -valued metric on X ”, and (d) might be called “path sufficiency” of the L -valued metric space X . Two other properties of importance are noted in

- (vi) Assuming that f respects finite nonempty joins, i.e., satisfies (d) above, show that it respects arbitrary nonempty joins if and only if
 - (e) for all $x, y \in X$, $d(x, y)$ is a compact element of L .
- (vii) Show that f is one-to-one if and only if
 - (f) L is generated under (not necessarily finite) joins by the elements $d(x, y)$ ($x, y \in X$).

Thus, to get various sorts of embeddings of a complete lattice L in one of the form $\mathbf{E}(X)$, it suffices to construct sets X with appropriate sorts of L -valued metrics.

How can one do this? Note that if we take a tree (in the graph-theoretic sense) with edges labeled in any way by elements of L , and define the distance between two vertices to be the join of the labels on the sequence of edges connecting those vertices, then we get an L -valued metric, such that the values assumed by this metric generate the same upper semilattice as do the set of labels. This can be used to get a system (X, d) satisfying (a)-(c) and (f). We also want condition (d). This can be achieved by adjoining additional vertices:

- Exercise 5.2:18.** (i) Suppose L is a nonempty complete lattice, X an L -valued metric space, x, y two points of X and p, q elements of L such that $d(x, y) \leq p \vee q$. Show that we can adjoin new points z_1, z_2, z_3 to X and extend the metric in a consistent way so that the steps of the path x, z_1, z_2, z_3, y have lengths p, q, p, q respectively. (The least obvious part is how to define the distance from z_2 to a point $w \in X$. To do this, verify that $d(w, x) \vee p \vee q = d(w, y) \vee p \vee q$, and use the common value.)
- (ii) Show that every L -valued metric space can be embedded in a path-sufficient one. (This will involve a countable sequence of steps $X = X_0 \subseteq X_1 \subseteq \dots$ such that each X_i “cures” all failures of path-sufficiency found in X_{i-1} , using the idea of part (i). The desired space is then $\bigcup X_i$.)

Now if the given complete lattice L is generated as a complete upper semilattice by the upper subsemilattice K of its compact elements, then one can carry out the above constructions as to get condition (e) above, and hence an embedding of L in $\mathbf{E}(X)$ that respects arbitrary meets and joins. Conversely, one sees that this assumption on K is necessary for such an embedding to exist. If we don’t make this assumption on K , we can still use the above construction to embed L in a lattice $\mathbf{E}(X)$ by a map respecting arbitrary meets, and finite joins.

We shall see in the next section that any lattice can be embedded by a lattice homomorphism in a complete lattice, so the above technique shows that any lattice can be embedded by a lattice homomorphism in a lattice of equivalence relations.

If L is finite, the construction of Exercise 5.2:18 gives, in general, a countable, but not a finite L -valued metric space X . It was for a long time an open question whether every finite lattice could be embedded in the lattice of equivalence relations of a finite set. This was finally proved in 1980 by P. Pudlák and J. Tůma [108]. However, good estimates for the size of an X such that even a quite small lattice L (e.g., the 15-element lattice $\mathbf{E}(4)^{\text{op}}$) can be embedded in $\mathbf{E}(X)$

remain to be found. The least m such that $\mathbf{E}(n)^{\text{op}}$ embeds in $\mathbf{E}(m)$ has been shown by Pudlák to grow *at least* exponentially in n ; the first *upper* bound obtained for it was $2^{2^{\dots}}$ with n^2 exponents! For subsequent better results see [91] and [79, in particular p.16, top].

5.3. Closure operators. We introduced this chapter by noting certain properties common to the partially ordered sets of all subsets of a set, of all subgroups of a group, and similar examples. But so far, we seem to have made a virtue of abstractness, defining semilattice, lattice, etc., without reference to systems of subsets of sets. Neither abstractness nor concreteness is everywhere a virtue; each makes its contribution, and it is time to turn to an important class of concrete lattices.

Lemma 5.3.1. *Let S be a set. Then the following data are equivalent:*

(i) *A lower semicomplete lower subsemilattice of $\mathbf{P}(S)$ which contains $1_{\mathbf{P}(S)} = S$, in other words, a set C of subsets of S closed under taking arbitrary intersections, including the empty intersection, S itself.*

(ii) *A function $\text{cl} : \mathbf{P}(S) \rightarrow \mathbf{P}(S)$ with the properties:*

$$\begin{array}{ll} (\forall X \subseteq S) \text{cl}(X) \supseteq X & (\text{cl is increasing}), \\ (\forall X, Y \subseteq S) X \subseteq Y \Rightarrow \text{cl}(X) \subseteq \text{cl}(Y) & (\text{cl is isotone}), \\ (\forall X \subseteq S) \text{cl}(\text{cl}(X)) = \text{cl}(X) & (\text{cl is idempotent}). \end{array}$$

Namely, given C , one defines cl as the operator taking each $X \subseteq S$ to the intersection of all members of C containing X , while given cl , one defines C as the set of $X \subseteq S$ satisfying $\text{cl}(X) = X$, equivalently, as the set of subsets of S of the form $\text{cl}(Y)$ ($Y \subseteq S$). \square

Exercise 5.3.1. Verify the above lemma. That is, show that the procedures described do carry families C with the properties of point (i) to operators cl with the properties of point (ii) and vice versa, and are inverse to one another, and also verify the assertion of equivalence in the final clause.

Definition 5.3.2. *An operator cl on the class of subsets of a set S with the properties described in point (ii) of the above lemma is called a closure operator on S . If cl is a closure operator on S , the subsets $X \subseteq S$ satisfying $\text{cl}(X) = X$, equivalently, the subsets of the form $\text{cl}(Y)$ ($Y \subseteq S$), are called the closed subsets of S under cl .*

We see that virtually every mathematical construction commonly referred to as “the ... generated by” (fill in the blank with subgroup, normal subgroup, submonoid, subring, sublattice, ideal, congruence, etc.) is an example of a closure operator on a set. The operation of topological closure on subsets of any topological space is another example. Some cases are called by other names: the *convex hull* of a set of points in Euclidean n -space, the *span* of a subset of a vector space (i.e., the vector subspace it generates), the set of *derived operations* of a set of operations on a set (§1.6). Incidentally, the constructions of subgroup and subring generated by a subset of a group or ring illustrate the fact that the closure of the empty set need not be empty.

