

Chapter 3. A Cook's tour of other universal constructions.

We shall now examine a number of other constructions having many similarities to that of free groups. In each case, the construction can be motivated by a question of the form, “Suppose we have a structure about which we know only that it satisfies such and such conditions. How much can we say about it based on this information alone?” In favorable cases, we shall find that if we collect the “data” we can deduce about such an object, this data itself can be made into an object F , which satisfies the given conditions, and satisfies no relations not implied by them (cf. remark 2.2.10). This F is then a “universal” example of these conditions, and that fact can be translated into a “universal mapping property” for F .

Although the original “What can we say about such an object?” question, and the “least set of relations” property, are valuable as motivation and intuition, the universal mapping property gives the characterization of these constructions that is most useful in applications. So, though I will sometimes refer explicitly to those motivating ideas, and other times leave them for you to see, we will always characterize our constructions by universal properties.

The existence of these universal objects may in most cases be proved from scratch by either of the methods of §§2.2 and 2.3: construction from below, as sets of terms modulo necessary identifications, or construction from above, as subobjects of big direct products. But often we will, alternatively, be able to combine previously described universal constructions to get the new one.

Where possible, we will get explicit information on the structure of the new object – a normal form or other such description. It is a mark of the skilled algebraist, when working with objects defined by universal properties, to know when to use the universal property, and when to turn to an explicit description.

As we move through this chapter, I shall more and more often leave standard details for the reader to fill in: the precise meaning of an object “universal for” a certain property, the verification that such an object exists, etc.. In the later sections, commutative diagrams illustrating universal properties will often be inserted without explanation. These diagrams are not substitutes for assertions, but aids to the reader in visualizing the situation of the assertion he or she needs to formulate.

Constructions of *groups* will receive more than their rightful share of attention here because groups give a wide range of interesting examples, and are more familiar to many students than lattices, noncommutative rings (my own love), Lie algebras, etc..

Let us begin by noting how some familiar elementary group-theoretic constructions can be characterized by universal properties.

3.1. The subgroup and normal subgroup of G generated by $S \subseteq |G|$. Suppose we are explicitly given a group G , and a subset S of $|G|$.

Consider a subgroup A of G about which we are told only that it contains all elements of the set S . How much can we say about A ?

Clearly A contains all elements of G that can be obtained from the elements of S by repeated formation of products and inverses, and also contains the neutral element. This is all we can deduce, for it is easy to see that the set of elements which can be so obtained will form the underlying set of a subgroup of G , namely the subgroup $\langle S \rangle$ generated by the set S . This description builds $\langle S \rangle$ up “from below”. We can also obtain it “from above” as the intersection of all subgroups of G containing S . Whichever way we obtain it, the defining universal property

of $\langle S \rangle$ is that it is a subgroup which contains S , and is contained in every subgroup A of G that contains S :

$$\begin{array}{ccc} S \subseteq & |\langle S \rangle| & \langle S \rangle \\ \supseteq & \cap & \cap \\ & |A| & A \end{array}$$

(In the second part of the above display, we symbolize the group homomorphism given by an inclusion map of underlying sets using an inclusion sign between the symbols for the groups; a slight abuse of notation.)

We know a somewhat better description of the elements of $\langle S \rangle$ than the one I just gave: Each such element is either e or the product of a sequence of elements of S and their inverses. A related observation is that $\langle S \rangle$ is the image of the map of the free group F on S into G induced by the inclusion-map $S \rightarrow |G|$. In particular cases one may get still better descriptions. For instance, if $S = \{a, b, c\}$ and a, b and c commute, then $\langle S \rangle$ consists of all elements $a^m b^n c^p$; if G is the additive group of integers, then the subgroup generated by $\{1492, 1974\}$ is the subgroup of all even integers; if G is the permutation group S_n and S consists of the two permutations (12) and $(12\dots n)$, then $\langle S \rangle$ is all of G .

There is likewise a least *normal* subgroup of G containing S . This is called “the normal subgroup of G generated by S ”, and has the corresponding universal property, with the word “normal” inserted before “subgroup”.

Exercise 3.1:1. Show that the normal subgroup $N \subseteq G$ generated by S is the subgroup of G generated by $\{gsg^{-1} \mid g \in |G|, s \in S\}$.

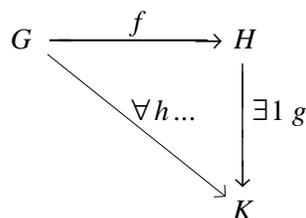
Can $|N|$ also be described as $\{ghg^{-1} \mid g \in |G|, h \in |\langle S \rangle|\}$?

Exercise 3.1:2. Let G be the free group on two generators x and y , and n a positive integer. Show that the normal subgroup of G generated by x^n and y is generated as a subgroup by x^n and $\{x^i y x^{-i} \mid 0 \leq i < n\}$, and is in fact a *free* group on this $(n+1)$ -element set. Also describe the normal subgroup we get if we let $n = 0$.

3.2. Imposing relations on a group; quotient groups. Suppose next that we are given a group G , and are interested in homomorphisms of G into other groups, $f: G \rightarrow H$, which make certain specified pairs of elements fall together. That is, let us be given a family of pairs of elements $\{(x_i, y_i) \mid i \in I\} \subseteq |G| \times |G|$ (perhaps only one pair, (x, y)) and consider homomorphisms f from G into other groups, which satisfy

$$(3.2.1) \quad (\forall i \in I) f(x_i) = f(y_i).$$

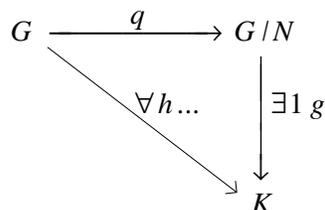
Note that given one homomorphism $f: G \rightarrow H$ with this property, we can get more such homomorphisms $G \rightarrow K$ by forming composites gf of f with arbitrary homomorphisms $g: H \rightarrow K$. It would be nice to know whether there exists one pair (H, f) which satisfies (3.2.1) and is *universal* for this condition, in the sense that given any other pair (K, h) satisfying it, there is a unique homomorphism $g: H \rightarrow K$ making the diagram below commute. (In that diagram, “ $\forall h \dots$ ” is short for “For all homomorphisms $h: G \rightarrow K$ such that $(\forall i \in I) h(x_i) = h(y_i)$ ”.)



It is not hard to prove the existence of such a universal pair directly, either by a “group-theoretic terms modulo an equivalence relation” construction, as in §2.2, or by an “image in a big direct product” construction, analogous to that of §2.3. But let us look at the problem another way. Condition (3.2.1) is clearly equivalent to

$$(3.2.2) \quad (\forall i \in I) f(x_i y_i^{-1}) = e.$$

So we are looking for a universal homomorphism which annihilates (sends to e) a certain family of elements of $|G|$. We know that the set of elements annihilated by a group homomorphism is always a normal subgroup, so this is equivalent to saying that f should annihilate the normal subgroup of G generated by $\{x_i y_i^{-1} \mid i \in I\}$, referred to at the end of the preceding section. And in fact, the pair $(G/N, q)$, where N is this normal subgroup, G/N is the quotient group, and $q: G \rightarrow G/N$ is the quotient map, has precisely the universal property we want:



So this quotient group is the solution to our problem.

If we had never seen the construction of the quotient of a group by a normal subgroup, an approach like the above would lead to a motivation of that construction. We would ask, “What do we know about a group H , given that it has a homomorphism of G into it satisfying (3.2.1)?” We would observe that H contains an image $f(a)$ of each $a \in G$, and that two such images are equal *if* they belong to the same coset of the normal subgroup generated by the $x_i y_i^{-1}$ ’s. We would discover how the group operations must act on these images-of-cosets, and conclude that this set of cosets, under these operations, was itself a universal example of this situation.

Let us assume even a little more naiveté in

Exercise 3.2:1. Suppose in the above situation that we had not been so astute, and had only noted that $f(a) = f(b)$ must hold in H whenever ab^{-1} lies in the *subgroup* generated by $\{x_i y_i^{-1}\}$. Attempt to describe the group operations on the set of equivalence classes under this relation, show where this description fails to be well-defined, and show how this “failure” could lead us to discover the *normality* condition needed.

The construction we have described is called *imposing the relations* $x_i = y_i$ ($i \in I$) on G . We can abbreviate the resulting group by $G/(x_i = y_i \mid i \in I)$.

For the next exercise, recall that if G is a group, then a G -set means a pair $S = (|S|, m)$, where $|S|$ is a set and $m: |G| \times |S| \rightarrow |S|$ is a map, which we shall abbreviate by writing $m(g, s) = gs$, satisfying

$$(3.2.3) \quad (\forall g, g' \in |G|, s \in |S|) \quad g(g's) = (gg')s,$$

$$(\forall s \in |S|) \quad es = s,$$

in other words, a set on which G acts by permutations [31, §I.5], [29, §II.4] [26, §1.7]. (We remark that this structure on the set $|S|$ can be described in two other ways: as a homomorphism from G to the group of permutations of $|S|$, and alternatively, as a system of unary operations \bar{g} on $|S|$, one for each $g \in |G|$, satisfying *identities* corresponding to all the *relations* holding in G .)

A *homomorphism* $S \rightarrow S'$ of G -sets (for a *fixed* group G) means a map $a: |S| \rightarrow |S'|$ satisfying

$$(3.2.4) \quad (\forall g \in |G|, s \in |S|) \quad a(gs) = ga(s).$$

If H is a subgroup of the group G , let $|G/H|$ denote the set of left cosets of H in G . We shall write a typical left coset as $[g] = gH$. Then $|G/H|$ can be made the underlying set of a left G -set G/H , by defining $g[g'] = [gg']$.

Exercise 3.2:2. Let H be any subgroup of G . Find a universal property characterizing the pair $(G/H, [e])$. In particular, what form does this universal mapping property take in the case where $H = \langle x_i^{-1}y_i \mid i \in I \rangle$ for some set $\{(x_i, y_i) \mid i \in I\} \subseteq |G| \times |G|$?

With the concept of imposing relations on a group under our belts, we are ready to consider

3.3. Groups presented by generators and relations. To start with a concrete example, suppose we are curious about groups G containing two elements a and b satisfying the relation

$$(3.3.1) \quad ab = b^2a.$$

One may investigate the consequences of this equation with the help of the group laws. What we would be investigating is, I claim, the structure of the group with a *universal* pair of elements satisfying (3.3.1).

More generally, let X be a set of symbols (in the above example, $X = \{a, b\}$), and let T be the set of all group-theoretic terms in the elements of X . Then formal group-theoretic *relations* in the elements of X mean formulae “ $s = t$ ”, where $s, t \in T$. Thus, given any set $R \subseteq T \times T$ of pairs (s, t) of terms, we may consider groups H with X -tuples of elements $v: X \rightarrow |H|$ satisfying the corresponding set of relations

$$(3.3.2) \quad (\forall (s, t) \in R) \quad s_v = t_v.$$

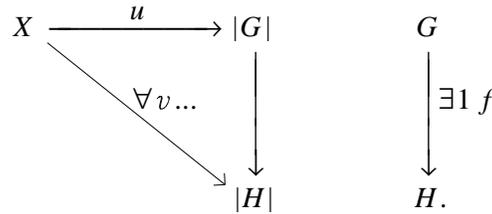
(For example, when $X = \{a, b\}$ and R is the singleton $\{(ab, b^2a)\}$, (3.3.2) becomes (3.3.1).) In this situation, we have

Proposition 3.3.3. *Let X be a set, T the set of group-theoretic terms in X , and R a subset of $T \times T$. Then there exists a universal example of a group with an X -tuple of elements satisfying the relations “ $s = t$ ” ($(s, t) \in R$). I.e., there exists a pair (G, u) , where G is a group, and u a map $X \rightarrow |G|$ such that*

$$(\forall (s, t) \in R) \quad s_u = t_u,$$

and such that for any group H , and any X -tuple v of elements of H satisfying (3.3.2), there exists a unique homomorphism $f: G \rightarrow H$ satisfying $v = fu$ (in other words, having the property

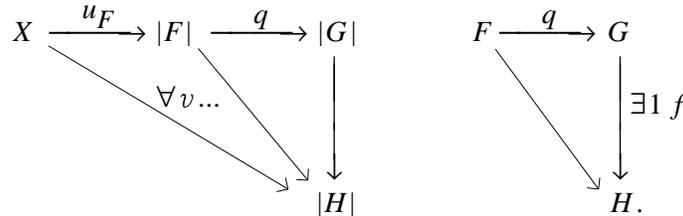
that the X -tuple v of elements of H is the image under f of the X -tuple u of elements of G).



Further, the pair (G, u) is determined up to canonical isomorphism by these properties, and the group G is generated by $u(X)$.

Three Methods of Proof. Two methods that we may use are the constructions of §2.2 and §2.3, applied essentially word-for-word, with the further condition (3.3.2) added to the group axioms throughout. (Note that unlike (2.2.1-8), the set of equations (3.3.2) involves no universal quantification over T , since we only require the family of relations R to hold for the particular X -tuple v of elements of each group H .)

However, we can now, alternatively, get the pair (G, u) with less work. Let (F, u_F) be the free group on X , let N be the normal subgroup of F generated by $\{s_{u_F} t_{u_F}^{-1} \mid (s, t) \in R\}$, i.e., by the set of elements of F that we want to annihilate. Let $G = F/N$, let $q: F \rightarrow F/N$ be the canonical map, and let $u = q u_F$. That (G, u) has the desired universal property follows immediately from the universal properties of free groups and quotient groups.



Having constructed (G, u) by any of these methods, let us now prove our uniqueness assertion. If (G', u') is another pair with the same universal property, then by the universal property of G there exists a homomorphism $i: G \rightarrow G'$ such that $i u = u'$, and by the universal property of G' , an $i': G' \rightarrow G$ such that $i' u' = u$. These are inverses of one another; indeed, note that $i' i u = u$, hence by the uniqueness condition in the universal property of G , $i' i$ equals the identity map of G ; by a like argument, $i i'$ is the identity of G' , so i is invertible, and gives the asserted isomorphism.

That G is generated by $u(X)$ can be seen from each of our constructions, but let us also show from the universal property that this *must* be so. Consider the subgroup $\langle u(X) \rangle$ of G generated by $u(X)$. The universal property of G gives a homomorphism $j: G \rightarrow \langle u(X) \rangle$ which is the identity on elements of $u(X)$. Following it by the inclusion of $\langle u(X) \rangle$ in G yields an endomorphism of G which agrees with the identity map on $u(X)$, and so, by the uniqueness assertion in the universal property, is the identity. So the inclusion of $\langle u(X) \rangle$ in G is surjective, as desired. \square

Though we implied above that the advantage of getting our construction by combining two known constructions was that this was less work than constructing it from scratch, an important general advantage to finding that a new construction can be obtained from known constructions is that one can apply results proved about the known constructions to the new one.

The group G of the preceding proposition is called the group *presented* by the *generators* X and *relations* R . A common notation for this is

$$(3.3.4) \quad G = \langle X \mid R \rangle.$$

For example, the universal group with a pair of elements satisfying (3.3.1) would be written

$$\langle a, b \mid ab = b^2a \rangle.$$

In a group presented by generators and relations (3.3.4), one often uses the same symbols for the elements of X and their images in $|G|$, even if the map u is not one-to-one. For instance, from the well-known lemma saying that if an element η of a group (or monoid) has both a left inverse ξ and a right inverse ζ , then $\xi = \zeta$, it follows that in the group $\langle x, y, z \mid xy = e = yz \rangle$, one has $u(x) = u(z)$. Unless there is a special need to be more precise, we may express this by saying “in $\langle x, y, z \mid xy = e = yz \rangle$, one has $x = z$ ”.

Recall that the concept of group-theoretic term was introduced both for the consideration of what relations hold among all families of elements of all groups, and to write specific relations that hold among particular families of elements in particular groups. For the purpose of discussing identities holding in all groups, it was necessary to distinguish between expressions such as $(xy)^{-1}$ and $y^{-1}x^{-1}$, between $(xy)z$ and $x(yz)$, etc.. But in considering relations in particular groups we can generally take for granted the group identities, i.e., not distinguish pairs of expressions that have the same evaluations in all groups. For example, in (3.3.1), the right hand side could be replaced by $b(ba)$ without changing the effect of the condition. Hence in considering groups presented by generators and relations, one often considers the relations to be given by pairs, not of *terms*, but of their equivalence classes under the relation of having equal values in all groups – in other words pairs $(s, t) \in |F| \times |F|$, where F is the free group on X . For such (s, t) , an X -tuple v of elements of a group G is considered to “satisfy $s = t$ ” if $h(s) = h(t)$, for h the homomorphism $F \rightarrow G$ induced by v as in Definition 2.1.3.

Whether s and t are group-theoretic terms as in Proposition 3.3.3, or elements of a free group as in the above paragraph, we should note that there is a certain abuse of language in saying that a family v of elements of a group G “satisfies the relation $s = t$ ”, and in writing equations “ $s = t$ ” in presentations of groups. What we mean in such cases is that a certain equation *obtained from* the pair (s, t) and the X -tuple v holds in G ; but the equality $s = t$ between terms or free group elements is itself generally false! As with other convenient but imprecise usages, once we are conscious of its impreciseness, we may use it, but should be ready to frame more precise statements when imprecision could lead to confusion (for instance, if we also want to discuss which of certain terms or elements of a free group are really equal).