A very common way of obtaining a closure operator on a set S , which includes most of the above examples, can be abstracted as follows: One specifies a certain subset

$$(5.3.3) \quad G \subseteq \mathbf{P}(S) \times S,$$

and then defines a subset $X \subseteq S$ to be *closed* if for all $(A, x) \in G$, $A \subseteq X \Rightarrow x \in X$. It is straightforward to verify that the class of “closed sets” under this definition is closed under arbitrary intersections, and so by Lemma 5.3.1, corresponds to a closure operator cl on S .

For example, if K is a group, the operator “subgroup generated by” on subsets of $|K|$ is of this form. One takes for (5.3.3) the set of all pairs of the forms

$$(5.3.4) \quad (\{x, y\}, xy), \quad (\{x\}, x^{-1}), \quad (\emptyset, e)$$

where x and y range over $|K|$. To get the operator “*normal* subgroup generated by –”, we use the above pairs, supplemented by the further family of pairs $(\{x\}, yxy^{-1})$ ($x, y \in |K|$). Clearly, the other “... generated by” constructions mentioned above can be characterized similarly. For a non-algebraic example, the operator giving the topological closure of a subset of the real line \mathbb{R} can be obtained by taking G to consist of all pairs (A, x) such that A is the set of points of a convergent sequence, and x is the limit of that sequence.

Exercise 5.3:2. Show that for any closure operator cl on a set S , there exists a subset $G \subseteq \mathbf{P}(S) \times S$ which determines cl in the sense we have been discussing.

Exercise 5.3:3. If T is a set, display a subset $G \subseteq \mathbf{P}(T \times T) \times (T \times T)$ such that the *equivalence relations* on T are precisely the subsets of $T \times T$ closed under the operator cl corresponding to G . (The previous exercise gives us a way of doing this “blindly”. But what I want here is an explicit set, which one might show to someone who didn’t know what “equivalence relation” meant, to provide a characterization of the concept.)

In Chapter 2 we contrasted the approaches of obtaining sets one is interested in “from above” as intersections of systems of larger sets, and of building them up “from below”. We have constructed the closure operator associated with a family (5.3.3) by noting that the class of subsets of S we wish to call closed is closed under arbitrary intersections; so we have implicitly obtained these closures “from above”. The next exercise constructs them “from below”.

Exercise 5.3:4. Let S be a set and G a subset of $\mathbf{P}(S) \times S$. For X a subset of S and α any ordinal, let us define $\text{cl}_G^{(\alpha)}(X)$ recursively by:

$$\begin{aligned} \text{cl}_G^{(0)}(X) &= X, \\ \text{cl}_G^{(\alpha+1)}(X) &= \text{cl}_G^{(\alpha)}(X) \cup \{x \mid (\exists A \subseteq \text{cl}_G^{(\alpha)}(X)) (A, x) \in G\}, \\ \text{cl}_G^{(\alpha)}(X) &= \bigcup_{\beta \in \alpha} \text{cl}_G^{(\beta)}(X) \text{ if } \alpha \text{ is a limit ordinal.} \end{aligned}$$

- (i) Show (for S, G as above) that there exists an ordinal α such that for all $\beta > \alpha$ and all $X \subseteq S$, $\text{cl}_G^{(\beta)}(X) = \text{cl}_G^{(\alpha)}(X)$, and that $\text{cl}_G^{(\alpha)}(X)$ is then $\text{cl}(X)$ in the sense of the preceding discussion. (Cf. the construction in §2.2 of the equivalence relation R on group-theoretic terms as the union of a chain of relations R_i .)
- (ii) If for all $(A, x) \in G$, A is finite, show that the α of part (i) can be taken to be ω .
- (iii) For each ordinal α , can you find an example of a set S and a $G \subseteq \mathbf{P}(S) \times S$ such that α is the least ordinal having the property of part (i)?

We have seen that there are restrictions on the sorts of lattices that can be embedded by lattice homomorphisms into lattices $(\mathbf{P}(S), \cup, \cap)$ (Exercise 5.1:15), or into lattices of submodules of modules (Exercise 5.1:14). In contrast, note points (ii) and (iii) of

Lemma 5.3.5. (i) *If cl is a closure operator on a set X , then the set of cl -closed subsets of X , partially ordered by inclusion, forms a complete lattice, with the meet of an arbitrary family given by its set-theoretic intersection, and the join of such a family given by the closure of its union. Conversely,*

(ii) *Every nonempty complete lattice L is isomorphic to the lattice of closed sets of a closure operator cl on some set S ; and*

(iii) *Every lattice L is isomorphic to a sublattice of the lattice of closed sets of a closure operator cl on some set S .*

Sketch of Proof. (i): It is straightforward to verify that the indicated operations give a greatest lower bound and a least upper bound to any family of closed subsets.

(ii): Take $S = |L|$, and for each $X \subseteq S$, define $\text{cl}(X) = \{y \mid y \leq \bigvee_{x \in X} x\}$. Then L is isomorphic to the lattice of closed subsets of S , by the map $y \mapsto \{x \mid x \leq y\}$.

(iii): Again take $S = |L|$, but since joins of arbitrary families may not be defined in L , define $\text{cl}(X)$ to be the set of all elements majorized by joins of finite subsets of X . Embed L in the lattice of cl -closed subsets of S by the same map as before. \square

Exercise 5.3:5. Verify that the constructions of (ii) and (iii) above give closure operators on $|L|$, and that the induced maps are respectively an isomorphism of complete lattices and a lattice embedding.

The second of the two closure operators used in the above proof can be thought of as closing a set X in $|L|$ under forming joins of its elements, and forming meets of its elements with elements of L . In the notation that denotes join by $+$ and writes meet as ‘multiplication’, this has the same form as the definition of an ideal of a ring. So lattice-theorists often call sets of elements in a lattice closed under these operations ‘ideals’. In particular, $\{y \mid y \leq x\}$ is called the *principal ideal* generated by x .

Exercise 5.3:6. (i) Show that assertion (iii) of the preceding lemma can also be proved by taking the same S and the same map, but taking $\text{cl}(X) \subseteq S$ to be the intersection of all principal ideals of L containing X .

(ii) Will the complete lattices generated by the images of L under these two constructions in general be isomorphic?

Exercise 5.3:7. Can the representation of a (complete) lattice L by closed sets of a closure operator given in Lemma 5.3.5(ii) and/or that given in Exercise 5.3:6 be characterized by any universal properties?

Exercise 5.3:8. Show that a lattice L is complete and nonempty if and only if every intersection of principal ideals of L (including the intersection of the empty family) is a principal ideal.

The concept of a set with a closure operator is not only general enough to allow representations of all lattices, it is a convenient tool for constructing examples. For example, recall that Exercise 5.1:10(ii), if solved by the hint given, shows that a lattice generated by 5 elements can be infinite. With more work, that method could have been made to give an infinite lattice with 4 generators, but one can show that any 3-generator sublattice of the lattice of affine subspaces of a vector space is finite. However, we shall now give an ad hoc construction of a closure operator whose lattice of closed sets has an infinite sublattice generated by 3 elements.

Exercise 5.3:9. Let $S = \omega \cup \{x, y\}$, where ω is regarded as the set of nonnegative integers, and x, y are two elements not in ω . Let $G \subseteq \mathbf{P}(S) \times S$ consist of all pairs

$$(\{x, 2m\}, 2m+1), \quad (\{y, 2m+1\}, 2m+2),$$

where m ranges over ω in each case. Let L denote the lattice of closed subsets of S under the induced closure operator, and consider the sublattice generated by $\{x\}$, $\{y, 0\}$, and ω . Show by induction that for every $n \geq 0$, this sublattice of L contains the set $\{0, \dots, n\}$. Thus, this 3-generator lattice is infinite.

(Exercise 7.10:6, which you can do at this point, will show that the same technique applied to complete lattices gives 3-generator lattices of arbitrarily large cardinalities.)

Exercise 5.3:10. The lattice of the above exercise contains an infinite chain. Does there exist a 3-generator lattice which is infinite but does not contain an infinite chain?

Exercise 5.3:11. If A is an abelian group, can a finitely generated sublattice of the lattice of all subgroups of A contain an infinite chain?

We now turn to a property which distinguishes the sort of closure operators commonly occurring in algebra from those arising in topology and analysis.

Lemma 5.3.6. *Let cl be a closure operator on a set S . Then the following conditions are equivalent:*

- (i) For all $X \subseteq S$, $\text{cl}(X) = \bigcup_{(\text{finite } X_0 \subseteq X)} \text{cl}(X_0)$.
- (ii) The union of every chain of closed subsets of S is closed.
- (iii) The closure of each singleton $\{s\} \subseteq S$ is compact in the lattice of closed subsets.
- (iv) cl is the closure operator determined by a set $G \subseteq \mathbf{P}(S) \times S$ having the property that the first component of each of its members is finite. \square

Exercise 5.3:12. Prove Lemma 5.3.6.

Definition 5.3.7. *A closure operator satisfying the equivalent conditions of the above lemma will be called finitary.*

This is because the lattice of subalgebras of an algebra A has this property if the operations of A are all *finitary*, i.e., have finite arity (§1.4). (Many authors call such closure operators “algebraic” instead of “finitary”, because, as noted, the property is typical of closure operators occurring in algebra.)

Exercise 5.3:13. (i) Show that an abstract nonempty complete lattice L is isomorphic to the lattice of all closed sets under a finitary closure operator if and only if every element of L is a (possibly infinite) join of compact elements.

(ii) For what complete lattices is it true that every closure operator cl , on any set, whose lattice of closed sets is isomorphic to L is finitary?

Exercise 5.3:14. Show that a closure operator cl is finitary if and only if the compact elements in the lattice of its closed subsets are precisely the closures of finite sets. For a not necessarily finitary closure operator, what is the relation between these two classes of closed sets?

Exercise 5.3:15. Consider the following three conditions on a closure operator cl on a set S . (a) cl is finitary. (b) The union of any two cl -closed subsets of S is cl -closed. (c) Every singleton subset of S is cl -closed.

For each subset of this set of three properties, find an example of a closure operator that has the properties in that subset, but not any of the others. (Thus, 8 examples are asked for.)

Where possible, use familiar or important examples.

Exercise 5.3:16. (i) Show that a closure operator cl on a set S is the operation of topological closure with respect to some topology on S if and only if it satisfies condition (b) of the preceding exercise, and: $(c_0) \emptyset$ is cl -closed in S .

(ii) Assuming S has more than one element, show that cl is closure with respect to a Hausdorff topology if and only if it satisfies conditions (b) and (c) of the preceding exercise.

Since the operation of topological closure determines the topology, this shows that topologies on a space are equivalent to closure operators satisfying the indicated conditions.

Exercise 5.3:17. It is well known that if a group K is generated by $\leq \gamma$ elements (γ a cardinal), then $\text{card}(|K|) \leq \gamma + \aleph_0$.

(i) Deduce this fact from simple properties of the set $G \subseteq \mathbf{P}(|K|) \times |K|$ defined in (5.3.4).

(ii) Try to generalize (i) to a result on the way the cardinalities of sets increase under application of a closure operator cl obtained from a set G as above, in terms of the properties of G . Can you show by example that your results are best possible?

When we described how to construct a closure operator cl from a subset $G \subseteq \mathbf{P}(S) \times S$, it would have been tempting to call cl “the closure operator generated by G ”. This would not quite have made sense, because a closure operator is not itself a subset of $\mathbf{P}(S) \times S$. However, we can show what this is “trying to say” by setting up a correspondence between closure operators on S and certain subsets of $\mathbf{P}(S) \times S$:

Exercise 5.3:18. Let S be a set.

If cl is a closure operator on S , let us write $\sigma(\text{cl}) = \{(A, x) \mid A \subseteq S, x \in \text{cl}(A)\}$ and let us call a subset $H \subseteq \mathbf{P}(S) \times S$ a *closure system* on S if $H = \sigma(\text{cl})$ for some closure operator cl on S .

(i) Show that closure systems on S are precisely the subsets of $\mathbf{P}(S) \times S$ closed under a certain closure operator, cl_{sys} (which you should describe).