We have noted that a relation (s, t) is satisfied by an X -tuple v of elements of a group G if and only if $(st^{-1})_v = e$ in G ; in other words, if and only if the relation (st^{-1}, e) is satisfied by v . Thus, every presentation of a group can be reduced to one in which the relations occurring all have the form (r, e) for terms (or free-group elements) r . The elements r are then called the *relators* in the presentation, and in expressing the group, one may list relators rather than relations. E.g., the group we wrote earlier as $\langle a, b \mid ab = b^2a \rangle$ would, in this notation, be written $\langle a, b \mid aba^{-1}b^{-2} \rangle$. However, I will stick to the former notation in these notes.

Exercise 3.3:1. Show that the three groups described below are isomorphic (as groups, ignoring the maps “ $X \rightarrow |G|$ ” coming from the presentations of the first two).

- (i) $G = \langle a, b \mid a^2 = e, ab = b^{-1}a \rangle$.
- (ii) $H = \langle s, t \mid s^2 = t^2 = e \rangle$.

(iii) The group of all distance-preserving permutations of the set \mathbb{Z} , i.e., the group consisting of all translation-maps $n \mapsto n+c$ ($c \in \mathbb{Z}$) and all reflection-maps $n \mapsto -n+d$ ($d \in \mathbb{Z}$).

The universal property of a group presented by generators and relations is extremely useful in considerations such as that of

Exercise 3.3:2. Find all endomorphisms of the group of the preceding exercise. Describe the structure of the monoid of these endomorphisms.

Returning to the example with which we started this section –

Exercise 3.3:3. Find a normal form or other convenient description for the group presented by two generators a, b and the one relation (3.3.1): $ab = b^2a$.

The following question, suggested by a member of the class some years ago, is harder, but has a nice solution:

Exercise 3.3:4. (D. Hickerson.) Do the same for $\langle a, b \mid ab = b^2a^2 \rangle$.

Any group G can be presented by some system of generators and relations. E.g., take $|G|$ itself for generating set, and the multiplication table of G as a set of relations. But it is often of interest to find concise presentations for given groups. Note that the *free* group on a set X may be presented by the generating set X and the empty set of relations!

Exercise 3.3:5. Suppose $f(x, y)$ and $g(y)$ are group-theoretic terms in two and one variables respectively. What can you prove about the group with presentation

$$\langle w, x, y \mid w = f(x, y), x = g(y) \rangle?$$

Generalize if you can.

Exercise 3.3:6. Consider the set $\mathbb{Z} \times \mathbb{Z}$ of “lattice points” in the plane. Let G be the group of “symmetries” of this set, i.e., maps $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z} \times \mathbb{Z}$ which preserve distances between points.

(i) Find a simple *description* of G . (Cf. the description of the group of symmetries of the set \mathbb{Z} in terms of translations and reflections in Exercise 3.3:1(iii).)

(ii) Find a simple *presentation* for G .

(iii) Find a *normal form* for elements of G , in terms of the generators used in your presentation.

Exercise 3.3:7. Suppose G is a group of n elements. Then the observation made above, on how to present any group by generators and relations, yields upper bounds on the minimum numbers of generators and relations needed to present G . Write down these bounds; then see to what extent you can improve on them.

The above observations show that every *finite* group is *finitely presented*, i.e., has a presentation in terms of finitely many generators and finitely many relations. Of course, there are also finitely presented groups which are infinite. The next two exercises, of which the first is not difficult, while the second requires some ingenuity or experience with infinite groups, concern this property of finite presentability.

Exercise 3.3:8. Show that if G is a group which has some finite presentation, and if $\langle x_1, \dots, x_n \mid R \rangle$ is any presentation of G using finitely many generators, then there is a finite subset $R_0 \subseteq R$ such that $\langle x_1, \dots, x_n \mid R_0 \rangle$ is also a presentation of G .

Exercise 3.3:9. Find a finitely generated group that is not finitely presented.

Another kind of question one can ask is typified by

Exercise 3.3:10. Is the group

$$\langle x, y \mid xyx^{-1} = y^2, yxy^{-1} = x^2 \rangle$$

trivial ($= \{e\}$)? What about

$$\langle x, y \mid xyx^{-1} = y^2, yxy^{-1} = x^3 \rangle?$$

(If you prove either or both of these groups trivial, you should present your calculation in a way that makes it clear at each stage which defining relation you are applying, and to what part of what expression.)

For the group-theory buff, here are two harder, but still tractable examples.

Exercise 3.3:11. (J. Simon [99].) (i) Is either of the groups

$$\langle a, b \mid (b^{-1}a)^4 a^{-3} = e = b^{10}(b^{-1}a)^{-3} \rangle \quad \text{or} \quad \langle a, b \mid (ba^{-1})^{-3} a^{-2} = e = b^9(ba^{-1})^4 \rangle$$

trivial?

(ii) In the group $\langle a, b \mid ba^{-4}bab^{-1}a = e \rangle$, is the subgroup generated by $ba(b^{-1}a)^2$ and a^3b^{-1} free abelian?

Let us observe a consequence of the universal property of Proposition 3.3.3, characterizing the group G with presentation $\langle X \mid R \rangle$: For any group H , the set of homomorphisms $\text{Hom}(G, H)$ is in natural one-to-one correspondence with the set of X -tuples of elements of H satisfying the relations R .

For instance, if n is a positive integer, observe that $\langle x \mid x^n = e \rangle$ is \mathbb{Z}_n , the cyclic group of order n ; it follows that for any group H , we get a natural bijection between $\text{Hom}(\mathbb{Z}_n, H)$ and $\{a \in |H| \mid a^n = e\}$, each a in the latter set corresponding to the unique homomorphism $\mathbb{Z}_n \rightarrow H$ carrying x to a . (Terminological note: A group element $a \in |H|$ which satisfies $a^n = e$ is said to have *exponent* n . This is equivalent to its having order *dividing* n .)

Similarly, one finds that $\langle x, y \mid xy = yx \rangle$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$, hence $\text{Hom}(\mathbb{Z} \times \mathbb{Z}, H)$ corresponds to the set of all ordered pairs of commuting elements of H .

Thus, presentations of groups by generators and relations provide a bridge between the internal structure of groups, and their “external” behavior under homomorphisms. This will be of particular importance when we turn to category theory, which treats mathematical objects in terms of the homomorphisms among them.

The last exercise of this section describes one of my favorite groups, though most of its interesting properties cannot be given here.

Exercise 3.3:12. Let $G = \langle x, y \mid y^{-1}x^2y = x^{-2}, x^{-1}y^2x = y^{-2} \rangle$.

(i) Find a normal form or other convenient description for elements of G . Verify from this description that G has no nonidentity elements of finite order.

(ii) Calling the group characterized in several ways in Exercise 3.3:1 “ D ”, show that G has exactly three normal subgroups N such that $G/N \cong D$, and that the intersection of these three subgroups is $\{e\}$.

(iii) It follows from (ii) above that G can be identified with a subgroup of $D \times D \times D$. Give a criterion for an element of $D \times D \times D$ to lie in this subgroup, and prove directly from this criterion that no element of this subgroup has finite order.

An interesting text on groups presented by generators and relations, which assumes only an undergraduate background but goes deep into the techniques of the subject, is [30]. This general area is often called “combinatorial group theory”, and there are several books with that title. There is also a web-page on group theory, [103], with a list of open questions, including a section

on questions in combinatorial group theory.

Though group presentations *often* yield groups for which a normal form can be found, it has been proved by Novikov, Boone and Britton that there exist finitely presented groups G such that no algorithm can decide whether an arbitrary pair of terms of G represent the same element. A proof of this result is given in the last chapter of [32].

3.4. Abelian groups, free abelian groups, and abelianizations. An *abelian group* is a group A satisfying the further identity

$$(\forall x, y \in |A|) \quad xy = yx.$$

The discussion of §2.1 carries over without essential change and gives us the concept of a *free abelian group* (F, u) on a set X ; the method of §2.2 establishes the existence of such groups by constructing them as quotients of sets T of terms by appropriate equivalence relations, and the method of §2.3 yields an alternative construction as subgroups of direct products of large enough families of abelian groups. We may clearly also obtain the free abelian group on a set X as the group presented by the generating set X and the relations $st = ts$, as s and t range over all elements of T . This big set of relations is easily shown to be equivalent, for any X -tuple of elements of any group, to the smaller family $xy = yx$ ($x, y \in X$), so the free abelian group on X may be presented as

$$\langle X \mid xy = yx \ (x, y \in X) \rangle.$$

To investigate the *structure* of free abelian groups, let us consider, say, three elements a, b, c of an arbitrary abelian group A , and look at elements $g \in A$ that can be obtained from these by group-theoretic operations. We know from §2.4 that any such g may be written either as e , or as a product of the elements $a, a^{-1}, b, b^{-1}, c, c^{-1}$. We can now use the commutativity of A to rearrange this product so that it begins with all factors a (if any), followed by all factors a^{-1} (if any), then all factors b (if any), etc.. Now performing cancellations if both a and a^{-1} occur, or both b and b^{-1} occur, or both c and c^{-1} occur, we can reduce g to an expression $a^i b^j c^k$, where i, j and k are integers (positive, negative, or 0; exponentiation by negative integers and 0 being defined by the usual conventions). Let us call the set of such expressions $T_{\text{ab-red}}$, and define composition and inverse operations and an identity element on this set by

$$\begin{aligned} (a^i b^j c^k) \odot (a^{i'} b^{j'} c^{k'}) &= a^{i+i'} b^{j+j'} c^{k+k'}, \\ (a^i b^j c^k)^{(-)} &= a^{-i} b^{-j} c^{-k}, \\ e &= a^0 b^0 c^0. \end{aligned}$$

Note that \odot and $(-)$ are here different operations from those represented by the same symbols in §2.4, but that the idea is as in that section; in particular, it is clear that for any map v of $\{a, b, c\}$ into an *abelian* group, one has $(s \cdot t)_v = (s \odot t)_v$ and $(s^{-1})_v = (s^{(-)})_v$. It is now easy to verify that under these operations, $T_{\text{ab-red}}$ itself forms an abelian group F . This verification does not require any analog of “van der Waerden’s trick” (§2.4); rather, the result follows from the known fact that the integers form an abelian group under $+$, $-$, and 0 .

It follows as in §2.4 that this F is the *free* abelian group on $\{a, b, c\}$, and thus that the set $T_{\text{ab-red}}$ of terms $a^i b^j c^k$ is a normal form for elements of the free abelian group on three generators.

The above normal form is certainly simpler than that of the free *group* on $\{a, b, c\}$. Yet there

is a curious way in which it is more complicated: It is based on our choice of “alphabetic order” for the generating set $\{a, b, c\}$. Using different orderings, we get different normal forms, e.g., $b^j c^k a^i$, etc.. If we want to generalize our normal form to the free abelian group on a finite set X without any particular structure, we must begin by ordering X , say writing $X = \{x_1, x_2, \dots, x_n\}$. Only then can we speak of “the set of all expressions $x_1^{i_1} \dots x_n^{i_n}$ ”. If we want a normal form in the free abelian group on an *infinite* set X , we must again choose a total ordering of X , and then either talk about “formally infinite products with all but finitely many factors equal to e ”, or modify the normal form, say to “ e or $x^{i(x)} y^{i(y)} \dots z^{i(z)}$ where $x < y < \dots < z \in X$, and all exponents shown are nonzero” (the last two conditions to ensure uniqueness!).

We may be satisfied with one of these approaches, or we may prefer to go to a slightly different kind of representation for F , which we discover as follows: Note that if g is a member of the free abelian group F on X , then for each $x \in X$, the exponent $i(x)$ to which x appears in our normal forms for g is the same for these various forms; only the position in which $x^{i(x)}$ is written (and if $i(x) = 0$, whether it is written) changes from one normal form to another. Clearly, any of our normal forms for g , and hence the element g itself, is determined by the X -tuple of exponents $(i(x))_{x \in X}$. So let us “represent” g by this X -tuple; that is, identify F with a certain set of integer-valued functions on X . It is easy to see that the group operations of F correspond to componentwise addition of such X -tuples, componentwise additive inverse, and the constant X -tuple 0 ; and that the X -tuple corresponding to each generator $x \in X$ is the function δ_x having value 1 at x and 0 at all other elements $y \in X$. The X -tuples that correspond to members of F are those which are nonzero at only finitely many components. Thus we get the familiar description of the free abelian group on X as the subgroup of \mathbb{Z}^X consisting of all functions having finite support in X . (The *support* of a function f means $\{x \mid f(x) \neq 0\}$.)

Exercise 3.4:1. If X is infinite, it is clear that the whole group \mathbb{Z}^X is *not* a free abelian group on X under the map $x \mapsto \delta_x$, since it is not generated by the δ_x . Show that \mathbb{Z}^X is in fact not a free abelian group on *any* set of generators.

(For further results on \mathbb{Z}^X and its subgroups when X is countably infinite, see Specker [114]. Among other things, it is shown there that the uncountable group \mathbb{Z}^X has only countably many homomorphisms into \mathbb{Z} , though its countable subgroup F clearly has uncountably many! It is also shown that the subgroup of *bounded* functions on X is free abelian, on uncountably many generators. This fact was generalized to not necessarily countable X by Nöbeling [104]. For a simpler proof of this result, using ring theory, see [39, §1].)

The concept of the abelian group presented by a system of generators and relations may be formulated exactly like that of a group presented by generators and relations. It may also be constructed analogously: as the quotient of the free abelian group on the given generators by the subgroup generated by the relators st^{-1} (we don't have to say “normal subgroup” because normality is automatic for subgroups of abelian groups); or alternatively, as the *group* presented by the given generators and relations, together with the additional relations saying that all the generators commute with one another.

Suppose now that we start with an arbitrary group G , and *impose relations* saying that for all $x, y \in |G|$, x and y should commute: “ $xy = yx$ ”. That is, we form the quotient of G by the normal subgroup generated by the elements $(yx)^{-1}(xy) = x^{-1}y^{-1}xy$. As noted in the paragraph introducing Exercise 2.4:2, these elements are called *commutators*, and often written

$$x^{-1}y^{-1}xy = [x, y].$$

(A still more common notation is (x, y) , but we will not use this, to avoid confusion with ordered

pairs.) The normal subgroup that they generate is called the *commutator subgroup*, or *derived subgroup* of G , written $[G, G]$, and often abbreviated by group theorists to G' . The quotient group, $G^{\text{ab}} = G/[G, G]$, is an abelian group with a homomorphism q of the given group G into it which is *universal* among homomorphisms of G into abelian groups A , the diagram for the universal property being

$$\begin{array}{ccc}
 G & \xrightarrow{q} & G^{\text{ab}} \\
 & \searrow \forall v & \downarrow \exists! f \\
 & & A.
 \end{array}$$

This group G^{ab} (or more precisely, the pair (G^{ab}, q) , or any isomorphic pair) is called the *abelianization* or *commutator factor group* of G .

Suppose now that we write down any system of generators and relations for a group, and compare the *group* G and the *abelian group* H that these same generators and relations define. By the universal property of G , there will exist a unique homomorphism $r: G \rightarrow H$ taking the generators of G to the corresponding generators of H . It is easy to check that (H, r) has the universal property characterizing the abelianization of G . So this gives another way of describing abelianization. Note, as a consequence, that given an arbitrary system of generators and group-theoretic relations, the *group* these present will determine, up to natural isomorphism, the abelian group that they present (but not vice versa).

Exercise 3.4:2. Find the structures of the abelianizations of the groups presented in Exercises 3.3:1, 3.3:3, 3.3:4, 3.3:10 and 3.3:11(i). (This is easier than determining the structures of the groups themselves, hence the one exercise here corresponding to the many earlier exercises.)

Exercise 3.4:3. Show that any group homomorphism $f: G \rightarrow H$ induces a homomorphism of abelian groups $f^{\text{ab}}: G^{\text{ab}} \rightarrow H^{\text{ab}}$. State precisely the condition relating f and f^{ab} . Show that for a composite of group homomorphisms, one has $(fg)^{\text{ab}} = f^{\text{ab}}g^{\text{ab}}$. Conclude that for any group G , there is a natural homomorphism of monoids, $\text{End}(G) \rightarrow \text{End}(G^{\text{ab}})$, and a natural homomorphism of groups $\text{Aut}(G) \rightarrow \text{Aut}(G^{\text{ab}})$.

Exercise 3.4:4. For G as in Exercises 3.3:1 and 3.3:2, is the natural homomorphism $\text{Aut}(G) \rightarrow \text{Aut}(G^{\text{ab}})$ of the above exercise one-to-one?

Exercise 3.4:5. If H is a subgroup of G , what can be said about the relation between H^{ab} and G^{ab} ? Same question if H is a homomorphic image of G .

Exercise 3.4:6. Let K be a field, n a positive integer, and $\text{GL}(n, K)$ the group of invertible $n \times n$ matrices over K . Determine as much as you can about the structure of $\text{GL}(n, K)^{\text{ab}}$.

Exercise 3.4:7. If G is a group, will there exist a universal homomorphism of G into a *solvable* group, $G \rightarrow G^{\text{solv}}$? What if G is assumed finite?

Does there exist a “free solvable group” on a set X , or some similar construction?

Exercise 3.4:8. Show that the free abelian group on n generators cannot be presented as a group by fewer than n generators and $n(n-1)/2$ relations.