(ii) Show that for any subset $G \subseteq \mathbf{P}(S) \times S$, if we write cl_G for the closure operator determined by G in the sense discussed earlier, then $\sigma(\text{cl}_G) = \text{cl}_{\text{sys}}(G)$.

So although we cannot call cl_G the closure operator generated by G , it is the operator corresponding to the *closure system* generated by G .

Of course, I cannot resist adding

(iii) Describe cl_{sys} as the closure operator on $\mathbf{P}(S) \times S$ determined (“generated”) by an appropriate set G_{sys} (of elements of what set?)

We now have three ways of looking at closure data on a set S : as certain families of subsets of S , as certain operators on subsets of S , and as certain “systems” contained in $\mathbf{P}(S) \times S$. We take a global look at this data in:

Exercise 5.3:19. Let S be a set. Call the set of all families of subsets of S that are closed under arbitrary intersections $\text{Clofam}(S)$, and order this set by inclusion. Call the set of all closure operators on S $\text{Clop}(S)$, and order it by putting $\text{cl}_1 \leq \text{cl}_2$ if for all X , $\text{cl}_1(X) \leq \text{cl}_2(X)$. Call the set of closure systems on S in the sense of the preceding exercise $\text{Closys}(S)$, and order it by inclusion.

Verify that $\text{Clofam}(S)$, $\text{Clop}(S)$ and $\text{Closys}(S)$ are all complete lattices. Do the natural correspondences between the three types of data constitute lattice isomorphisms? If not, state precisely the relationships involved. Describe the meet and join operations of $\text{Clop}(S)$ explicitly.

Exercise 5.3:20. Investigate the subset of *finitary* closure operators within the set $\text{Clop}(S)$ defined in the preceding exercise. Will it be closed under meets (finite? arbitrary?) – joins (ditto)? Given any $\text{cl} \in \text{Clop}(S)$, will there be a least finitary closure operator containing cl ?

A greatest finitary closure operator contained in cl ?

Descending from the abstruse to the elementary, here is a problem on closure operators that could be explained to a bright High School student, but which has so far defied solution:

Exercise 5.3:21. (Péter Frankl’s question) Let S be a finite set, and cl a closure operator on S such that $\text{cl}(\emptyset) \neq S$. Must there exist an element $s \in S$ which belongs to *not more than half* of the sets closed under cl ?

(I generally state this conjecture to people not in this course in terms of “a system C of subsets of S which is closed under pairwise intersections, and contains at least one proper subset of S ”. There are still other formulations; for instance, as asking whether every nontrivial finite lattice has an element which is join-irreducible (not a join of two smaller elements) and which is majorized by no more than half the elements of the lattice.)

One occasionally encounters the dual of the type of data defining a closure operator – a system U of subsets of a set S closed under forming arbitrary *unions*; equivalently, an operator f on subsets of S which is *decreasing*, idempotent, and isotone. In this situation, the *complements* in S of the sets in U will be the closed sets of a closure operator, namely $X \mapsto {}^c(f({}^cX))$ (where c denotes complementation). When such an operator is discovered, it is often convenient to change viewpoints and work with the dual operator ${}^c f^c$, to which one can apply the theory of closure operators. However, U and f may be more natural in some situations than the dual family and map. In such cases one may refer to f as an *interior operator* (though the term is not widely used), since in a topological space, the complement of the closure of the complement of X is called the interior of X . Clearly, every result about closure operators gives a dual result on interior operators.

(Péter Frankl’s question, described in the last exercise, is most often stated in dual form, asking whether, given a system C of subsets of a finite set S which is closed under pairwise unions and contains at least one nonempty subset of S , there must exist a member of S belonging to at least half the members of C . As such, it is called the “union-closed set” question, and papers on the topic can be found by searching for the keyword “union-closed”.)

5.4. Digression: a pattern of threes. It is curious that many basic mathematical definitions involve similar systems of three parts.

A *group structure* on a set is given by (1) a neutral element, (2) an inverse-operation and (3) a multiplication; these must satisfy (1) the neutral-element laws, (2) the inverse laws and (3) the associative law.

A *partial ordering* on a set is a binary relation that is (1) reflexive, (2) antisymmetric and (3) transitive, while an *equivalence relation* is (1) reflexive, (2) symmetric and (3) transitive.

The operation of a *semilattice* is (1) idempotent, (2) commutative and (3) associative.

A *closure operator* is (1) increasing, (2) isotone and (3) idempotent.

In a *metric space*, the metric satisfies (1) a condition on when distances are 0, (2) symmetry and (3) the triangle inequality.

This parallelism is not just numerical. The general pattern seems to be that the simplest conditions or operations, those marked (1) above, have to do with the relation of an element to itself; the intermediate ones, marked (2), tell us, if we know how two elements relate in one order, how they relate in the reverse order; while the strongest, those marked (3), tell us how to use the relation of one element to a second and this second to a third to get a relation between the first and the third.

Let us see this in the examples listed above. We must distinguish in some cases between abstract structures and the “concrete” structures that motivated them.

The concrete situation motivating the concept of a group is that of a group of permutations of a set. For a set of permutations to form a group, (1) it should contain the permutation e that takes every element of the set to itself, (2) if it contains a permutation x , it should also contain the permutation x^{-1} which carries q to p whenever x carries p to q , and (3) along with any permutations x and y it should contain the permutation xy , which carries p to r whenever y carries p to q and x carries q to r . So this fits the pattern described.

When we look at the definition of an *abstract* group G , the above *closure conditions* are replaced by *operations* of neutral element, inverse, and composition. The conditions on these operations needed to mimic the internal properties of permutation groups say that when G acts on itself by left or right multiplication, the three operations of G actually behave like the constructions they are modeled on: left or right multiplication by the neutral element leaves all elements of $|G|$ fixed, left or right multiplication by x is “reversed” by the action of x^{-1} , and left or right multiplication by x followed by multiplication on the same side by y is equivalent to multiplication by yx , respectively xy . These are the neutral-element, inverse and associative laws (slightly reformulated). Finally, when we *return* from this abstract concept to its concrete origins via the concept of a G -set X , we again have three conditions, saying that the actions of the neutral element, of inverses of elements, and of composites of elements of G behave on X in the proper manner. (However, the condition for inverses is a consequence of the other two plus the group identities of G , and so is usually omitted from the definition of a G -set.)