3.5. The Burnside problem. In 1902, W. Burnside [55] asked whether a finitely generated group, all of whose elements have finite order, must be finite. This problem was hard to approach because, with nothing assumed about the *values* of the finite orders of the elements, one had no

place to begin a calculation. So Burnside also posed this question under the stronger hypothesis that there be a *common* finite bound on the orders of all elements of G .

The original question with no bound on the orders was suddenly answered negatively in 1964, with a counterexample arising from the Golod-Shafarevich construction [71]; there is a short and fairly self-contained presentation of this material in the last chapter of [28]. In the opposite direction, Burnside himself proved that if G is a finitely generated group of *matrices* over a field, and all elements of G have finite order, then G is finite [56].

Turning to the question of a general group G with a common bound on the orders of its elements, note that if m is such a bound, then $m!$ is a common *exponent* for these elements; while if n is a common exponent, it is also a bound on their orders. So “there is a common bound on the orders of all elements” is equivalent to “all elements have a common exponent”. The latter condition is more convenient to study, since the statement that x has exponent n has the form of an equation. So for any positive integer n , one defines the *Burnside problem for exponent n* to be the question of whether every finitely generated group satisfying the identity

$$(3.5.1) \quad (\forall x) x^n = e$$

is finite.

For $n = 1$, the answer is trivially yes, for $n = 2$ the same result is an easy exercise, for $n = 3$ it is not very hard to show, and it has also been proved for $n = 4$ and $n = 6$. On the other hand, it has been shown in recent years that the answer is negative for all odd $n \geq 665$ [34], and for all $n > 8000$ [94]. This leaves a large but finite set of cases still open: all odd values from 5 to 663, and all even values from 8 to 8000. We won't get involved in these hard group-theoretic problems here. But the concept of universal constructions does allow us to understand the nature of the question better. Call a group G an *n -Burnside group* if it satisfies (3.5.1). One may define the *free n -Burnside group* on any set X by the obvious universal property, and it will exist for the usual reasons. In particular, it can be presented, as a group, by the generating set X , and the infinite family of relations equating the n th powers of *all* terms in the elements of X to e . I leave it to you to think through the following relationships:

Exercise 3.5:1. Let n and r be positive integers.

- (i) What implications can you prove among the following statements?
 - (a) Every n -Burnside group which can be generated by r elements is finite.
 - (b) The free n -Burnside group on r generators is finite.
 - (c) The group $\langle x_1, \dots, x_r \mid x_1^n = \dots = x_r^n = e \rangle$ is finite.
 - (d) There exists a finite r -generator group having a finite presentation in which all relators are n th powers, $\langle x_1, \dots, x_r \mid w_1^n = \dots = w_s^n = e \rangle$ (where each w_i is a term in x_1, \dots, x_r . Cf. Exercises 3.3:7 and 3.3:8.)
 - (e) There exists an integer N such that all n -Burnside groups generated by r elements have order $\leq N$.
 - (f) There exists an integer N such that all *finite* n -Burnside groups generated by r elements have order $\leq N$ (“the restricted Burnside problem”).
- (ii) What implications can you prove among cases of statement (a) involving the same value of n but different values of r ? involving the same value of r but different values of n ?

Note that if for a given n and r we could find a *normal form* for the free n -Burnside group on r generators, we would know whether (b) was true! But except when n or r is very small, such normal forms are not known. For further discussion of these questions, see [27, Chapter 18]. Recent results, including a solution to the restricted Burnside problem ((f) above), and some

negative results on the *word problem* for free Burnside groups can be found in [86], [100], [121], [122], and references given in those works.

A group G is called *residually finite* if for any two elements $x \neq y \in |G|$, there exists a homomorphism f of G into a *finite* group such that $f(x) \neq f(y)$.

Exercise 3.5:2. Investigate implications involving conditions (a)-(f) of the preceding exercise, together with

(g) The free n -Burnside group on r generators is residually finite.

Exercise 3.5:3. (i) Restate Exercise 2.4:3 as a question about residual finiteness (showing, of course, that your restatement is equivalent to the original question).

(ii) If G is a group, does there exist a universal homomorphism $G \rightarrow G^{\text{rf}}$, of G into a residually finite group?

We remark that the original Burnside problem, with no restriction on orders, is still open for *finitely presented* groups [130].

3.6. Products and coproducts. Let G and H be groups. Consider the following two situations:

(a) a group P given with a homomorphism $p_G: P \rightarrow G$ and a homomorphism $p_H: P \rightarrow H$, and

(b) a group Q given with a homomorphism $q_G: G \rightarrow Q$, and a homomorphism $q_H: H \rightarrow Q$ (cf. diagrams below).

Note that if in situation (a) we choose a homomorphism a of any other group P' into P , then P' also acquires homomorphisms into G and H , namely $p_G a$ and $p_H a$; similarly, if in situation (b) we choose any homomorphism b of Q into a group Q' , then Q' acquires homomorphisms $b q_G$ and $b q_H$ of G and H into it:



So we may ask whether there exists a *universal* example of a P with maps into G and H , that is, a 3-tuple (P, p_G, p_H) such that for any group P' , every pair of maps $p'_G: P' \rightarrow G$ and $p'_H: P' \rightarrow H$ arises by composition of p_G and p_H with a unique homomorphism $a: P' \rightarrow P$; and, dually, whether there exists a universal example of a group Q with maps of G and H into it.

In both cases, the answer is yes. The universal P is simply the direct product group $G \times H$, with its projection maps p_G and p_H onto the two factors; the universal property is easy to verify. The universal Q , on the other hand, can be constructed by generators and relations. It has to have for each $g \in |G|$ an element $q_G(g)$ – let us abbreviate this to \bar{g} – and for each $h \in |H|$ an element $q_H(h)$ – call this \tilde{h} . So let us take for generators a set of symbols

$$(3.6.1) \quad \{\bar{g}, \tilde{h} \mid g \in |G|, h \in |H|\}.$$

The relations these must satisfy are those saying that q_G and q_H are homomorphisms:

$$(3.6.2) \quad \bar{g}\bar{g}' = \overline{gg'} \quad (g, g' \in |G|), \quad \tilde{h}\tilde{h}' = \widetilde{hh'} \quad (h, h' \in |H|).$$

It is immediate that the group presented by generators (3.6.1) and relations (3.6.2) has the desired universal mapping property. (We might have supplemented (3.6.2) with the further

relations $\overline{e_G} = e$, $\widetilde{e_H} = e$, $\overline{g^{-1}} = \overline{g}^{-1}$, $\widetilde{h^{-1}} = \widetilde{h}^{-1}$. But these are implied by the relations listed, since, as is well known, any set map between groups which preserves products also preserves neutral elements and inverses.) More generally, if G is a group which can be presented as $\langle X \mid R \rangle$, and if similarly $H = \langle Y \mid S \rangle$, then we may take for generators of Q a disjoint union $X \sqcup Y$, and for relations the union of R and S . For instance, if

$$G = \mathbb{Z}_3 = \langle x \mid x^3 = e \rangle \quad \text{and} \quad H = \mathbb{Z}_2 = \langle x \mid x^2 = e \rangle,$$

then Q may be presented as

$$\langle x, x' \mid x^3 = e, x'^2 = e \rangle,$$

with q_G and q_H given by $x \mapsto x$ and $x \mapsto x'$, respectively. You should be able to verify the universal property of Q from this presentation.

(If you are not familiar with the concept of a ‘‘disjoint union’’ $X \sqcup Y$ of two sets X and Y , I hope that the above context suggests the meaning. Explicitly, it means the union of a bijective copy of X and a bijective copy of Y , chosen to be disjoint. So, if $X = \{a, b, c\}$, $Y = \{b, c, d, e\}$ where a, b, c, d, e are all distinct, then their ordinary set-theoretic union is the 5-element set $X \cup Y = \{a, b, c, d, e\}$, but an example of a ‘‘disjoint union’’ would be any set of the form $X \sqcup Y = \{a, b, c, b', c', d', e'\}$ where a, b, c, b', c', d', e' are all distinct, given with the obvious maps taking X to the 3-element subset $\{a, b, c\}$ of this set and Y to the disjoint 4-element subset $\{b', c', d', e'\}$. Though there is not a unique way of choosing a disjoint union of two sets, the construction is unique in the ways we care about; e.g., note that in the above example, any disjoint union of X and Y will have $|X| + |Y| = 7$ elements. Hence one often speaks of ‘‘the’’ disjoint union. We will see, a few sections from now, that disjoint union of sets is itself a universal construction – of set theory.)

To see for general G and H what the group determined by the above universal property ‘‘looks like’’, let us again think about an arbitrary group Q with homomorphisms of G and H into it, abbreviated $g \mapsto \overline{g}$ and $h \mapsto \widetilde{h}$. The elements of Q which we can name in this situation are, of course, the products

$$(3.6.3) \quad x_n^{\pm 1} x_{n-1}^{\pm 1} \dots x_1^{\pm 1} \quad \text{with} \quad x_i \in \{\overline{g}, \widetilde{h} \mid g \in |G|, h \in |H|\} \quad \text{and} \quad n \geq 0.$$

(Notational remark: In §2.4, I generally kept $n \geq 1$, and introduced ‘‘ e ’’ as a separate kind of expression. Here I shall adopt the convenient convention that the product of the empty (length 0) sequence of factors is e , so that the case ‘‘ e ’’ may be absorbed in the general case.)

Now for any $g \in |G|$ or $h \in |H|$ we have noted that $\overline{g^{-1}} = \overline{g}^{-1}$ and $\widetilde{h^{-1}} = \widetilde{h}^{-1}$ in Q ; hence the inverse of any member of the generating set $\{\overline{g}, \widetilde{h} \mid g \in |G|, h \in |H|\}$ is another member of that set. So we may simplify any product (3.6.3) to one in which all exponents are $+1$, and so write it without showing these exponents. We also know that $\overline{e} = \widetilde{e} = e$, so wherever instances of \overline{e} or \widetilde{e} occur in such a product, we may drop them. Finally, the relations (3.6.2) allow us to replace any occurrence of two successive factors coming from $\{\overline{g} \mid g \in |G|\}$ by a single such factor, and to do the same if two factors from $\{\widetilde{h} \mid h \in |H|\}$ occur together. So the elements of Q that we can construct can all be reduced to the form

$$(3.6.4) \quad x_1 \dots x_n$$

where $n \geq 0$, $x_i \in \{\bar{g} \mid g \in |G| - \{e\}\} \cup \{\tilde{h} \mid h \in |H| - \{e\}\}$, and no two successive x 's come from the same set, $\{\bar{g} \mid g \in |G| - \{e\}\}$ or $\{\tilde{h} \mid h \in |H| - \{e\}\}$.

We can express the *product* of two elements (3.6.4) as another such element, by putting the sequences of factors together, and reducing the resulting expression to the above form as described above; likewise it is clear how to find expressions of that form for inverses of elements (3.6.4), and for the element e . In any particular group Q with homomorphisms of G and H into it, there may be other elements than those expressed by (3.6.4), and there may be some equalities among such products. But as far as we can see, there don't seem to be any cases left of two expressions (3.6.4) that must represent the same element in *every* such group Q . If in fact there are none, then, as in §2.4, the expressions (3.6.4) will correspond to the distinct elements of the *universal* Q we are trying to describe, and thus will give a normal form for the elements of this group.

We can use the same stratagem as in §2.4 to show that there are no undiscovered necessary equalities – it was for this situation that van der Waerden devised it!

Proposition 3.6.5 (van der Waerden [123]). *Let G, H be groups, and Q the group with a universal pair of homomorphisms $G \rightarrow Q, H \rightarrow Q$, written $g \mapsto \bar{g}, h \mapsto \tilde{h}$. Then every element of Q can be written uniquely in the form (3.6.4).*

Proof. Let us, as before, introduce an additional symbol a , and now denote by A the set of all symbols

$$(3.6.6) \quad x_n \dots x_1 a, \quad \text{where } x_1, \dots, x_n \text{ are as in (3.6.4) (the } n = 0 \text{ case now being interpreted as the bare symbol } a).$$

We would like to describe actions of G and H on this set. It is clear what these actions *should* be, but an explicit description is a bit messy, because of the need to state separately the cases where the element of A on which we are acting does or does not begin with an element of the group we are acting by, and if it does, the cases where this beginning element is or is not the inverse of the element by which we are acting. This messiness in the definition makes still more messy the verification that the ‘‘actions’’ give homomorphisms of G and of H into the permutation group of A .

We shall get around these annoyances (which are in any case minor compared with the difficulties of doing things *without* van der Waerden's method) by another trick. Let us describe a set A_G which is in bijective correspondence with A : For those elements $b \in A$ which already begin with a symbol \bar{g} ($g \in |G| - \{e\}$), we let A_G contain the same element b . For elements b which do not, let the corresponding element of A_G be the expression $\bar{e}b$. Thus *every* element of A_G begins with a symbol \bar{g} ($g \in |G|$), and we can now describe the action of $g' \in |G|$ on A_G as simply taking an element $\bar{g}c$ to $\overline{g'g}c$. It is trivial to verify that *this* is a homomorphism of G into the permutation group of A_G . This action on A_G now induces, in an obvious way, an action on the bijectively related set A .

Likewise, an action of H on A can be defined, via an action on the analogously constructed set A_H .

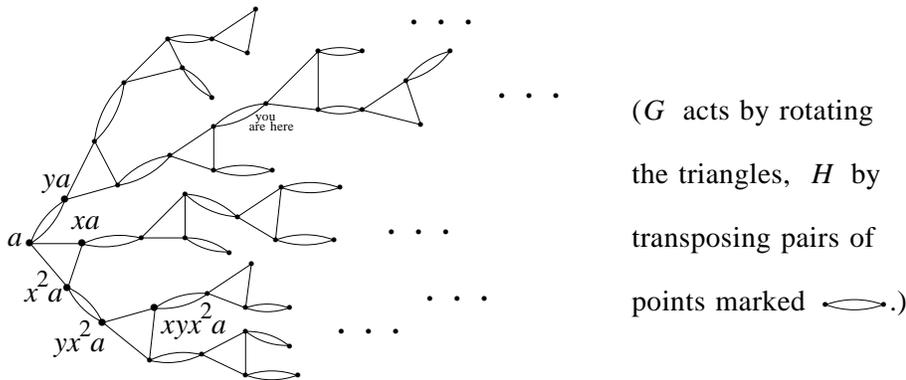
Thus we have homomorphisms of both G and H into the permutation group of A ; this is equivalent to giving a homomorphism of the group Q we are interested in into this permutation group. Further, given any element (3.6.4) of Q , it is easy to see by induction on n that its

image in the permutation group of A sends the “starting point” element a to precisely $x_n \dots x_1 a$. Hence two distinct expressions (3.6.4) correspond to elements of Q having distinct actions on a , hence these elements of Q are themselves distinct. So not only can every element of Q be written in the form (3.6.4), but distinct expressions (3.6.4) correspond to distinct elements of Q , proving the proposition. \square

For a concrete example, again let $G = \mathbb{Z}_3 = \langle x \mid x^3 = e \rangle$ and let $H = \mathbb{Z}_2 = \langle y \mid y^2 = e \rangle$. Then A will consist of strings such as a, ya, xyx^2a , etc.. (We can drop “-” and “~” here because $|G| - \{e\}$ and $|H| - \{e\}$ use no symbols in common.) The element x of $G = \mathbb{Z}_3$ will

act on this set by 3-cycles $b \begin{matrix} \nearrow xb \\ \downarrow \\ \searrow x^2b \end{matrix}$, one for each string b not beginning with x , while the

element y of $H = \mathbb{Z}_2$ acts by transposing pairs of symbols $b \rightleftarrows yb$ where b does not begin with y . If we want to see that say, $yxyx^2$ and x^2yxy have distinct actions on A , we simply note that the first sends the symbol a to the symbol $yxyx^2a$, while the second takes it to x^2yxya . A picture of the Q -set A , for this G and H , looks like some kind of seaweed:



We recall that the universal group “ P ” considered at the beginning of this section turned out to be the direct product of G and H . Since Q is characterized by a dual universal property, we shall call it the *coproduct* of G and H .

Because of the similarity of the normal form of this construction to that of *free groups*, group-theorists have long called it the *free product* of the given groups. However, the constructions for *sets, commutative rings, abelian groups, topological spaces*, etc. characterized by this same universal property show a great diversity of forms, and have been known under different names in the respective disciplines. The general name “coproduct” introduced by category theory (Chapter 6 below) unifies the terminology, and we shall follow it. On the other hand, the “ P ” constructions look very similar in all these cases, and have generally all had the name “direct product”, which is retained (shortened to “product”) by category theory.

In both our product and coproduct constructions, the pair of groups G and H may be replaced by an arbitrary family $(G_i)_{i \in I}$. The universal example of a group P given with an I -tuple of maps $p_i: P \rightarrow G_i$ is again the classical direct product $\prod_I G_i$ with its I -tuple of projection maps. The coproduct $Q = \coprod_I G_i$, generated by the images of a universal family of maps of the groups G_i 's into Q , can be constructed, just as above, using strings of nonidentity elements from a disjoint union of the underlying sets of these groups, such that two factors from the same group G_i never occur consecutively. The coproduct symbol \coprod is, of course, the direct product symbol \prod turned upside-down.

Exercise 3.6:1. If X is a set, then a coproduct of copies of the infinite cyclic group \mathbb{Z} , indexed by X , $\coprod_X \mathbb{Z}$, will be a free group on X . Show this by universal properties, and describe the correspondence of normal forms. Can you find any other families of groups whose coproduct is a free group?

Exercise 3.6:2. Let us (following group-theorists' notation) write coproducts of finite families of groups as $Q = G * H$, $Q = F * G * H$, etc.. Prove that for any three groups F , G and H , one has $(F * G) * H \cong F * G * H \cong F * (G * H)$, using (a) universal properties, and (b) normal forms.

Exercise 3.6:3. For any two groups G and H , show how to define natural isomorphisms $i_{G,H}: G \times H \cong H \times G$, and $j_{G,H}: G * H \cong H * G$. What form do these isomorphisms take when $G = H$? (Describe them on elements.)

It is sometimes said that "We may identify $G \times H$ with $H \times G$, and $G * H$ with $H * G$, by treating the isomorphisms $i_{G,H}$ and $j_{G,H}$ as the identity, and identifying the corresponding group elements." Is this reasonable when $G = H$?

Exercise 3.6:4. Show that in a coproduct group $G * H$, the only elements of finite order are the conjugates of the images of elements of finite order of G and H . (First step: Find how to determine, from the normal form of an element of $G * H$, whether it is a conjugate of an element of G or H .)

Can you similarly describe all finite *subgroups* of $G * H$?

There is a fact about the direct product group which one would not at first expect from its universal property: It also has two natural maps *into* it: $f_G: G \rightarrow G \times H$ and $f_H: H \rightarrow G \times H$, given by $g \mapsto (g, e)$ and $h \mapsto (e, h)$. (Note that there are no analogous maps into a direct product of *sets*.) To examine this phenomenon, we recall that the universal property of $G \times H$ says that to map a group A into $G \times H$ is equivalent to giving a map $A \rightarrow G$ and a map $A \rightarrow H$. Looking at f_G , we see that the two maps it corresponds to are the identity map $\text{id}_G: G \rightarrow G$, defined by $\text{id}_G(g) = g$, and the trivial map $e: G \rightarrow H$, defined by $e(g) = e$. The map f_H is characterized similarly, with the roles of G and H reversed.

The group $G \times H$ has, in fact, a second universal property, in terms of this pair of maps. The 3-tuple $(G \times H, f_G, f_H)$ is universal among 3-tuples (K, a, b) such that K is a group, $a: G \rightarrow K$ and $b: H \rightarrow K$ are homomorphisms, and the images in K of these homomorphisms *centralize* one another:

$$(\forall g \in |G|, h \in |H|) \quad a(g)b(h) = b(h)a(g),$$

equivalently:

$$[a(G), b(H)] = \{e\}.$$

If $P = \prod_I G_i$ is a direct product of arbitrarily many groups, one similarly has maps $f_i: G_i \rightarrow P$, but if the index set I is infinite, the images of the f_i will not in general generate P , and it follows from this that P cannot have the same universal property. But one finds that the subgroup P_0 of P generated by the images $f_i(G_i)$ (which consists of those elements of P having only finitely many coordinates $\neq e$) is again a universal group with maps of the G_i into it having images that centralize one another.

Exercise 3.6:5. (i) Prove the above new universal property of $G \times H$.

(ii) Describe the map

$$m: G * H \rightarrow G \times H$$

which the universal property of $G * H$ associates to the above pair of maps f_G, f_H and deduce that this map is surjective, and that its kernel is the normal subgroup of $G * H$ generated by the

commutators $[\bar{g}, \tilde{h}]$ ($g \in |G|, h \in |H|$).

(iii) Give versions of the above results for products and coproducts of possibly infinite families $(G_i)_{i \in I}$.

One may wonder why commutativity suddenly came up like this, since the original universal property by which we characterized $G \times H$ had nothing to do with it. The following observation throws a little light on this. The set of relations that will be satisfied in $G \times H$ by the images of elements of G and H under the two maps f_G and f_H defined above will be the *intersection* of the sets of relations satisfied by their images in K under $a: G \rightarrow K, b: H \rightarrow K$, in the two cases

- (iv) $K = G; a = \text{id}_G, b = e,$
- (v) $K = H; a = e, b = \text{id}_H.$ (Why?)

And what are such relations? Clearly $a(g)b(h) = b(h)a(g)$ holds in each case. The above second universal property of $G \times H$ is equivalent to saying that no relations hold in both cases *except* these relations and their consequences.

A coproduct group $G * H$ similarly has natural maps $u_G: G * H \rightarrow G$ and $u_H: G * H \rightarrow H$, constructed from the identity maps of G and H and the trivial maps between them; but u_G and u_H have no unexpected properties that I know of.

Exercise 3.6:6. If G is a group, construct maps $G \rightarrow G \times G$ and $G * G \rightarrow G$ using universal properties, and the identity map of G , but *not* using the trivial map of G . Describe how these maps behave on elements.

Exercise 3.6:7. Suppose $(G_i)_{i \in I}$ is a family of groups, and we wish to consider groups G given with homomorphisms $G_i \rightarrow G$ such that the images of *certain* pairs $G_i, G_{i'}$ commute, while no condition is imposed on the remaining pairs. To formalize this, let $J \subseteq I \times I$ be a symmetric antireflexive relation on our index set I (antireflexive means $(\forall i \in I) (i, i) \notin J$); and let H be the universal group with homomorphisms $r_i: G_i \rightarrow H (i \in I)$ such that for $(i, i') \in J, [r_i(G_i), r_{i'}(G_{i'})] = \{e\}$.

Study the structure of this H , and obtain a normal form if possible. You may assume the index set I finite if this helps.

3.7. Products and coproducts of abelian groups. Let A and B be abelian groups. Following the model of the preceding section, we may look for abelian groups P and Q having universal pairs of maps:



Again abelian groups with both these properties exist – but this time, they turn out to be the same group, namely $A \times B$! (The reader should verify both universal properties.) To look at this another way, if we construct abelian groups P and Q with the universal properties of the direct product and coproduct of A and B respectively, and then form the homomorphism $m: P \rightarrow Q$ analogous to that of Exercise 3.6:5, this turns out to be an isomorphism.

Note that though $A \times B$ is the universal *abelian* group with homomorphisms of A and B into it, this is not the same as the universal *group* with homomorphisms of A and B into it – that group, $A * B$, constructed in the preceding section, will generally not be abelian when A and B are. Thus, the coproduct of two abelian groups A and B as *abelian groups* is generally not the

same as their coproduct as *groups*. Rather, we can see by comparing universal properties that the coproduct as abelian groups is the abelianization of the coproduct as groups: $A \times B = (A * B)^{\text{ab}}$.

Hence, in using the coproduct symbol “ \amalg ”, we have to specify what kind of coproduct we are talking about, $\amalg_{\text{gp}} A_i$ or $\amalg_{\text{ab gp}} A_i$, unless this is clear from context. On the other hand, a *direct product* of abelian groups as abelian groups is the same as their direct product as groups.

For a not necessarily finite family $(A_i)_{i \in I}$ of abelian groups, the coproduct still *embeds in* the direct product under the map “ m ”. It can in fact be described as the subgroup of that direct product group consisting of those elements almost all of whose coordinates are e . When abelian groups are written additively, this coproduct is generally called the “direct sum” of the groups, and denoted $\bigoplus_I A_i$; in the case of two groups we write $A \oplus B = A \times B$.

Notes on confused terminology: Some people extend the term “direct sum” to mean “coproduct” in all contexts – groups, rings, etc.. Other writers, because of the form that “direct sum” has for finite families of abelian groups, use the phrase “direct sum” as a synonym of “direct product”, even in the case of infinite families of groups! The coproduct of an infinite family of abelian groups is sometimes called their “restricted direct product” or “restricted direct sum”, the direct product then being called the “complete direct product” or “complete direct sum”. In these notes, we shall stick with the terms “product” and “coproduct”, as defined above (except that we shall often expand “product” to “direct product”, to avoid possible confusion with meanings such as a product of elements under a multiplication).

What is special about abelian groups, that makes finite products and coproducts come out the same; and why only *finite* products and coproducts? One may consider the key property to be the fact that homomorphisms of abelian groups can be “added”; i.e., that given two homomorphisms $f, g: A \rightarrow B$, the map $f+g: A \rightarrow B$ defined by $(f+g)(a) = f(a) + g(a)$ is again a homomorphism. (The corresponding statement is not true for nonabelian groups.) Temporarily writing $*_{\text{ab gp}}$ for the coproduct of two abelian groups, one finds, in fact, that the map $m: G *_{\text{ab gp}} H \rightarrow G \times H$ referred to earlier has an inverse, given by the sum

$$q_G p_G + q_H p_H: G \times H \rightarrow G *_{\text{ab gp}} H.$$

For coproducts of noncommutative groups, the corresponding map is not a group homomorphism, while for coproducts of infinite families of abelian groups, no analog of the above map can be constructed because one cannot in general make sense of an infinite sum of homomorphisms. So it is only when the coproduct is taken in the class of abelian groups, and the given family is finite, that we get in this way an inverse to m .

Part (ii) of the next exercise concerns a subtle but interesting distinction.

Exercise 3.7:1. (i) Show that for any groups G and H one has $(G * H)^{\text{ab}} \cong (G \times H)^{\text{ab}} \cong G^{\text{ab}} \times H^{\text{ab}}$.

(ii) Given an infinite family of groups (G_i) , is it similarly true that $(\amalg_{\text{gp}} G_i)^{\text{ab}} \cong \amalg_{\text{ab gp}} G_i^{\text{ab}}$ (i.e., $\bigoplus G_i^{\text{ab}}$), and that $(\prod G_i)^{\text{ab}} \cong \prod G_i^{\text{ab}}$? If one of these isomorphisms is not always true, can you establish any general results on when it holds and when it fails?

3.8. Right and left universal properties. The universal property of direct products differs in a basic way from the other universal properties we have looked at so far. In all other cases, we constructed an object (e.g., a group) F with specified “additional structure” or conditions (e.g., a map of a given set X into $|F|$), such that any instance of a structure of that sort on any object U could be obtained by a unique homomorphism *from* the universal object F *to* the object U . A

direct product $P = G \times H$ is an object with the opposite type of universal property: all groups with the specified additional structure (a map into G , and a map into H) are obtained by mapping arbitrary groups U into the universal example P . Thus, while the free group on a set X , the abelianization of a group G , the coproduct of two groups G and H , etc., can be thought of as “first” or diagrammatically “leftmost” groups with given kinds of structure, the direct product $G \times H$ is the “last” or “rightmost” group with maps into G and H . We shall refer to these two types of conditions as “left” and “right” universal properties respectively. (This terminology is based on thinking of arrows as going from left to right, though it happens that in most of the diagrams in preceding sections, the arrow from the left universal object to the general object was drawn downward.)

The philosophy of how to construct objects with properties of either kind is in broad outline the same: Figure out as much information as possible about an arbitrary object (*not* assumed universal) with the given “additional structure”, and see whether that information can itself be considered as a description of an object. If it can, this object will in general turn out to be universal for the given structure! In the case of “left universal” constructions (free groups, coproducts, etc.), this “information” means answers to the question, “What elements do we know exist, and what equalities must hold among them?” (cf. remark 2.2.10). In the right universal case, on the other hand, the corresponding question is, “Given an element of our object, what data can one describe about it in terms of the additional structure?”

Let us illustrate this with the case of the direct product of groups. Given groups G and H , consider any group P with specified homomorphisms p_G, p_H into G and H respectively. What data can we find about an element x of P using these maps? Obviously, we can get from x a pair of elements $(g, h) \in |G| \times |H|$, namely

$$g = p_G(x) \in |G|, \quad h = p_H(x) \in |H|.$$

Can we get any more data? We can also obtain elements $p_G(x^2), p_H(x^{-1})$, etc.; but these can be found by group operations from the elements $g = p_G(x)$ and $h = p_H(x)$, so they give no new information about x . All right then, let us agree to classify elements of P according to the pairs $(g, h) \in |G| \times |H|$ which they determine.

Now suppose $x \in |P|$ gives the pair (g, h) , and y gives the pair (g', h') . Can we find from these the pair given by $xy \in |P|$? the pair given by x^{-1} ? Clearly so: these will be (gg', hh') , and (g^{-1}, h^{-1}) respectively. And we can likewise write down the pair that $e \in |P|$ yields: (e_G, e_H) .

Very well, let us take the “data” by which we have classified elements of our arbitrary P , namely the set of pairs (g, h) ($g \in |G|, h \in |H|$) – together with the law of composition we have found for these pairs, namely $(g, h) \cdot (g', h') = (gg', hh')$, the inverse operation $(g, h) \mapsto (g^{-1}, h^{-1})$, and the neutral element pair (e_G, e_H) – and ask whether this data forms a group. It does! And, because of the way this group was constructed, it will have homomorphisms into G and H , and we find it is universal for this property. It is, of course, the product group $G \times H$.

Here is a pair of examples we have not yet discussed. Suppose we are given a homomorphism of groups

$$f: G \rightarrow H.$$

Now consider

(a) homomorphisms $a: A \rightarrow G$, from arbitrary groups A into G , whose composites with f are

the trivial homomorphism, i.e., which satisfy $fa = e$; and

(b) homomorphisms $b: H \rightarrow B$, from H into arbitrary groups B , whose composites with f are the trivial homomorphism, i.e., which satisfy $bf = e$.

Given a homomorphism of the first sort, one can get further homomorphisms with the same property by composing with homomorphisms $A' \rightarrow A$, for arbitrary groups A' ; so one may look for a pair (A, a) with the *right* universal property that every such pair (A', a') arises from (A, a) via a unique homomorphism $A' \rightarrow A$. For (b), one would want a corresponding *left* universal B .

To try to find the right-universal A , we ask: Given an arbitrary homomorphism $A \rightarrow G$ with $fa = e$ as in (a), what data can we attach to any element $x \in |A|$? Its image $g = a(x)$, certainly. This must be an element which f carries to the neutral element, since $fa = e$; thus the set of possibilities is $\{g \in |G| \mid f(g) = e\}$. We find that this set forms a group (with a map into G , namely the inclusion) having the desired universal property. This is the *kernel* of f .

We get the left universal example of (b) by familiar methods: Given arbitrary $b: H \rightarrow B$ with $bf = e$ as in (b), B must contain an image $\bar{h} = b(h)$ of each element $h \in |H|$. The fact that $bf = e$ tells us that the images in B of all elements of $f(G)$ must be the neutral element, and we quickly discover that the universal example is the quotient group $B = H/N$, where N is the normal subgroup of H generated by $f(G)$. This group H/N is called the *cokernel* of the map f .

Right universal constructions are not as conspicuous in algebra as left universal constructions. When they occur, they are often fairly elementary and familiar constructions (e.g., the direct product of two groups; the kernel of a homomorphism). However, we shall see less trivial cases in later chapters; some of the exercises below also give interesting examples.

Exercise 3.8:1. Let G be a group, and X a set. Show that there exist

- (i) a G -set S with a universal map $f: |S| \rightarrow X$, and
- (ii) a G -set T with a universal map $g: X \rightarrow |T|$,

and describe these G -sets. Begin by stating the universal properties explicitly.

(Hint to (i): Given any G -set S with a map $f: |S| \rightarrow X$, an element $s \in |S|$ will determine not only an element $x = f(s) \in X$, but for every $g \in |G|$ an element $x_g = f(gs) \in X$. From the *family* of elements, $(x_g)_{g \in |G|}$ determined by an $s \in S$, can one describe the family associated with hs for any $h \in |G|$?)

One can carry the idea of the above exercise further in several directions:

(a) Given a group homomorphism $\varphi: G_1 \rightarrow G_2$, note that from any G_2 -set S one can get a G_1 -set S_φ , by taking the same underlying set, and defining for $g \in |G_1|$, $s \in |S|$

$$gs = \varphi(g)s.$$

Now given a G_1 -set X , one can look for a G_2 -set S with a universal homomorphism of G_1 -sets $S_\varphi \rightarrow X$, or for a G_2 -set T with a universal homomorphism of G_1 -sets $X \rightarrow T_\varphi$. The above exercise corresponds to the cases where $G_1 = \{e\}$, since an $\{e\}$ -set is essentially a set with no additional structure. You should verify that for $G_1 = \{e\}$, the universal questions just mentioned reduce to those of that exercise.

(b) Instead of looking at *sets* S on which a group G acts by *permutations*, one can consider abelian groups or vector spaces on which G acts by *automorphisms*. Such structures are called *linear representations* of G . In this case, the universal constructions analogous to those of (a) above are still possible, and they give two concepts of “induced representations” of a group,

important in modern group theory.

(c) The preceding point introduced extra structure on the *sets* on which our groups act. One can also consider the situation where one's groups G have additional structure, say topological or measure-theoretic, and restrict attention to continuous, measurable, etc., G -actions on appropriately structured spaces S . The versions of "induced representation" that one then obtains are at the heart of the modern representation theory of topological groups.

Exercise 3.8:2. Let G be a group. As discussed in the last two sentences of point (a) above, the ideas described there, applied to the unique homomorphism $\{e\} \rightarrow G$, lead to the two universal constructions of Exercise 3.8:1. Apply the same ideas to the unique homomorphism $G \rightarrow \{e\}$ (again combining them with the observation that an $\{e\}$ -set is essentially the same as a set) and describe the resulting constructions explicitly.

Exercise 3.8:3. Formulate right universal properties analogous to the left universal property defining free groups and the abelianization of a group, and show that no constructions exist having these properties. What goes wrong when we attempt to apply the general approach of this section?