In the definitions of *partial ordering*, of *equivalence relation*, and of *metric*, we do not have an abstraction of a structure on a set, but such a structure itself. The reader can easily verify that these 3-part definitions each have the form we have described.

In the cases of *semilattices* and *closure operators*, one can say roughly that closure operators are the concrete origins and semilattices the abstraction. My general characterization of the three components of these definitions does not, as we shall see, give quite as good a fit in this case. The condition that a closure operator be idempotent, $\text{cl}(\text{cl}(X)) = \text{cl}(X)$, may be considered a “transitivity” type condition, since it says that if you can get some elements from elements of X , and some further elements from these, then you get those further elements from X . The “reflexivity” type condition is the one saying $\text{cl}(X) \supseteq X$, since it means that what one gets from X includes all of X itself. But I cannot see a way of interpreting the remaining condition, $X \subseteq Y \Rightarrow \text{cl}(X) \subseteq \text{cl}(Y)$, as describing the relation between elements considered in two different orders.

In the abstracted concept, that of a semilattice, the three conditions of idempotence, commutativity, and associativity of the operation \vee do fit the pattern described, but they do not seem to come in a systematic way from the corresponding properties of closure operators.

When one looks at important weakenings of the concepts of group etc., the middle operation or condition seems to be the one most naturally removed: Monoids are a useful generalization of groups, and preorders are a useful generalizations both of partial orders and of equivalence relations.

The folklorist Alan Dundes argued that the number “three” holds a fundamental place in the culture of Western civilization, in ways ranging from traditional stories (three brothers go out to seek their fortune; Goldilocks and the three bears), superstitions (“third time’s a charm”), verbal formulas (“Tom, Dick and Harry”) etc., to our 3-word personal names. (See essay in [63].) He

raised the challenge of how many of the “threes” occurring in science (archeologists’ division of each epoch into an “early”, a “middle” and a “late” period; the three-stage polio vaccination; the three dimensions of physics, etc.) represent circumstances given to us by nature, and how many we have imposed on nature through cultural prejudice!

In the situation we have been discussing, I would argue that the similarity between the various sets of definitions *does* represent a genuine pattern in “mathematical nature”; that the way the pattern appears, in terms of systems of *three* conditions, in contemporary developments of these topics, is *not* the only natural way these topics could be developed; but that the fact that they are developed in this way is not a consequence of a prejudice toward the number three, but of chance. As a simple example of how these topics might be differently developed, if basic textbooks regularly first defined “monoid”, and then defined a group as a monoid with an inverse operation, and similarly first defined “preorder”, then defined partial orders and equivalence relations as preorders with certain properties, and so on, then, though we would still have a recurring pattern, it would not be a pattern of “threes”. More radically, we might define composition in a group or monoid as an operation taking each ordered n -tuple of elements ($n \geq 0$) to its product, and formulate the associative law accordingly, letting the neutral element simply be the empty product, and the neutral-element law a special case of the associative law; and again, no “threes” would be apparent. As to the reason we develop the topics as we do, rather than in one of the above ways, I think this comes out of certain choices regarding pedagogy and notation that have evolved in Western mathematics, for better or worse, without anyone’s looking ahead at the form this would yield for such definitions. (On the other hand, I freely admit that my choice in §2.1 to motivate the idea of a free group with the 3-generator case was culturally influenced.)

Let me close this discussion by noting that many of the more complicated objects of mathematical study arise by combining one structure that fits, or partially fits, the pattern we have noted, with another. Thus, a lattice is a set with two *semilattice* structures that satisfy compatibility identities; a ring is given by an *abelian group*, together with a bilinear binary operation on this group under which it is a *monoid*.

The reader familiar with the definition of a Lie algebra over a commutative ring R (§8.7 below) will note similarly that it is an R -module (a concept which fits into the above pattern in the same way as that of G -set), with an R -bilinear operation, the Lie bracket, which satisfies the alternating identity (which tells *both* the result of bracketing an element with itself, and the relation between bracketings in opposite orders), and the Jacobi identity (which describes how the bracket of an element with the bracket of two others can be described in terms of the operations of bracketing with those elements successively).

Returning to the description of a ring as an abelian group given with a bilinear operation under which it is a monoid, it is interesting to note that various refinements of the concept of ring involve adding one (or more!) conditions that can be thought of as filling in the missing “middle slot” in the monoid structure, concerning how elements relate in opposite orders: a multiplicative inverse operation on nonzero elements gives a *division ring* structure; *commutativity* of multiplication determines the favorite class of rings of contemporary algebra; both together give the class of fields. Another important ring-theoretic concept which can be thought of in this way is that of an *involution* on a (not necessarily commutative) ring, that is, an abelian group automorphism $*$: $|R| \rightarrow |R|$ satisfying $x^{**} = x$ and $(xy)^* = y^*x^*$. The complex numbers have all three structures: multiplicative inverses, commutativity, and the involution of complex conjugation.

Finally, the concept of a closure operator has an important special case gotten by imposing an additional condition on “how elements relate in opposite orders”, the “exchange axiom”:

$$(5.4.1) \quad y \notin \text{cl}(X), y \in \text{cl}(X \cup \{z\}) \Rightarrow z \in \text{cl}(X \cup \{y\}) \quad (X \subseteq S, y, z \in S).$$

This is the condition which in the theory of vector spaces allows one to prove that bases have unique cardinalities, and in the theory of transcendental field extensions yields a similar result for transcendence bases. Closure operators satisfying (5.4.1) are called (among other names) *matroids*. Cf. [125], and for a ring-theoretic application, [44].

I do not attach great importance to the observations of this section. But I have noticed them for years, and thought this would be a good place to mention them.

5.5. Galois connections. Let us introduce this very general concept using the case from which it gets its name:

Galois theory deals with the situation where one is given a field F and a finite group G of automorphisms of F . Given any subset A of F , let A^* denote the set of elements of the group G fixing all elements of A , and given any subset B of G , likewise let B^* be the set of elements of the field F fixed by all members of B . It is not hard to see that in these situations, A^* is always a subgroup of G , and B^* a subfield of F . The Fundamental Theorem of Galois Theory says that the groups A^* give *all* the subgroups of G , and similarly that the sets B^* are *all* the fields between the fixed field of G in F and the whole field F , and gives further information on the relation between corresponding subgroups and subfields.

Some parts of the proof of this theorem use arguments specific to fields and their automorphism groups; but certain other parts can be carried out without even knowing what the words mean. For instance, the result, ‘‘If A is a set of elements of the field F , and A^{**} is the set of elements of F fixed by all automorphisms in G that fix all elements of A , then $A^{**} \supseteq A$ ’’ is clearly true independent of what is meant by a ‘‘field’’, an ‘‘automorphism’’, or ‘‘to fix’’!

This suggests that one should look for a general context to which the latter sort of arguments apply. Replacing the set of elements of our field F by an arbitrary set S , the set of elements of the group G by any set T , and the condition of elements of F being fixed by elements of G by any relation $R \subseteq S \times T$, we get the following set of observations:

Lemma 5.5.1. *Let S, T be sets, and $R \subseteq S \times T$ a relation. For $A \subseteq S, B \subseteq T$, let us write*

$$(5.5.2) \quad \begin{aligned} A^* &= \{t \in T \mid (\forall a \in A) aRt\} \subseteq T, \\ B^* &= \{s \in S \mid (\forall b \in B) sRb\} \subseteq S, \end{aligned}$$

thus defining two operations written $$, one from $\mathbf{P}(S)$ to $\mathbf{P}(T)$ and the other from $\mathbf{P}(T)$ to $\mathbf{P}(S)$. Then for $A, A' \subseteq S, B, B' \subseteq T$, we have*

- (i) $A \subseteq A' \Rightarrow A^* \supseteq A'^*$ $B \subseteq B' \Rightarrow B^* \supseteq B'^*$ ($*$ reverses inclusions),
- (ii) $A^{**} \supseteq A$ $B^{**} \supseteq B$ ($**$ is increasing),
- (iii) $A^{***} = A^*$ $B^{***} = B^*$ ($*** = *$),
- (iv) $** : \mathbf{P}(S) \rightarrow \mathbf{P}(S)$ and $** : \mathbf{P}(T) \rightarrow \mathbf{P}(T)$ are closure operators on S and T respectively.
- (v) The sets A^* ($A \subseteq S$) are precisely the closed subsets of T , and the sets B^* ($B \subseteq T$) are precisely the closed subsets of S with respect to these closure operators $**$.
- (vi) The maps $*$, restricted to closed sets, give an antiisomorphism (an order-reversing,

*equivalently, \vee -and- \wedge -interchanging, bijection) between the complete lattices of $**$ -closed subsets of S and of T .*

Proof. (i) and (ii) are immediate. We shall prove the remaining assertions from those two.

If we apply $*$ to both sides of (ii), so that the inclusions are reversed by (i), we get $A^{***} \subseteq A^*$, $B^{***} \subseteq B^*$; but if we put B^* for A and A^* for B in (ii) we get $B^{***} \supseteq B^*$, $A^{***} \supseteq A^*$. Together these inclusions give (iii). To get (iv), note that by (i) applied twice, the operators $**$ are inclusion-preserving, by (ii) they are increasing, and by applying $*$ to both sides of (iii) we find that they are idempotent. To get (v) note that by (iii) every set B^* respectively A^* is closed, and of course every closed set X has the form Y^* for $Y = X^*$. (vi) now follows from (v), (iii) and (i). \square

If for each $t \in T$ we consider the relation $-Rt$ as a condition satisfied by some elements $s \in S$, then for $A \subseteq S$ we can interpret A^{**} as ‘‘the set of elements of S which satisfy all conditions (of this sort) that are satisfied by the elements of A ’’. From this interpretation, the fact that $**$ is a closure operator is intuitively understandable.

Definition 5.5.3. *If S and T are sets, then a pair of maps $*$: $\mathbf{P}(S) \rightarrow \mathbf{P}(T)$ and $*$: $\mathbf{P}(T) \rightarrow \mathbf{P}(S)$ satisfying conditions (i) and (ii) of Lemma 5.5.1 (and hence the consequences (iii)-(vi)) is called a Galois connection between the sets S and T .*

Exercise 5.5:1. Show that every Galois connection between sets S and T arises from a relation R as in Lemma 5.5.1, and that this relation R is in fact unique.

Thus, a Galois connection on a pair of sets S, T can be characterized either abstractly, by Definition 5.5.3, or as a structure arising from some relation $R \subseteq S \times T$. In all naturally occurring cases that I know of, the relation R is what we start with, and the Galois connection is obtained from it. On the other hand, the characterization as in Definition 5.5.3 has the advantage that it can be generalized by replacing $\mathbf{P}(S)$ and $\mathbf{P}(T)$ by other partially ordered sets.

Here is another order-theoretic characterization of Galois connections:

Exercise 5.5:2. If S and T are sets, show that a pair of maps $*$: $\mathbf{P}(S) \rightarrow \mathbf{P}(T)$, $*$: $\mathbf{P}(T) \rightarrow \mathbf{P}(S)$ is a Galois connection if and only if for $X \subseteq S$, $Y \subseteq T$, one has

$$X \subseteq Y^* \Leftrightarrow Y \subseteq X^*.$$

More generally, you can show that given two partially ordered sets $(|P|, \leq)$ and $(|Q|, \leq)$, and a pair of maps $*$: $|P| \rightarrow |Q|$, $*$: $|Q| \rightarrow |P|$, these maps will satisfy conditions (i)-(ii) of Lemma 5.5.1 if and only if they satisfy the above condition (with ‘‘ \leq ’’ in place of ‘‘ \subseteq ’’ throughout).

Exercise 5.5:3. Show that for every closure operator cl on a set S , there exists a set T and a relation $R \subseteq S \times T$ such that the closure operator $**$ on S induced by R is cl . Can one in fact take for T any set given with any closure operator whose lattice of closed subsets is antiisomorphic to the lattice of cl -closed subsets of S ?

A Galois connection between two sets S and T becomes particularly valuable when the $**$ -closed subsets have characterizations of independent interest. Let us give a number of examples, beginning with the one that motivated our definition. (The reader should not worry if he or she is not familiar with all the concepts and results mentioned in these examples.) In describing these examples, I will for brevity often ignore the distinction between algebraic objects and their underlying sets.