Exercise 3.8:4. If X is a set and S a subset of X , then given any set map $f: Y \rightarrow X$, one gets a subset of Y , $T = f^{-1}(S)$. Does there exist a *universal* pair (X, S) , such that for any set Y , every subset $T \subseteq Y$ is induced in this way via a unique set map $f: Y \rightarrow X$?

Exercise 3.8:5. Let A, B be fixed sets. Suppose X is another set, and $f: A \times X \rightarrow B$ is a set map. Then for any set Y , and map $m: Y \rightarrow X$, a set map $A \times Y \rightarrow B$ is induced. (How?) Does there exist, for each A and B , a universal set X and map f as above, i.e., an X and an f such that for any Y , all maps $A \times Y \rightarrow B$ are induced by unique maps $Y \rightarrow X$?

Exercise 3.8:6. Let R be a ring with 1. (Commutative if you like. If you consider general R , then for "module" understand "left module" below.) Before attempting each of the following questions, formulate precisely the universal property desired.

- (i) Given a set X , does there exist an R -module M with a universal set map $|M| \rightarrow X$?
- (ii) If M is an R -module, let M_{add} denote the underlying additive group of M . Given an abelian group A , does there exist an R -module M with a universal homomorphism of abelian groups $M_{\text{add}} \rightarrow A$?
- (iii) and (iv): What about the left universal analogs of the above right universal questions?

3.9. Tensor products. Let A, B and C be abelian groups, which we shall write additively. Then by a *bilinear map* $\beta: (A, B) \rightarrow C$ we shall mean a set map $\beta: |A| \times |B| \rightarrow |C|$ such that

- (i) for each $a \in |A|$, the map $\beta(a, -): |B| \rightarrow |C|$ (that is, the map taking each element $b \in |B|$ to $\beta(a, b) \in |C|$) is a *linear* map (homomorphism of abelian groups) from B to C , and
- (ii) for each $b \in |B|$, the map $\beta(-, b): |A| \rightarrow |C|$ is a linear map from A to C .

This is usually called a bilinear map "from $A \times B$ to C ". (I usually call it that myself.) However, that terminology misleads many students into thinking that it has something to do with the *group* $A \times B$. In fact, although the definition of bilinear map involves the group structures of A and B , and involves the set $|A| \times |B|$, it has nothing to do with the structure of direct product group that one can put on this set. This is illustrated by:

Exercise 3.9:1. Show that for any abelian groups A, B, C , the only map $|A| \times |B| \rightarrow |C|$ which is both a linear map $A \times B \rightarrow C$, and a bilinear map $(A, B) \rightarrow C$ is the zero map.

As an example to keep in mind, take any ring $R = (|R|, +, \cdot, -, 0, 1)$, and let R_{add} denote the additive group $(|R|, +, -, 0)$. Then the maps $(x, y) \mapsto x + y$ and $(x, y) \mapsto x - y$ are *group*

homomorphisms $R_{\text{add}} \times R_{\text{add}} \rightarrow R_{\text{add}}$, but not bilinear maps; while the multiplication map $(x, y) \mapsto x \cdot y$ is a bilinear map $(R_{\text{add}}, R_{\text{add}}) \rightarrow R_{\text{add}}$, but not a group homomorphism $R_{\text{add}} \times R_{\text{add}} \rightarrow R_{\text{add}}$.

I am speaking about abelian groups to keep the widest possible audience. However, abelian groups can be regarded as \mathbb{Z} -modules, and everything I have said and will say about bilinear maps of abelian groups applies, more generally, to bilinear maps of modules over an arbitrary commutative ring, and in particular, to bilinear maps of vector spaces over a field, with the adjustment that “linear map” in (i) and (ii) above should be understood to mean module homomorphism. (There are also extensions of all these concepts to left modules, right modules, and bimodules over noncommutative rings, which we will look at with the help of a more sophisticated perspective in §9.8; but we won’t worry about these till then.)

Given two abelian groups A and B , let us construct an abelian group $A \otimes B$ (called the *tensor product* of A and B) as follows: We present it using a set of generators which we write $a \otimes b$, one for each $a \in |A|$, $b \in |B|$, and defining relations which are precisely the conditions required to make the map $(a, b) \mapsto a \otimes b$ bilinear; namely

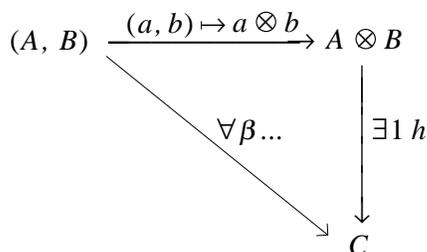
$$\begin{aligned} (a+a') \otimes b &= a \otimes b + a' \otimes b, & (a, a' \in |A|, b, b' \in |B|). \\ a \otimes (b+b') &= a \otimes b + a \otimes b' \end{aligned}$$

(If we are working with R -modules, we also need the R -module relations

$$(ra) \otimes b = r(a \otimes b) = a \otimes (rb) \quad (a \in |A|, b \in |B|, r \in |R|).$$

To indicate that one is referring to the tensor product as R -modules rather than the tensor product as abelian groups, one often writes this $A \otimes_R B$.)

By construction, $A \otimes B$ will be an abelian group with a bilinear map $\otimes: (A, B) \rightarrow A \otimes B$, and the universal property arising from its presentation translates to say that the map \otimes will be universal among bilinear maps on (A, B) .



We can get a simpler presentation of this group if we are given presentations of A and B . To describe this, let us write our presentation of A as a representation $A = F(X)/\langle S \rangle$, where $F(X)$ is the free abelian group on the set X of generators, and $\langle S \rangle$ is the subgroup of $F(X)$ generated by the family S of relators (elements that are required to go to 0). If A is so presented, and likewise B is written as $F(Y)/\langle T \rangle$, then it is not hard to show (and you may do so as the next exercise) that

$$(3.9.1) \quad A \otimes B \cong F(X) \otimes F(Y) / \langle S \otimes Y \cup X \otimes T \rangle$$

where $S \otimes Y$ means $\{s \otimes y \mid s \in S, y \in Y\} \subseteq |F(X) \otimes F(Y)|$, and $X \otimes T$ is defined analogously. One finds that $F(X) \otimes F(Y)$ is a free abelian group on the family $X \otimes Y$ (more precisely: it is a free abelian group on $X \times Y$ via the mapping $(x, y) \mapsto u(x) \otimes v(y)$, where $u: X \rightarrow |F(X)|$ and $v: Y \rightarrow |F(Y)|$ are the universal maps associated with the free groups $F(X)$ and $F(Y)$). Hence

(3.9.1) is equivalent to a presentation of $A \otimes B$ by the generating set $X \times Y$ and a certain set of relations.

In the following exercises, unless the contrary is stated, you can, if you wish, substitute “ R -module” for “abelian group”, and prove the results for this more general case.

Exercise 3.9:2. Prove (3.9.1), and the assertion that $F(X) \otimes F(Y)$ is free abelian on $X \otimes Y$. Can the “denominator” of (3.9.1) be replaced simply by $\langle S \otimes T \rangle$?

Exercise 3.9:3. (i) Given abelian groups A and C , is there a universal pair (B, β) , of an abelian group B and a bilinear map $\beta: (A, B) \rightarrow C$?

(ii) Given an abelian group C , is there a universal 3-tuple (A, B, β) , such that A and B are abelian groups and β a bilinear map $(A, B) \rightarrow C$?

Before answering each part, say what the universal property would be and whether it would be a right or left universal property. Try the approach suggested in the preceding section for finding such objects.

Why have we defined bilinear maps only for *abelian* groups? This is answered by

Exercise 3.9:4. Let F , G and H be not necessarily abelian groups (so this exercise has no generalization to R -modules), and suppose $\beta: |F| \times |G| \rightarrow |H|$ is a map such that

$$(3.9.2) \quad \begin{aligned} (\forall f \in |F|) \quad & \text{the map } g \mapsto \beta(f, g) \text{ is a group homomorphism } G \rightarrow H; \\ (\forall g \in |G|) \quad & \text{the map } f \mapsto \beta(f, g) \text{ is a group homomorphism } F \rightarrow H. \end{aligned}$$

(i) Show that the subgroup H_0 of H generated by the image of β is abelian.

(ii) Deduce that the map β has a natural factorization

$$|F| \times |G| \rightarrow |F^{\text{ab}}| \times |G^{\text{ab}}| \xrightarrow{\beta'} |H_0| \hookrightarrow |H|,$$

where β' is bilinear. Thus, the study of maps satisfying (3.9.2) is reduced to the study of bilinear maps of *abelian* groups. This makes it easy to do

(iii) For general groups F and G , deduce a description of the group H with a universal map β satisfying (3.9.2), in terms of tensor products of abelian groups.

Remark: the above exercise, together with the observation that the multiplication map of a ring is a *bilinear* operation with respect to the ring's additive group structure, show why, though one often deals with rings having noncommutative multiplication, one does not have a natural concept of “ring with noncommutative addition”.

(Nonetheless, there are sometimes ways of generalizing a concept other than the obvious ones, and some group-theorists have introduced a version of the concept of bilinear map which does not collapse in the manner described above in the noncommutative case. The student interested in this can look at [54] and papers referred to there.)

Although the image of every group homomorphism is a subgroup of the codomain group, this is not true of images of bilinear maps:

Exercise 3.9:5. (i) Let U, V, W be finite-dimensional vector spaces over a field, and consider composition of linear maps as a set map $\text{Hom}(U, V) \times \text{Hom}(V, W) \rightarrow \text{Hom}(U, W)$. Note that if we regard these hom-sets as additive groups, this map is bilinear. Suppose V is one-dimensional; describe the range of this composition map. Is it a subgroup of $\text{Hom}(U, W)$?

(ii) If A and B are abelian groups, does every element of $A \otimes B$ have the form $a \otimes b$ for some $a \in |A|, b \in |B|$? (Prove your answer, of course.)

Another important property of tensor products is noted in

Exercise 3.9:6. If A , B and C are abelian groups, show that there is a natural isomorphism $\text{Hom}(A \otimes B, C) \cong \text{Hom}(A, \text{Hom}(B, C))$.

State an analogous result holding for sets A , B , C and set maps.

To motivate the next exercise, let n be a positive integer, and \mathbb{Z}_n the cyclic group of order n , which can be presented by one generator y and one relation $ny = 0$. (Since we are using additive notation here, ny means $y + \dots + y$, with n summands.) If A is any abelian group, you should not find it hard to verify that $A \otimes \mathbb{Z}_n$ is isomorphic to A/nA , where nA is the subgroup of A consisting of all elements na ($a \in |A|$); precisely, that the homomorphism $x \mapsto x \otimes y$ is surjective, and has kernel nA .

To generalize this observation, let us replace \mathbb{Z}_n by an arbitrary abelian group B with a presentation

$$(3.9.3) \quad B = F(Y) / \langle T \rangle,$$

where $F(Y)$ is the free abelian group on Y , and T a subset of $|F(Y)|$. Given any abelian group A , one finds that $A \otimes F(Y)$ is a direct sum of copies of A , indexed by Y , $\bigoplus_Y A$. I claim now that we can get $A \otimes B$ from this group by dividing out by another sum of homomorphic images of A , indexed by T . To describe this sum, we need a way of specifying certain maps of A into $\bigoplus_Y A$. Because the latter group is a coproduct, it has associated with it a universal Y -tuple of maps,

$$q_y: A \rightarrow \bigoplus_Y A \quad (y \in Y).$$

Since $\text{Hom}(A, \bigoplus_Y A)$ is an abelian group, this Y -tuple of elements determines a homomorphism $\psi: F(Y) \rightarrow \text{Hom}(A, \bigoplus_Y A)$, such that for each $y \in Y$, $\psi(y) = q_y$. Having defined this homomorphism ψ , we can now apply it to general elements of $F(Y)$; in particular, we can define

$$(3.9.4) \quad C = (\bigoplus_Y A) / \sum_{t \in T} \psi(t)(A),$$

where the denominator means the subgroup of $\bigoplus_Y A$ generated by the images of A under all the homomorphisms $\psi(t): A \rightarrow \bigoplus_Y A$, as t ranges over the relator-set T of (3.9.3). Now for any $f \in |F(Y)|$, let $[f]$ denote its image in B (cf. (3.9.3)), and for any $x \in \bigoplus_Y A$, let $[x]$ denote its image in C (cf. (3.9.4)). The rest I leave to you:

Exercise 3.9:7. Show that the formula $\tau(a, [f]) = [\psi(f)(a)]$ gives a well-defined map $\tau: |A| \times |B| \rightarrow |C|$, that this map is bilinear, and that the pair (C, τ) has the universal property of $(A \otimes B, \otimes)$. Conclude that the abelian group described by (3.9.4) is isomorphic to $A \otimes B$.

Apply this to the case $B = \mathbb{Z}_n$, and recover the description of $A \otimes \mathbb{Z}_n$ given in the motivating remarks above.

Another interesting problem is

Exercise 3.9:8. Investigate conditions on abelian groups (or R -modules) A and B under which $A \otimes B = \{0\}$.

Although I began this discussion of bilinear maps by noting that the condition that a set map $\beta: |A| \times |B| \rightarrow |C|$ be a bilinear map $(A, B) \rightarrow C$ had nothing to do with the group structure of the direct product group, $A \times B$, there are certain relations between these concepts:

Exercise 3.9:9. (i) Show that if A and B are abelian groups, and $\beta: (A, B) \rightarrow C$ a bilinear map, then β , regarded as a map on underlying sets of groups, $|A \times B| \rightarrow |C|$, satisfies nontrivial identities. That is, show that for some m and n one can find a derived n -ary operation s , and n derived m -ary operations t_1, \dots, t_n for abelian groups, such that

$$s(\beta(t_1(x_1, \dots, x_m)), \dots, \beta(t_n(x_1, \dots, x_m))) = 0$$

holds for all $x_1, \dots, x_n \in |A \times B|$ (with the t 's evaluated using the product-group structure on $|A \times B|$), but such that the corresponding equation does not hold for arbitrary maps $\beta: |D| \rightarrow |C|$ of underlying sets of abelian groups.

(ii) On the other hand, show that the bilinearity of β cannot be characterized in terms of such identities; in other words, that there exist maps $\beta: |A| \times |B| \rightarrow |C|$ which are not bilinear maps $(A, B) \rightarrow C$, but which satisfy all identities that are satisfied by bilinear maps.

(iii) Can you find a list of identities which imply all identities satisfied by bilinear maps β , in the sense described in (i)?

In subsequent sections, we shall occasionally refer again to bilinear maps. In those situations, we may use either the notation “ $(A, B) \rightarrow C$ ” introduced here, or the more standard notation “ $A \times B \rightarrow C$ ”. (Of course, if all we have to say is something like “this map $|A| \times |B| \rightarrow |C|$ is bilinear”, we will not need either notation.)

3.10. Monoids. So far, we have been moving within the realm of groups. It is time to broaden our horizons. We begin with semigroups and monoids, objects which are very much like groups in some ways, yet quite different in others.

We recall that a *semigroup* means an ordered pair $S = (|S|, \cdot)$ such that $|S|$ is a set and \cdot a map $|S| \times |S| \rightarrow |S|$ satisfying the associative identity, while a *monoid* is a 3-tuple $S = (|S|, \cdot, e)$ where $|S|$ and \cdot are as above, and the third component, e , is a *neutral element* for the operation \cdot . As with groups, the multiplication of semigroups and monoids is most often written without the “ \cdot ” when there is no need to be explicit. A *homomorphism* of semigroups $f: S \rightarrow T$ means a set map $f: |S| \rightarrow |T|$ which respects “ \cdot ”; a *homomorphism of monoids* is required to respect neutral elements as well: $f(e_S) = e_T$.

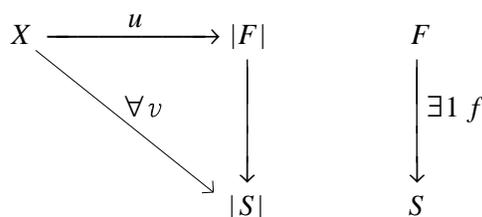
(I have long considered the use of two unrelated terms, “semigroup” and “monoid”, for these very closely related types of objects to be an unnecessary proliferation of terminology. In most areas of mathematics, distinctions between related concepts are made by modifying phrases, e.g., “abelian group” versus “not necessarily abelian group”, “ring with 1” versus “ring without 1”, “manifold with boundary” versus “manifold without boundary”. The author of a paper considering one of these concepts will generally begin by setting conventions, such as “In this note, unless the contrary is stated, rings will have unit element, and ring homomorphisms will be understood to respect this element”. In papers of mine where monoids came up, I followed the same principle for a long time, calling them “semigroups with neutral element” or, after saying what this would mean, simply “semigroups”. I did the same in these notes through 1995. However, it appears that the term “monoid” is here to stay; so since Edition/Printing 2.1, I have followed standard usage, given above.)

The concept of monoid seems somewhat more basic than that of semigroup. If X is any set, then the set of all maps $X \rightarrow X$ has a natural monoid structure, with functional composition as the multiplication and the identity map as the neutral element, and more generally, this is true of the set of endomorphisms of any mathematical object. Sets whose natural structure is one of semigroup and not of monoid tend to arise as subsidiary constructions, when one considers those elements of a naturally occurring monoid that satisfy some restriction that excludes the neutral

element; e.g., the set of maps $X \rightarrow X$ having finite range, or the set of even integers under multiplication. However, “semigroup” is the older of the two terms, so the study of semigroups and monoids is called “semigroup theory”.

If $(|S|, \cdot, e)$ is a monoid, one can, of course, look at the semigroup $(|S|, \cdot)$, while if $(|S|, \cdot)$ is a semigroup, one can “adjoin a neutral element” and get a monoid $(|S| \sqcup \{e\}, \cdot, e)$. Thus, results on monoids yield results on semigroups, and vice versa. To avoid repetitiveness, I will focus here on monoids, and mention semigroups only when there is a contrast to be made. Most of our observations on monoids will have obvious analogs for semigroups, the exceptions being those relating to invertible elements.

The concept of a free monoid (F, u) on a set X is defined using the expected universal property (diagram below).



Free monoids on all sets exist, by the general arguments of §2.2 and §2.3. One also has a normal form in the free monoid on X , analogous to that of §2.4, but without any negative exponents. That is, every element can be written uniquely as a product,

$$x_n \dots x_1,$$

where $x_1, \dots, x_n \in |X|$, and $n \geq 0$ (the product of 0 factors being understood to mean the neutral element). Multiplication is performed by juxtaposing such products. “Van der Waerden’s trick” is not needed to establish this normal form, since there is no cancellation to complicate a direct verification of associativity. Note that the free monoid on X is isomorphic to the submonoid generated by X within the free group on X .

If X is a set, and R a set of pairs of *monoid terms* in the elements of X , there will likewise exist a monoid determined by “generators X and relations R ”, i.e., a monoid S with a map $u: X \rightarrow |S|$ such that for each of the pairs of terms $(s, t) \in R$, one has $s_u = t_u$ in S , and which is universal for this property. As in the group case, this S can be obtained by a direct construction, using terms modulo identifications deducible from the monoid laws and the set of relations R , or as a submonoid of a large direct product, or by taking the free monoid F on the set X , and imposing the given relations.

But how does one “impose relations” on a monoid? In a group, we noted that any relation $x = y$ was equivalent to $xy^{-1} = e$, so to study relations satisfied in a homomorphic image of a given group G , it sufficed to study the set of elements of G that went to e , hence the construction of imposing relations reduced to that of dividing out by an appropriate normal subgroup. But for monoids, the question of which elements fall together does not come down to that of which elements go to e . For instance, let S be the free monoid on $\{x, y\}$, and map S homomorphically to the free monoid on $\{x\}$ by sending both x and y to x . Note that any product of m x ’s and n y ’s goes to x^{m+n} under this map. Thus the only element going to e is e itself, but the homomorphism is far from one-to-one.

So to study relations satisfied in the image of a monoid homomorphism $f: S \rightarrow T$, one should look at the whole set

$$K_f = \{(s, t) \mid f(s) = f(t)\} \subseteq |S| \times |S|.$$

We note the following properties of K_f :

$$(3.10.1) \quad (\forall s \in S) \ (s, s) \in K_f.$$

$$(3.10.2) \quad (\forall s, t \in S) \ (s, t) \in K_f \Rightarrow (t, s) \in K_f.$$

$$(3.10.3) \quad (\forall s, t, u \in S) \ (s, t) \in K_f, (t, u) \in K_f \Rightarrow (s, u) \in K_f.$$

$$(3.10.4) \quad (\forall s, t, s', t' \in S) \ (s, t) \in K_f, (s', t') \in K_f \Rightarrow (ss', tt') \in K_f.$$

Here (3.10.1-3) say that K_f is an equivalence relation, and (3.10.4) says that it “respects” the monoid operation.

I claim, conversely, that if S is a monoid, and $K \subseteq |S| \times |S|$ is any subset satisfying (3.10.1-4), then there exists a homomorphism f of S into a monoid T such that $K_f = K$. Indeed, since K is an equivalence relation on $|S|$, we may define $|T| = |S|/K$ and let $f: |S| \rightarrow |T|$ be the quotient map $x \mapsto [x]$. It is now easy to see from (3.10.4) that the formula $[s] \cdot [t] = [st]$ defines an operation on $|T|$, and to verify that this makes $T = (|T|, \cdot, [e])$ a monoid such that f is a homomorphism, and $K_f = K$.

Exercise 3.10:1. (i) Compare this construction with that of §2.2. Why did we need the conditions (2.2.1-3) in that construction, but not the corresponding conditions here?

(ii) Given two monoid homomorphisms $f: S \rightarrow T$ and $f': S \rightarrow T'$, show that there exist isomorphisms between their images making the diagram below commute if and only if $K_f = K_{f'}$.

$$\begin{array}{ccc} & & f(S) \subseteq T \\ S & \begin{array}{c} \nearrow \\ \searrow \end{array} & \parallel \\ & & f'(S) \subseteq T' \end{array}$$

Definition 3.10.5. For any monoid S , a binary relation K on $|S|$ satisfying (3.10.1-4) above is called a congruence on S . The equivalence class of an element is called its congruence class under K ; the monoid T constructed above is called the quotient or factor monoid of S by K , written S/K .

Given a set R of pairs of elements of a monoid S , it is clear that one can construct the *least* congruence K containing R by closing R under four operations corresponding to conditions (3.10.1-4). The quotient S/K has the correct universal property to be called the monoid obtained by imposing the relations R on the monoid S . We shall sometimes denote this S/R , or $S/(s = t \mid (s, t) \in R)$, or, if the elements of R are listed as (s_i, t_i) ($i \in I$), as $S/(s_i = t_i \mid i \in I)$.

In particular, by imposing relations on a *free* monoid, we can, as asserted earlier, get a monoid presented by any families of generators X and relations R . Like the corresponding construction for groups, this is written $\langle X \mid R \rangle$. If there is danger of ambiguity, the group- and monoid-constructions can be distinguished as $\langle X \mid R \rangle_{\text{gp}}$ and $\langle X \mid R \rangle_{\text{md}}$.

Exercise 3.10:2. Given congruences K and K' on a monoid S , will there exist a least congruence containing both K and K' ? A greatest congruence contained in both? Will set-theoretic union and intersection give such congruences? If not, what useful descriptions can you find for them? Is there a least congruence on S ? A greatest?

If K is a congruence on S , characterize congruences on $T = S/K$ in terms of congruences

on S .

Exercise 3.10:3. If S is a monoid and X a subset of $|S| \times |S|$, will there be a largest congruence contained in X ? If not, will this become true under additional assumptions, such as that X is an equivalence relation on $|S|$, or is the underlying set of a submonoid of $S \times S$?

Some observations on congruences: One can speak similarly of congruences on *groups*, *rings*, *lattices*, etc.. They are defined in each case by conditions (3.10.1-3), plus a family of conditions analogous to (3.10.4), one for each operation of positive arity on our algebras. The special fact that allowed us to give a simpler treatment in the case of groups can now be reformulated: “A congruence K on a group G is uniquely determined by the congruence class of the neutral element $e \in |G|$, which can be any *normal subgroup* N of G . The congruence classes of K are then the cosets of N in G .” Hence in group theory, rather than considering congruences, one almost always talks about normal subgroups.

Since a ring R has an additive group structure, a congruence on a ring will in particular be a congruence on its additive group, and hence will be determined by the congruence class J of the additive neutral element 0 . The possibilities for J turn out to be precisely the *ideals* of R , so in ring theory, one works with ideals rather than congruences. However, historically, the congruence relation “ $a \equiv b \pmod{n}$ ” on the ring of integers \mathbb{Z} was talked about before one had the concept of the ideal $n\mathbb{Z}$. Ring theorists still occasionally find it suggestive to write $a \equiv b \pmod{J}$ rather than $a - b \in J$.

On the other hand, for objects such as monoids and lattices, congruences cannot be reduced to anything simpler, and are studied as such.

As usual, questions of the *structure* of monoids presented by generators and relations must be tackled case by case. For example:

Exercise 3.10:4. Find a normal form or other description for the monoid presented by two generators a and b and the one relation $ab = e$.

(Note that in the above and the next few exercises, letters a through d denote general monoid elements, but e is always the neutral element. If you prefer to write 1 instead of e in your solutions, feel free to do so.)

Exercise 3.10:5. (i) Same problem for generators a, b, c, d and relations

$$ab = ac = dc = e.$$

(ii) Same problem for generators a, b, c, d and relations

$$ab = ac = cd = e.$$

Exercise 3.10:6. Same problem for generators a, b, c and relations

$$ab = ac, \quad ba = bc, \quad ca = cb.$$

Exercise 3.10:7. Same problem for generators a, b and the relation $ab = b^2a$.

One may define the *product* and the *coproduct* of two or of an arbitrary family of monoids, by the same universal properties as for groups,



These also turn out to have the same descriptions as for groups: The direct product of an I -tuple of monoids consists of all I -tuples such that for each $i \in I$, the i th position is occupied by a member of the i th monoid, with operations defined componentwise; the coproduct consists of formal products of strings of elements other than the neutral element taken from the given monoids, such that no two successive factors come from the same monoid. Van der Waerden's method is used in establishing this normal form, since multiplication of two such products can involve "cancellation" if any of the given monoids have elements satisfying $ab = e$.

On monoids, as on groups, one has the construction of *abelianization*, gotten by imposing the relations $ab = ba$ for all $a, b \in |S|$.

One may also define the *kernel* and *cokernel* of a monoid homomorphism $f: S \rightarrow S'$ as for groups:

$$(3.10.6) \quad \text{Ker } f = \text{submonoid of } S \text{ with underlying set } \{s \in |S| \mid f(s) = e\},$$

$$(3.10.7) \quad \text{Cok } f = S' / (f(s) = e \mid s \in |S|).$$

But we have seen that the structure of the image of a monoid homomorphism f is not determined by the kernel of f , and it follows that not every homomorphic image T of a monoid S' can be written as the cokernel (3.10.7) of a homomorphism of another monoid S into S' (e.g., the image of S' under a non-one-to-one homomorphism with trivial kernel cannot). Hence these concepts of kernel and cokernel are not as important in the theory of monoids as in group theory.

We have seen that for f a homomorphism of monoids, a better analog of the group-theoretic concept of kernel is the congruence

$$(3.10.8) \quad K_f = \{(s, t) \mid f(s) = f(t)\} \subseteq |S| \times |S|.$$

Note that K_f is the underlying set of a submonoid of $S \times S$, which we may call $\text{Cong } f$. Likewise, since to impose relations on a monoid we specify, not that some elements should go to e , but that some pairs of elements should fall together, it seems reasonable that a good generalization of the cokernel concept should be, not an image $q(S)$ universal for the condition $qf = e$, where f is a given monoid homomorphism into S , but an image $q(S)$ universal for the condition $qf = qg$, for some *pair* of homomorphisms

$$(3.10.9) \quad f, g: T \rightarrow S.$$

Given f and g as above, $q(S)$ may be constructed as the quotient of the monoid S by the congruence generated by all pairs $(f(t), g(t))$ ($t \in |T|$). Postponing, for the moment, the question of what $q(S)$ will be called, let us note that there is a dual construction: Given f, g as in (3.10.9), one can get a universal map p into T such that $fp = gp$. This will be given by inclusion in T of the submonoid whose underlying set is $\{t \mid f(t) = g(t)\}$, called the *equalizer* of f and g . Dually, the $q(S)$ constructed above is called the *coequalizer* of f and g .

Exercise 3.10:8. Let $f: S \rightarrow T$ be a monoid homomorphism.

- (i) Note that there is a natural pair of monoid homomorphisms from $\text{Cong } f$ to S . Characterize the 3-tuple formed by $\text{Cong } f$ and these two maps by a universal property.
- (ii) What can be said of the equalizer and coequalizer of the above pair of maps?
- (iii) Can you construct from f a monoid $\text{CoCong } f$ with a pair of maps into it, having a dual universal property? If so, again, look at the equalizer and coequalizer of this pair.

Exercise 3.10:9. The definition of equalizer can be applied to groups as well as monoids. If G is a group, investigate which subgroups of G can occur as equalizers of pairs of homomorphisms on G .

3.11. Groups to monoids and back again. If S is a monoid, we can get a group S^{gp} from S by “adjoining inverses” to all its elements in a universal manner. Thus, S^{gp} is a group G having a map $q: |S| \rightarrow |G|$ which respects products and neutral elements, and is universal among all such maps from S to groups.

But what kind of a map, exactly, is q ? Since $S = (|S|, \cdot, e)$ is a monoid while $S^{\text{gp}} = G = (G, \cdot, {}^{-1}, e)$ is a group, we cannot call it a group homomorphism or a monoid homomorphism from S to G . But it is more than just a set map, since it respects \cdot and e . The answer is that q is a monoid homomorphism from S to the monoid $(|G|, \cdot, e)$ (i.e., $(|G|, \mu_G, e_G)$). So for an arbitrary group $H = (|H|, \mu_H, \iota_H, e_H)$, let us write H_{md} for $(|H|, \mu_H, e_H)$, that is, “ H considered as a monoid”. We can now state the universal property of S^{gp} and q neatly: S^{gp} is a group G , and q is a monoid homomorphism from S to G_{md} , such that for any group H and any monoid homomorphism $a: S \rightarrow H_{\text{md}}$, there exists a unique group homomorphism $f: G \rightarrow H$ such that $a = fq: S \rightarrow H_{\text{md}}$.

$$\begin{array}{ccc}
 S & \xrightarrow{q} & G_{\text{md}} & & G \\
 & \searrow \forall a & \downarrow & & \downarrow \exists! f \\
 & & H_{\text{md}} & & H
 \end{array}$$

We shall call S^{gp} the *universal enveloping group* of the monoid S . It may be presented as a group by taking a generator for each element of $|S|$, and taking for defining relations the full multiplication table of S . More generally, if we are given some presentation of S by generators and relations as a monoid, G will be a group presented by the same generators and relations.

Exercise 3.11:1. Show that a monoid S is “embeddable in a group” (meaning embeddable in the monoid H_{md} for some group H) if and only if the universal map $q: |S| \rightarrow |S^{\text{gp}}|$ is one-to-one.

Exercise 3.11:2. Describe the universal enveloping groups of the monoids of Exercises 3.10:4-7, and also of the monoid presented by generators a, b, c and the one relation $ab = ac$.

The last part of the above exercise reveals one necessary condition for the one-one-ness of the exercise preceding it to hold: The monoid S must have the *cancellation* property $xy = xy' \Rightarrow y = y'$. An interesting way of obtaining a full set of necessary and sufficient conditions for the universal map of a given monoid into a group to be one-to-one was found by A. I. Mal'cev ([96], [97] also described in [6, §VII.3]).

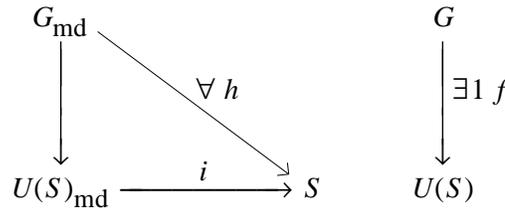
Exercise 3.11:3. Let G be a group and S a *submonoid* of G_{md} , which generates G as a group. Observe that the inclusion of S in G_{md} induces a homomorphism $S^{\text{gp}} \rightarrow G$. Will this in general be one-to-one? Onto?

If you have done Exercise 3.3:3, consider the case where G is the group of that exercise, and S the submonoid generated by a and ba . Describe the structure of S and of S^{gp} .

Suppose S is an *abelian* monoid. In this situation, important applications of the universal enveloping group construction have been made by A. Grothendieck; the group S^{gp} for S an

abelian monoid is therefore often called “the Grothendieck group $K(S)$ ”. This group is also abelian, and has a simple description: Using additive notation, and writing \bar{a} for $q(a)$, one finds that every element of $K(S)$ can be written $\bar{a} - \bar{b}$ ($a, b \in |S|$), and that one has equality $\bar{a} - \bar{b} = \bar{a}' - \bar{b}'$ between two such elements if and only if there exists $c \in |S|$ such that $a + b' + c = a' + b + c$ [31, p.40]. (If you have seen the construction of the *localization* RS^{-1} of a commutative ring at a multiplicative subset S , you will see that these constructions are closely related. In particular, the multiplicative group of nonzero elements of the field of fractions F of a commutative integral domain R is the Grothendieck group of the multiplicative monoid of nonzero elements of R .) The application of this construction to the abelian monoid of isomorphism classes of finite-dimensional vector bundles on a topological space X , with monoid operation corresponding to the operation “ \oplus ” on vector bundles, is the starting point of “ K -theory”. But perhaps this idea has been pushed too much – it is annoyingly predictable that when I mention to a fellow algebraist a monoid of isomorphism classes of modules under “ \oplus ”, he or she will say, “oh, and then you take its Grothendieck group,” when in fact I wanted to talk about the monoid itself.

Given a monoid S , there is also a *right-universal* way of obtaining a group: The set of *invertible elements* (“units”) of S can be made a group $U(S)$ in an obvious way, and the inclusion $i: U(S) \rightarrow S$ is universal among “homomorphisms of groups into S ”, in the sense indicated in the diagram below.



Exercise 3.11:4. Let S be the monoid defined by generators x, y, z and relations $xyz = e, zxy = e$. Investigate the structures of S and its abelianization S^{ab} . Describe the groups $U(S), U(S)^{\text{ab}}$, and $U(S^{\text{ab}})$.