Example 5.5.4. Take for S the underlying set of a field F , and for T the underlying set of a finite group G of automorphisms of F . For $a \in S$ and $g \in T$ let aRg mean that g fixes a , that is, $g(a) = a$. If we write $K \subseteq F$ for the subfield G^* , then, as noted earlier, the Fundamental Theorem of Galois Theory tells us that the closed subsets of F are precisely the subfields of F containing K , while the closed subsets of G are all the subgroups of G . One finds that properties of the field extension F/K are closely related to properties of the group G , and can be studied with the help of group theory ([29, Chapter V], [31, Chapter VI]). These further relations between group structure and field structure are not, of course, part of the general theory of Galois connections. That theory gives the underpinnings, over which these further results are built.

Example 5.5.5. Let us take for S a vector space over a field K , for T the dual space $\text{Hom}_K(S, K)$, and let us take xRf to mean $f(x) = 0$. In this case, one finds that the closed subsets of S are precisely all its vector subspaces, while those of T are the vector subspaces that are closed in a certain topology. In the finite-dimensional case, this topology is discrete, and so the closed subsets of T are all its subspaces. The resulting correspondence between subspaces of a finite-dimensional vector space and of its dual space is a basic tool which is taught (or should be!) in undergraduate linear algebra. Some details of the infinite-dimensional case are developed in an exercise below.

Example 5.5.6. A superficially similar example: Let $S = \mathbb{C}^n$ (complex n -space), $T = \mathbb{Q}[x_0, \dots, x_{n-1}]$, the polynomial ring in n indeterminates over the rationals, and let $(a_0, \dots, a_{n-1})Rf$ mean $f(a_0, \dots, a_{n-1}) = 0$. This case is the starting-point for classical algebraic geometry, and still the underlying inspiration for much of the modern theory. The closed subsets of \mathbb{C}^n are the solution-sets of systems of polynomial equations, while the Nullstellensatz says that the closed subsets of $T = \mathbb{Q}[x_0, \dots, x_{n-1}]$ are the ‘‘radical ideals’’.

Example 5.5.7. Let S be a finite-dimensional vector space over the real numbers \mathbb{R} , T the set of pairs (f, a) , where f is a linear functional on S and $a \in \mathbb{R}$, and define $xR(f, a)$ to mean $f(x) \leq a$. Then the closed subsets of S turn out to be the closed *convex* sets.

If we restrict a to the value 1, so that we can regard T simply as the dual space of S , and write xRf for the condition $f(x) \leq 1$, we get a Galois connection between S and its dual space, under which the closed subsets, on each side, are the closed convex subsets containing 0. For instance, if we take $S = \mathbb{R}^3$ and identify it with its dual via the natural inner product, we find that the dual of a cube centered at the origin is an octahedron centered at the origin. The regular dodecahedron and icosahedron are similarly dual to one another.

Example 5.5.8. Let M be an abelian group (or more generally, a module over a commutative ring k), and $S = T =$ the ring of endomorphisms of M (as an abelian group, respectively a k -module). Let sRt denote the condition $st = ts$. It is easy to verify that the subsets of $S = T$ closed under the resulting Galois connection are certain subrings (respectively k -subalgebras). For every subring X , the subring X^* is called by ring-theorists the *commutant* of X . If, in this situation, we regard M as an X -module, then X^* is the ring (respectively, k -algebra) of X -module endomorphisms of M . The ring $X^{**} \supseteq X$, the commutant of the commutant, is called the *bicommutant* of X .

Example 5.5.9. Let S be a set of mathematical objects, T a set of propositions about an object of this sort, and sRt the relation “the object s satisfies the proposition t ”; in logician’s notation, $s \models t$. Then the closed subsets of S are those classes of objects definable by sets of propositions from T , which model theorists call *axiomatic classes*, while the closed subsets of T are what they call *theories*. The theory B^{**} generated by a set B of propositions consists of those members of T that are *consequences* of the propositions in B , in the sense that they hold in all members of S satisfying the latter.

(Actually, in the naturally occurring cases of this example, S is often a proper class rather than a set of mathematical objects; e.g., the class of all groups. We will see how to deal comfortably with such situations in the next chapter.)

There are, of course, cases where it is preferable to use symbols other than “ $*$ ” for the operators of a Galois connection. In Example 5.5.5, it is usual to write the set obtained from a set A as $\text{Ann}(A)$ or A^0 or A^+ (the *annihilator* or *null space* of A) because “ $*$ ” is commonly used for the dual space. More seriously, whenever $S = T$ but R is not a symmetric relation on S , the two constructions $\{s' \mid (\forall s \in A) s'Rs\}$ and $\{s' \mid (\forall s \in A) sRs'\}$ will be distinct, so one must denote them by different symbols, such as A^* and A_* . An example of such a case is

Exercise 5.5:4. (i) If $S = T = \mathbb{Q}$, the set of rational numbers, and R is the relation \leq , characterize the two systems of closed subsets of \mathbb{Q} . Describe in as simple a way as possible the structure of the lattices of closed sets.

(ii) Same question with “ $<$ ” in place of “ \leq ”.

The next exercise gives, as promised, some details on the infinite-dimensional case of Example 5.5.5. The one following it is related to Example 5.5.8.

Exercise 5.5:5. Let K be a field, S a K -vector-space, and T its dual space.

(i) Show that the subsets of S closed under the Galois connection of Example 5.5.5 are indeed all the vector subspaces of S .

To characterize the subsets of T closed under this connection, let us, for each $s \in S$ and $c \in K$, define $U_{s,c} = \{t \in T \mid t(s) = c\}$, and topologize T by making the $U_{s,c}$ a subbasis of open sets.

(ii) Show that the resulting topology is the weakest such that for each $s \in S$, the evaluation map $t \mapsto t(s)$ is a continuous map from T to the discrete topological space K .

(iii) Show that the subsets of T closed under the Galois connection described above are the vector subspaces of T closed in the above topology.

(There is an elegant characterization of the class of topological vector spaces that arise in this way. They are called *linearly compact* vector spaces. See [92, Chapter II, 27.6 and 32.1], or for a summary, [2, first half of §24].)