The two constructions that relate *semigroups* to monoids mentioned near the beginning of the preceding section are related to each other in a way paralleling the relation between $()^{\text{gp}}$ and $()_{\text{md}}$:

Exercise 3.11:5. (i) If $S = (|S|, \cdot)$ is a semigroup, describe how to extend the multiplication “ \cdot ” to $|S| \sqcup \{e\}$ so that $(|S| \sqcup \{e\}, \cdot, e)$ becomes a monoid.

Let us call the monoid resulting from the above construction S^{md} , while if $S' = (|S'|, \cdot, e)$ is a monoid, let us write S'_{sg} for the semigroup $(|S'|, \cdot)$.

(ii) Show that given a semigroup S , the monoid S^{md} is universal among monoids T given with semigroup homomorphisms $S \rightarrow T_{\text{sg}}$.

(iii) Given a monoid $S = (|S|, \cdot, e)$, what is the relation between the monoids S and $(S_{\text{sg}})^{\text{md}}$? Is there a natural homomorphism in either direction between them?

3.12. Associative and commutative rings. An *associative ring* R means a 6-tuple

$$R = (|R|, +, \cdot, -, 0, 1)$$

such that $(|R|, +, -, 0)$ is an abelian group, $(|R|, \cdot, 1)$ is a monoid, and the monoid operation $\cdot : |R| \times |R| \rightarrow |R|$ is *bilinear* with respect to the additive group structure. (Dropping the “1” from this definition, one gets a concept of “ring without 1”, but we shall not consider these except in one exercise near the end of this section.) A *ring homomorphism* is a map of underlying sets respecting all the operations, including 1. (Some writers, although requiring their rings to have 1, perversely allow “homomorphisms” that may not preserve 1; but we shall stick to the above more sensible definition.) An associative ring is called *commutative* if the multiplication \cdot is so.

“Commutative associative ring” is usually abbreviated to “commutative ring”. Depending on the focus of a given work, *either* the term “associative ring” *or* the term “commutative ring” is usually shortened further to “ring”; an author should always make clear what his or her usage will be. Here, I shall generally shorten “associative ring” to “ring”; though I will sometimes retain the word “associative” when I want to emphasize that commutativity is not being assumed.

(When one deals with *nonassociative* rings – which we shall not do in this chapter – it is the associativity condition on the *multiplication* that is removed; frequently one then considers in its place other identities, which may involve both addition and the multiplication; for instance, the identities of Lie rings, which we will introduce in §8.7, or of Jordan rings, mentioned at the end of that section. In the definition of a given kind of nonassociative ring, it may or may not be natural to have a 1 or other distinguished element. The assumption that $(|R|, +, -, 0)$ is an abelian group, and that multiplication is a bilinear map with respect to this group structure, is made in all versions of ring theory: commutative, associative and nonassociative. If weaker assumptions are made – in particular, if this abelian group structure is replaced by a monoid or semigroup structure, and/or if multiplication is only assumed linear in one of its two arguments – the resulting structures are given names such as “semiring”, “half-ring” or “near-ring”.)

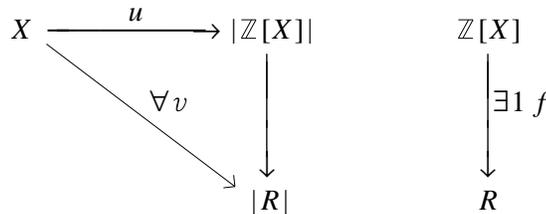
If k is a fixed commutative ring, then k -modules form a natural generalization of abelian groups, on which a concept of bilinear map is also defined, as noted parenthetically in §3.9 above. Hence one can generalize the definition of associative ring by replacing the abelian group structure by a k -module structure, and the bilinear map of abelian groups by a bilinear map of k -modules. The result is the definition of an *associative algebra* over k . The reader familiar with these concepts may note that everything I shall say below for rings remains valid, *mutatis mutandis*, for k -algebras. (An associative k -algebra is sometimes defined differently, as ring R given with a homomorphism of k into its center; but the two formulations are equivalent: Given a k -algebra R in the present sense of a ring with appropriate k -module structure, the map $c \mapsto c1_R$ ($c \in k$) is easily shown to be a homomorphism of k into the center of R ; while given a homomorphism g of k into the center of a ring R , the definition $c \cdot r = g(c)r$ gives an appropriate module structure; and these constructions are inverse to one another. For algebras without 1, and for nonassociative algebras, this equivalence does not hold, and the “ring with k -module structure” definition is then the useful one.)

The subject of universal constructions in ring theory is a vast one. In this section and the next, we will mainly look at the analogs of some of the constructions we have considered for groups and monoids.

First, free rings. Let us begin with the commutative case, since that is the more familiar one. Suppose R is a commutative ring, and x, y, z are three elements of R . Given any ring-theoretic combination of x, y and z , we can use the distributive law of multiplication (i.e., bilinearity of \cdot) to expand this as a sum of products of x, y and z (monomials) and additive inverses of such products. Using the commutativity and associativity of multiplication, we can write each monomial so that all factors x come first, followed by all y 's, followed by all z 's. We can then

use commutativity of addition to bring together all occurrences of each monomial (arranging the distinct monomials in some specified order), and finally combine occurrences of the same monomial using integer coefficients. If we now consider all ring-theoretic terms in *symbols* x, y and z , of the forms to which we have just shown we can reduce any combination of *elements* x, y and z in any ring, we see, by the same argument as in §2.4, that the set of these “reduced terms” should give a normal form for the free commutative ring on three generators x, y and z – if they form a commutative ring under the obvious operations. It is, of course, well known that the set of such expressions *does* form a commutative ring, called the *polynomial ring* in three indeterminates, and written $\mathbb{Z}[x, y, z]$.

So polynomial rings over \mathbb{Z} are free commutative rings. (More generally, the free commutative k -algebra on a set X is the polynomial algebra $k[X]$.) The universal mapping property corresponds to the familiar operation of *substituting values* for the indeterminates in a polynomial.



When we drop the commutativity assumption and look at general associative rings, the situation is similar, except that we cannot rewrite each monomial so that “all x ’s come first” etc.. Thus we end up with linear combinations with coefficients in \mathbb{Z} of arbitrary products of our generators. We claim that formal linear combinations of such products give a normal form for elements of the free associative ring on the set X . This ring is written $\mathbb{Z}\langle X \rangle$, and sometimes called the ring of *noncommuting polynomials* in X .

We were sketchy in talking about $\mathbb{Z}[X]$ because it is a well-known construction, but let us stop and sort out just what we mean by the above description of $\mathbb{Z}\langle X \rangle$, before looking for a way to prove it.

We could choose a particular way of arranging the parentheses in every monomial term (say, nested to the right), a particular way of arranging the different monomials, and of arranging the parentheses in every sum or difference, and so obtain a set of *ring-theoretic terms* to which every term could be reduced, which we would prove constituted a normal form for the free ring. But observe that the question of putting parentheses into monomial terms is really just one of how to write elements in a *free monoid*, while the question of expressing sums and differences is that of describing an element of the free abelian group on a set of generators. Let us therefore assume that we have chosen one or another way of calculating in free abelian groups – whether using a normal form, or a representation by integer-valued functions with only finitely many nonzero values, or whatever – and likewise that we have chosen a way of calculating in free monoids. Then we can calculate in free rings! A precise statement is

Lemma 3.12.1. *Let $\mathbb{Z}\langle X \rangle$ denote the free ring on the set X . Then the additive group of $\mathbb{Z}\langle X \rangle$ is a free abelian group on the set of products in $\mathbb{Z}\langle X \rangle$ of elements of X (including the empty product, 1), and this set of products forms, under the multiplication of $\mathbb{Z}\langle X \rangle$, a free monoid on X .*

Proof. Let S denote the free monoid on X , and $F(|S|)$ the free abelian group on the

underlying set of this monoid. We shall begin by describing a map $F(|S|) \rightarrow |\mathbb{Z}\langle X \rangle|$.

If we write u for the universal map $X \rightarrow |\mathbb{Z}\langle X \rangle|$, then by the universal property of free monoids, u induces a homomorphism u' from the free monoid S into the multiplicative monoid of $\mathbb{Z}\langle X \rangle$. Hence by the universal property of free abelian groups, there exists a unique abelian group homomorphism u'' from the free abelian group $F(|S|)$ into the additive group of $\mathbb{Z}\langle X \rangle$ whose restriction to $|S|$ is given by u' . Clearly the image of the monoid S in $\mathbb{Z}\langle X \rangle$ is closed under multiplication and contains the multiplicative neutral element; it is easy to deduce from this and the distributive law that the image of the abelian group $F(|S|)$ is closed under all the ring operations. (Note that our considerations so far are valid with $\mathbb{Z}\langle X \rangle$ and u replaced by any ring R and set map $X \rightarrow |R|$.) Since this image contains X , and $\mathbb{Z}\langle X \rangle$ is generated as a ring by X , the image is all of $|\mathbb{Z}\langle X \rangle|$, i.e., u'' is surjective. (The above argument formalizes our observation that every element of the subring generated by an X -tuple of elements of an arbitrary ring R can be expressed as a linear combination of products of elements of the given X -tuple.)

We now wish to show that u'' is one-to-one. To do this it will suffice to show that there is *some* ring R with an X -tuple v of elements, such that under the induced homomorphism $\mathbb{Z}\langle X \rangle \rightarrow R$, any two elements of $\mathbb{Z}\langle X \rangle$ which are images of distinct elements of $F(|S|)$ are mapped to distinct elements of R .

How do we find such an R ? Van der Waerden's trick for groups suggests that we should obtain it from some natural *representation* of the desired free ring. We noted in §2.4 that the *group* operations and identities arise as the operations and identities of the permutations of a set, so for "representations" of groups, we used actions on sets. The operations and identities for associative rings arise as the natural structure on the set of all endomorphisms of an abelian group A – one can compose such endomorphisms, and add and subtract them, and under these operations they form a ring $\text{End}(A)$. So we should look for an appropriate family of endomorphisms of some abelian group to represent $\mathbb{Z}\langle X \rangle$.

Let us, as in (2.4.5), introduce a symbol a ; let Sa denote the set of symbols $x_n \dots x_1 a$ ($x_i \in X$, $n \geq 0$); and this time let us further write $F(Sa)$ for the free abelian group on this set Sa . For every $x \in X$, let $\bar{x}: Sa \rightarrow Sa$ denote the map carrying each symbol $b \in Sa$ to the symbol xb . This extends uniquely (by the universal property of free abelian groups) to an additive group homomorphism $\bar{x}: F(Sa) \rightarrow F(Sa)$. Thus $(\bar{x})_{x \in X}$ is an X -tuple of elements of the associative ring $\text{End}(F(Sa))$.

Taking $R = \text{End}(F(Sa))$, the above X -tuple induces a homomorphism

$$f: \mathbb{Z}\langle X \rangle \rightarrow R.$$

Now given any element of $F(|S|)$, which we may write

$$(3.12.2) \quad r = \sum_{s \in |S|} n_s s \quad (n_s \in \mathbb{Z}, \text{ almost all } n_s = 0),$$

we verify easily that the element $f(u''(r)) \in \text{End}(F(Sa))$ carries a to $\sum n_s sa$. Hence distinct elements (3.12.2) must give distinct elements $u''(r) \in \mathbb{Z}\langle X \rangle$, which proves the one-one-ness of u'' and establishes the lemma. \square

For many fascinating results and open problems on free algebras, see [62]. For a smaller dose, you could try my paper [38], which answers the question, "When do two elements of a free algebra commute?" That problem is not of great importance itself, but it leads to the development of a number of beautiful and useful ring-theoretic tools.

Exercise 3.12:1. Let α denote the automorphism of the polynomial ring $\mathbb{Z}[x, y]$ which interchanges x and y . It is a standard result that the fixed ring of α , i.e., $\{a \in \mathbb{Z}[x, y] \mid \alpha(a) = a\}$, can be described as the polynomial ring in the two elements $x+y$ and xy .

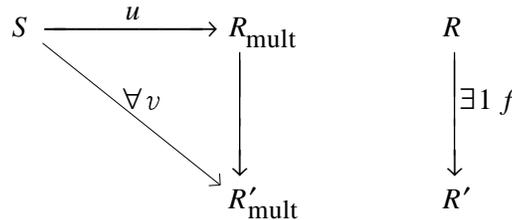
(i) Consider analogously the automorphism β of the free associative ring $\mathbb{Z}\langle x, y \rangle$ interchanging x and y . Show that the fixed ring of β is generated by the elements $x+y, x^2+y^2, x^3+y^3, \dots$ and is a free ring on this infinite set.

(ii) Observe that the homomorphism $\mathbb{Z}\langle x, y \rangle \rightarrow \mathbb{Z}[x, y]$ taking x to x and y to y must take the fixed ring of β into the fixed ring of α . Will it take it *onto* the fixed ring of α ?

(iii) If G is the free group on generators x and y , and if γ is the automorphism interchanging x and y in this group, describe the fixed subgroup of γ . Do the same for the free *abelian* group on x and y . (The analog of (ii) is trivial to answer when this has been done.)

The preceding description of the free ring on a set X involved the free monoid on X , and we see that our earlier description of the free commutative ring (the polynomial ring) has an analogous relationship to the free commutative monoid. These connections between rings and monoids can be explained in terms of another universal construction:

If $R = (|R|, +, \cdot, -, 0, 1)$ is an associative ring, let R_{mult} denote its multiplicative monoid, $(|R|, \cdot, 1)$. Then for any monoid S , there will exist, by the usual arguments, a ring R with a *universal* monoid homomorphism $u: S \rightarrow R_{\text{mult}}$.



To study this object, let us fix S , and consider any ring R' with a homomorphism $S \rightarrow R'_{\text{mult}}$. The elements of R' that we can capture using this map are the linear combinations of images of elements of S , with integer coefficients. (Why is there no need to mention products of such linear combinations?) One finds that the *universal* such ring R will have as additive structure the free abelian group on $|S|$, with multiplicative structure determined by the condition that the given map $|S| \rightarrow |R|$ respect multiplication, together with the bilinearity of multiplication. The result is called the *monoid ring* on S , denoted $\mathbb{Z}S$.

Given a presentation of S by generators and relations (written multiplicatively), a presentation of $\mathbb{Z}S$ as a ring will be given by the same generators and relations. In particular, if we take for S the *free* monoid on a set X , presented by the generator set X and no relations, then $\mathbb{Z}S$ will be presented as a *ring* by generators X and no relations, and so will be the free ring on X , which is just what we saw in Lemma 3.12.1. If we take for S a free *abelian* monoid, then S may be presented as a monoid by generators X and relations $xy = yx$ ($x, y \in X$), hence this is also a presentation of $\mathbb{Z}S$ as a ring. Since commutativity of a set of generators of a ring is equivalent to commutativity of the whole ring, the above presentation makes $\mathbb{Z}S$ the free commutative ring on X .

If S is a monoid, then a “linear action” or “representation” of S on an abelian group A means a homomorphism of S into the multiplicative monoid of the endomorphism ring $\text{End}(A)$ of A . By the universal property of $\mathbb{Z}S$, this is equivalent to a ring homomorphism of $\mathbb{Z}S$ into $\text{End}(A)$, which is in turn equivalent to a structure of left $\mathbb{Z}S$ -module on the abelian group A . In

particular, to give an action of a *group* G by automorphisms on an abelian group A corresponds to making A a left module over the *group ring* $\mathbb{Z}G$. Much of modern group theory revolves around linear actions, and hence is closely connected with the properties of $\mathbb{Z}G$ (and more generally, with group *algebras* kG where k is a commutative ring, so that left kG -modules correspond to actions of G on k -modules). For some of the elementary theory, see [31, Chapter XVIII]. A major work on group algebras is [105].

Above, we “factored” the construction of the free associative or commutative ring on a set X into two constructions: the free (respectively, free abelian) monoid construction, which universally closes X under a multiplication with a neutral element, and the monoid-ring construction, which brings in an additive structure in a universal way. These constructions can also be factored the other way around! Given a set X , we can first map it into an abelian group in a universal way, getting the free abelian group A on X , then form a ring (respectively a commutative ring) R with a universal additive group homomorphism $A \rightarrow R_{\text{add}}$. For any abelian group A , the associative ring with such a universal homomorphism is called the *tensor ring* on A , because its additive group structure turns out to have the form

$$\mathbb{Z} \oplus A \oplus (A \otimes A) \oplus (A \otimes A \otimes A) \oplus \dots,$$

though we shall not show this here. The corresponding universal *commutative* ring is called the *symmetric ring* on A ; its structure for general A is more difficult to describe. For more details see [31, §§XVI.7, 8] or [49]. Thus, a free associative ring can be described as the tensor ring on a free abelian group, and a polynomial ring as the symmetric ring on a free abelian group.