Exercise 5.5:6. Let M be the underlying abelian group of the polynomial ring $\mathbb{Q}[t]$ in one indeterminate t , let $x: M \rightarrow M$ be the abelian group endomorphism given by *multiplication by t* , and $d: M \rightarrow M$ the endomorphism given by *differentiation with respect to t* . Find the commutant and bicommutant (as defined in Example 5.5.8) of each of the following subrings of $\text{End}(M)$:

(i) $\mathbb{Z}[x]$.

(ii) $\mathbb{Z}[x^2, x^3]$.

(iii) $\mathbb{Z}[d]$.

(iv) $\mathbb{Z}\langle x, d \rangle$ (the ring generated by x and d . Angle brackets are used to indicate generators of not necessarily commutative rings.)

Exercise 5.5:7. If G is a group and X a subset of G , then $\{g \in G \mid (\forall x \in X) gx = xg\}$ is called the *centralizer* of X in G , often denoted $C_G(X)$. This is easily seen to be a subgroup of G .

- (i) Show that if H is a subgroup of a group G then the following conditions are equivalent:
 (a) H is commutative, and is the centralizer of its centralizer. (b) H is the intersection of some nonempty family of maximal commutative subgroups of G .
 (ii) Give a result about Galois connections of which the above is a particular case.

(You may either state and prove in detail the result of (i), and then for (ii) formulate a general result which can clearly be proved the same way, in which case you need not repeat the argument; or do (ii) in detail, then note briefly how to apply your result to get (i).)

We recall that for a general closure operator on a set S , the union of two closed subsets of S is not in general closed; their join in the lattice of closed sets is the *closure* of this union. However, if we consider the Galois connection between a set of objects and a set of propositions, and if these propositions are the sentences in a language that contains the operator \vee (“or”), then the set of objects satisfying the proposition $s \vee t$ will be precisely the union of the set of objects satisfying s and the set satisfying t :

$$\{s \vee t\}^* = \{s\}^* \cup \{t\}^*.$$

Likewise, if the language contains the operator \wedge (“and”), then

$$\{s \wedge t\}^* = \{s\}^* \cap \{t\}^*.$$

In fact, the choice of the symbols \vee and \wedge (modifications of \cup and \cap) by logicians to represent these operators was probably suggested by these properties of the sets of objects satisfying such relations. (At least, so I thought when I wrote this. But a student told me he had heard a different explanation: that \vee is an abbreviation of Latin *vel* “or”, and \wedge was formed by inverting it. If so, \cup and \cap may have been created as modifications of \vee and \wedge , and the fact that \cup looks like the first letter of “union” may be a coincidence.)

If we look at closed sets of propositions rather than closed sets of objects, these are, of course, ordered in the reverse fashion: The set of propositions implied by a proposition $s \vee t$ is the *intersection* of those implied by s and those implied by t , while the set implied by $s \wedge t$ is the *closure of the union* of the sets implied by s and by t . Thus the use of the words “and” (which implies something “bigger”) and “or” (which suggests a weakening) is based on the proposition-oriented viewpoint, while the choice of symbols \wedge and \vee corresponds to the object viewpoint.

The conflict between these two viewpoints explains the problem students in precalculus courses have when they are asked, say, to describe by inequalities the set of real numbers x satisfying $x^2 \geq 1$. We want the answer “ $x \leq -1$ or $x \geq 1$ ”, meaning $\{x \mid x \leq -1 \text{ or } x \geq 1\}$. But they often put “ $x \leq -1$ and $x \geq 1$ ”. What they have in mind could be translated as “ $\{x \mid x \leq -1\}$ and $\{x \mid x \geq 1\}$ ”. We can hardly tell them that their difficulty arises from the order-reversing nature of the Galois connection between propositions and objects! But the more thoughtful students might be helped if, without going into the formalism, we pointed out that there is a kind of “reverse relation” between statements and the things they refer to: the larger a set of statements, the smaller the set of things satisfying it; the larger a set of things, the smaller the set of statements they all satisfy; so that “and” for sets of real numbers translates to “or” among formulas defining them.

I point out this “reverse relation” in a handout on set theory and mathematical notation that I give out in my upper division courses [46, in particular, §2]. Whether it helps, I don’t know.

Logicians often write the propositions $(\forall x \in X) P(x)$ and $(\exists x \in X) Q(x)$ as $\bigwedge_{x \in X} P(x)$

and $\bigvee_{x \in X} Q(x)$. Here the universal and existential quantifications are being represented as (generally infinite) conjunctions and disjunctions, corresponding to intersections and unions respectively of the classes of models defined by the given families of conditions $P(x)$ and $Q(x)$, as x ranges over X .

We have noted that for many naturally arising types of closure operators cl , the closure of a set X can be constructed both “from above” and “from below” – either by taking the intersection of all closed sets containing X , or by “building” elements of $\text{cl}(X)$ from elements of X by iterating some procedure in terms of which cl was defined. Closure operators determined by Galois connections, however, are born with only a construction “from above”: For $X \subseteq S$, X^{**} is the intersection of those sets $\{t\}^*$ ($t \in T$) which contain X ; the definition of a Galois connection does not provide any way of constructing this set “from below”. Rather, this is a recurring type of mathematical problem for the particular Galois connections of mathematical interest! Typically, given such a Galois connection, one looks for operations that all the sets $\{t\}^*$ are closed under, and when one suspects one has found enough of these, one seeks to prove that for every X , the set X^{**} is the closure of X under these operations. For instance, the fixed set of an automorphism of a field extension F/K is easily seen to contain all elements of K and to be closed under the field operations; the Fundamental Theorem of Galois Theory says that under appropriate hypotheses, the closed subsets of F are precisely the subsets closed under these operations. When one considers mathematical objects and propositions, then the problem of finding a way to “build up” the closure of a set of propositions is that of finding an adequate set of *rules of inference* for the type of proposition under consideration, while to construct the closure operator on objects is to characterize intrinsically the axiomatic model classes.

The definition of Galois connection is unfortunately seldom presented in courses, and many mathematicians who discover examples of it have not heard of the general concept. Of course, Lemma 5.5.1 is a set of easy observations which can be verified in any particular case without referring to a general result. But it is useful to have the general concept as a guide, and once one proves the lemma, one can skip those trivial verifications from then on.