On to other constructions. Suppose R is a commutative ring, and (f_i, g_i) ($i \in I$) a family of pairs of elements of R . To impose the relations $f_i = g_i$ on R , one forms the factor-ring R/J , where J is the ideal generated by the elements $f_i - g_i$. This ideal is often written $(f_i - g_i)_{i \in I}$. Another common notation, preferable because it is more explicit, is $\sum_{i \in I} R(f_i - g_i)$, or, if we set $U = \{f_i - g_i \mid i \in I\}$, simply RU . It consists of all sums

$$(3.12.3) \quad \sum r_i(f_i - g_i) \quad (r_i \in |R|, \text{ nonzero for only finitely many } i \in I).$$

The construction of imposing relations on a *noncommutative* ring R is of the same form, but with “ideal” taken to mean a *two-sided ideal* – an additive subgroup of R closed under both left and right multiplication by members of R . The two-sided ideal generated by $\{f_i - g_i \mid i \in I\}$ is also often written $(f_i - g_i)_{i \in I}$, and again there is a more expressive notation, $\sum_{i \in I} R(f_i - g_i)R$, or simply RUR . This ideal consists of all sums of products of the form $r(f_i - g_i)r'$ ($i \in I$, $r, r' \in R$). Note, however, that in the noncommutative case, it is not in general enough to have, as in (3.12.3), *one* such summand for each $i \in I$. For instance, in $\mathbb{Z}\langle x, y \rangle$, the ideal generated by the one element x contains the element $xyx^2 + y^2xy$, which cannot be simplified to a single product rxr' .

Exercise 3.12:2. Let R be a commutative ring. Will there, in general, exist a universal homomorphism of R into an *integral domain* R' ? If not, can you find conditions on R for such a homomorphism to exist? Suggestion: Consider the cases $R = \mathbb{Z}$, \mathbb{Z}_6 , \mathbb{Z}_4 .

Exercise 3.12:3. (i) Obtain a normal form for elements of the associative ring A presented by two generators x, y , and one relation $yx - xy = 1$.

(ii) Let $\mathbb{Z}[x]_{\text{add}}$ be the *additive group* of polynomials in one indeterminate x . Show that there exists a homomorphism f of the ring A of part (i) into the endomorphism ring of this abelian group, such that $f(x)$ is the operation of multiplying polynomials by x in $\mathbb{Z}[x]$, and

$f(y)$ the operation of differentiating with respect to x . Is this homomorphism one-to-one?

The ring of the above example, or rather the corresponding algebra over a field k , is called the *Weyl algebra*. It is of importance in quantum mechanics, where multiplication by the coordinate function x corresponds to determining the x -coordinate of a particle, and differentiating with respect to x corresponds to determining its momentum in the x -direction. The fact that these operators do not commute leads, via the mysterious reasoning of quantum mechanics, to the impossibility of measuring those two quantities simultaneously, the simplest case of the ‘‘Heisenberg uncertainty principle’’.

Direct products $\prod_I R_i$ of associative rings and of commutative rings turn out, as expected, to be gotten by taking direct products of underlying sets, with componentwise operations.

Exercise 3.12:4. (Andreas Dress) (i) Find all subrings of $\mathbb{Z} \times \mathbb{Z}$. (Remember: a subring must have the same multiplicative neutral element 1. Try to formulate your description of each such subring R as a necessary and sufficient condition for an arbitrary $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ to lie in $|R|$.)

A much harder problem is:

(ii) Is there a similar characterization of all subrings of $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$?

Exercise 3.12:5. Show that the commutative ring presented by one generator x , and one relation $x^2 = x$, is isomorphic (as a ring) to the product ring $\mathbb{Z} \times \mathbb{Z}$.

Exercise 3.12:6. Given generators and relations for two rings, R and S , show how to obtain generators and relations for $R \times S$.

Exercise 3.12:7. Describe

(i) the commutative ring A presented by one generator x , and one relation $2x = 1$, and

(ii) the commutative ring B presented by one generator x and two relations $4x = 2$, $2x^2 = x$. (Note that both of these relations are implied by the relation of (i).)

Are these two rings isomorphic? Show that each of them has the property that for any ring R (commutative if you wish) there is *at most* one homomorphism of the indicated ring (A , respectively B) into R .

Exercise 3.12:8. Suppose R is a ring whose underlying abelian group is finitely generated. Show that as a ring, R is finitely presented. (You may use the fact that every finitely generated abelian group is finitely presented.)

If you are comfortable with algebras over a commutative ring k , try to generalize this result to that context.

In discussing universal properties, I have neglected to mention some trivial cases. Let me give these in the next two exercises. Even if you do not write them up, think through the ‘‘ring’’ cases of parts (i) and (iii) of the next exercise, since some later exercises use them.

Exercise 3.12:9. (i) Consider the free group, the free monoid, the free associative ring, and the free commutative ring on the *empty* set of generators. Reformulate the universal properties of these objects in as simple a form as possible. Display the group, monoid, ring, and commutative ring characterized by these properties, if they exist.

(ii) State, similarly, the universal properties that would characterize the product and coproduct of an empty family of groups, monoids, rings, or commutative rings, and determine these objects, if any.

(iii) Give as simple as possible a system of defining generators and relations for the rings \mathbb{Z} and \mathbb{Z}_n .

The next exercise concerns semigroups, and rings without neutral elements. Note that when we say ‘‘without 1’’ etc., this does not *forbid* the existence of an element 1 satisfying $(\forall x) 1x =$

$x = x1$. It just means that we don't require the existence of such elements, and that when they exist, we don't give them a special place in the definition, or require homomorphisms to respect them.

Exercise 3.12:10. Same as parts (i) and (ii) of the preceding exercise, but for semigroups, and for rings without 1. Same for sets. Same for G -sets for a fixed group G .

Now back to rings with 1.

3.13. Coproducts and tensor products of rings. We have noted that the descriptions of *coproducts* vary from one sort of algebraic object to another, so it will not be surprising to find that they have different forms for commutative and noncommutative rings. Let us again start with the commutative case.

Suppose S and T are fixed commutative rings, and we are given homomorphisms $s \mapsto \bar{s}$ and $t \mapsto \tilde{t}$ of these into a third commutative ring R . What elements of R can we capture? Obviously, elements \bar{s} ($s \in |S|$) and \tilde{t} ($t \in |T|$); from these we can form products $\bar{s}\tilde{t}$, and we can then form sums of elements of all these sorts:

$$(3.13.1) \quad \bar{s} + \tilde{t} + \bar{s}_1\tilde{t}_1 + \dots + \bar{s}_n\tilde{t}_n.$$

We don't get more elements by multiplying such sums together, because a product $(\bar{s}\tilde{t})(\bar{s}'\tilde{t}')$ reduces to $\overline{ss'}\widetilde{tt'}$. Let us note that the lone summands \bar{s} and \tilde{t} in (3.13.1) can actually be written in the same form as the others, for $\overline{1_S} = \widetilde{1_T} = 1_R$, hence $\bar{s} = \overline{s}1_T$ and $\tilde{t} = \widetilde{t}1_S$. So the subring of R that we get is generated as an additive group by the image of the map

$$(3.13.2) \quad (s, t) \mapsto \bar{s}\tilde{t}$$

of $|S| \times |T|$ into $|R|$. If we look for equalities among sums of elements of this form, we find

$$\overline{(s+s')}\tilde{t} = \bar{s}\tilde{t} + \bar{s}'\tilde{t}, \quad \text{and} \quad \overline{s(t+t')} = \bar{s}\tilde{t} + \bar{s}\tilde{t}',$$

in other words, relations saying that (3.13.2) is bilinear. These relations and their consequences turn out to be *all* we can find, and one can show that the *universal* R with ring homomorphisms of S and T into it, that is, the coproduct of S and T as commutative rings, has the additive structure of the *tensor product* of the additive groups of S and T . The elements that we have written \bar{s} and \tilde{t} are, as the above discussion implies, $s \otimes 1_T$ and $1_S \otimes t$ respectively; the multiplication is determined by the formula

$$(3.13.3) \quad (s \otimes t)(s' \otimes t') = ss' \otimes tt'$$

which specifies how to multiply the additive generators of the tensor product group. For a proof that this extends to a bilinear operation on all of $S \otimes T$, and that this operation makes the additive group $S \otimes T$ into a ring, see Lang [31, §XVI.6]. (Note: Lang works in the context of algebras over a ring k , and he defines such an algebra as a homomorphism f of k into the center of a ring R – what I prefer to call, for intuitive comprehensibility, a ring R given with a homomorphism of k into its center; cf. parenthetical remark near the beginning of §3.12 above. Thus, when he defines the coproduct of two commutative k -algebras to be a certain *map*, look at the *codomain* of the map to see the ring that he means. Alternatively, instead of looking in Lang for this construction, you might do Exercise 3.13:5 below, which gives a generalization of this result.)

This coproduct construction is called the “tensor product of commutative rings”.

Exercise 3.13:1. If m and n are integers, find the structure of the tensor product ring $\mathbb{Z}_m \otimes \mathbb{Z}_n$ by two methods:

- (i) By constructing the tensor product of the abelian groups \mathbb{Z}_m and \mathbb{Z}_n , and describing the multiplication characterized above.
- (ii) By using the fact that a presentation of a coproduct can be obtained by “putting together” presentations for the two given objects. (Cf. Exercise 3.12:9.)

Exercise 3.13:2. Let $\mathbb{Z}[i]$ denote the ring of *Gaussian integers* (complex numbers $a+bi$ such that a and b are integers). This may be presented as a commutative ring by one generator i , and one relation $i^2 = -1$. Examine the structures of the rings $\mathbb{Z}[i] \otimes \mathbb{Z}_p$ (p a prime). E.g., will they be integral domains for all p ? For some p ?

The next two exercises concern tensor products of algebras over a field k , for students familiar with this concept. Tensor products of this sort are actually simpler to work with than the tensor products of rings described above, because every algebra over a field k is free as a k -module (since every k -vector-space has a basis), and tensor products of free modules are easily described (cf. lines following (3.9.1) above).

Exercise 3.13:3. Let K and L be extensions of a field k . A *compositum* of K and L means a 3-tuple (E, f, g) where E is a field extension of k , and $f: K \rightarrow E$, $g: L \rightarrow E$ are k -algebra homomorphisms such that E is generated by $f(|K|) \cup g(|L|)$ as a field (i.e., under the ring operations, and the *partial operation* of multiplicative inverse).

- (i) Suppose K and L are *finite-dimensional* over k , and we form their tensor product algebra $K \otimes_k L$, which is a commutative k -algebra, but not necessarily a field. Show that up to isomorphism, all the composita of K and L over k are given by the factor rings $(K \otimes L)/P$, for prime ideals $P \subseteq K \otimes L$. (First write down what should be meant by an isomorphism between composita of K and L .)
- (ii) What if K and L are not assumed finite-dimensional?

Exercise 3.13:4. (i) Determine the structure of the tensor product $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$, where \mathbb{C} is the field of complex numbers and \mathbb{R} the field of real numbers. In particular, can it be written as a nontrivial direct product of \mathbb{R} -algebras?

- (ii) Do the same for $\mathbb{Q}(2^{1/3}) \otimes_{\mathbb{Q}} \mathbb{Q}(2^{1/3})$.
- (iii) Relate the above results to the preceding exercise.

You can carry this exercise much farther if you like – find a general description of a tensor product of a finite Galois extension with itself; then of two arbitrary finite separable field extensions (by taking them to lie in a common Galois extension, and considering the subgroups of the Galois group they correspond to); then try some examples with inseparable extensions... . In fact, one modern approach to the whole subject of Galois theory is via properties of such tensor products.

If S and T are arbitrary (not necessarily commutative) associative rings, one can still make the tensor product of the additive groups of S and T into a ring with a multiplication satisfying (3.13.3). It is not hard to verify that this will be universal among rings R given with homomorphisms $f: S \rightarrow R$, $g: T \rightarrow R$ such that all elements of $f(S)$ *commute with* all elements of $g(T)$. (Cf. the “second universal property” of the direct product of two groups, end of §3.6 above. In fact, some early ring-theorists wrote $S \times T$ for what we now denote $S \otimes T$, considering this construction as the ring-theoretic analog of the direct product construction on groups.) This verification is the exercise mentioned earlier as an alternative to looking in Lang for the universal property of the tensor product of commutative rings:

Exercise 3.13:5. Verify the above assertion that given rings S and T , the universal ring with mutually commuting homomorphic images of S and T has the additive structure of $S \otimes T$, and multiplication given by (3.13.3). (Suggestion: Map the additive group $S \otimes T$ onto that universal ring R , then use “van der Waerden’s trick” to show the map is an isomorphism.) Obtain as a corollary the characterization of the coproduct of commutative rings referred to earlier.

Exercise 3.13:6. Show that if S and T are monoids, then $\mathbb{Z}S \otimes \mathbb{Z}T \cong \mathbb{Z}(S \times T)$.

Exercise 3.13:7. Suppose S and T are associative rings, and we form the additive group $R_{\text{add}} = S_{\text{add}} \otimes T_{\text{add}}$. Is the multiplication described above in general the unique multiplication on R_{add} which makes it into a ring R such that the maps $s \mapsto s \otimes 1_T$ and $t \mapsto 1_S \otimes t$ are ring homomorphisms? You might look, in particular, at the case $S = \mathbb{Z}[x]$, $T = \mathbb{Z}[y]$.

Let us now look at coproducts of not necessarily commutative rings, writing these $S * T$ as for groups and monoids. They exist by the usual general nonsense, and again, a presentation of $S * T$ can be gotten by putting together presentations of S and T . But the explicit description of these coproducts is more complicated than for the constructions we have considered so far. For S and T arbitrary associative rings, there is no neat explicit description of $S * T$. Suppose, however, that as abelian groups, S is free on a basis $\{1_S\} \cup B_S$, and T is free on a basis $\{1_T\} \cup B_T$. (For example, the rings $S = \mathbb{Z}[x]$ and $T = \mathbb{Z}[i]$ have such bases, with $B_S = \{x, x^2, \dots\}$ and $B_T = \{i\}$.) Then we see that given a ring R and homomorphisms $S \rightarrow R$, $T \rightarrow R$, written $s \mapsto \bar{s}$ and $t \mapsto \tilde{t}$, the elements of R that we get by ring operations from the images of S and T can be written as linear combinations, with integer coefficients, of products $x_n \dots x_1$ where $x_i \in \overline{B_S} \cup \widetilde{B_T}$ (i.e., $\{\bar{b} \mid b \in B_S\} \cup \{\tilde{b} \mid b \in B_T\}$), and no two factors from the same basis-set occur successively. (In thinking this through, note that a product of two elements from $\overline{B_S}$ can be rewritten as a linear combination of single elements from $\overline{B_S} \cup \{\overline{1_S}\}$, and that occurrences of $\overline{1_S}$ can be eliminated because in R , $\overline{1_S} = 1_R$; and the same considerations apply to elements from $\widetilde{B_T}$. In this description we are again considering 1_R as the “empty” or “length 0” product.) And in fact, the coproduct of S and T as associative rings turns out to have precisely the set of such products as an additive basis.

Exercise 3.13:8. Verify the above assertion, using an appropriate modification of van der Waerden’s trick.

Exercise 3.13:9. (i) Study the structure of the coproduct ring $\mathbb{Z}[i] * \mathbb{Z}[i]$, where $\mathbb{Z}[i]$ denotes the Gaussian integers as in Exercise 3.13:2. In particular, try to determine its center, and whether it has any zero-divisors.

(ii) In general, if S and T are rings free as abelian groups on *two-element* bases of the forms $\{1, s\}$ and $\{1, t\}$, what can be said about the structure and center of $S * T$?

The next part shows that the above situation is exceptional.

(iii) Suppose as in (ii) that S and T each have additive bases containing 1, and neither of these bases consists of 1 alone, but now suppose that at least one of them has more than two elements. Show that in this situation, the center of $S * T$ is just \mathbb{Z} .

When the rings in question are not free \mathbb{Z} -modules, the above description does not work, but in some cases the result is nonetheless easy to characterize.

Exercise 3.13:10. (i) Describe the rings $\mathbb{Q} * \mathbb{Q}$ and $\mathbb{Q} \otimes \mathbb{Q}$.

(ii) Describe the rings $\mathbb{Z}_n * \mathbb{Q}$ and $\mathbb{Z}_n \otimes \mathbb{Q}$, where n is a positive integer.

Some surprising results on the module theory of ring coproducts are obtained in [41]. (That

paper presumes familiarity with basic properties of semisimple artin rings and their modules. The reader who is familiar with such rings and modules, but not with homological algebra, should not be deterred by the discussion of homological properties of coproducts in the first section; that section gives homological applications of the main result of the paper, but the later sections where the main result is proved do not require homological methods.)

3.14. Boolean algebras and Boolean rings. Let S be a set, and let $\mathbf{P}(S)$ denote the power set of S , that is, $\{T \mid T \subseteq S\}$. There are various natural operations on $\mathbf{P}(S)$: union, intersection, complement (i.e., ${}^c T = \{s \in S \mid s \notin T\}$), and the two zeroary operations, $\emptyset \in \mathbf{P}(S)$ and $S = {}^c \emptyset \in \mathbf{P}(S)$. Thus we can regard $\mathbf{P}(S)$ as the underlying set of an algebraic structure

$$(3.14.1) \quad (\mathbf{P}(S), \cup, \cap, {}^c, \emptyset, S).$$

This structure, or more generally, any 6-tuple consisting of a set, and five operations of arities 2, 2, 1, 0, 0 satisfying all the identities satisfied by structures of the form (3.14.1) for sets S , is called a *Boolean algebra*.

Such 6-tuples do not quite fit any of the pigeonholes we have considered so far. For instance, neither of the operations \cup, \cap is the composition operation of an abelian group, hence a ‘‘Boolean algebra’’ is not a ring.

However, there is a way of looking at $\mathbf{P}(S)$ which reduces us to ring theory. There is a standard one-to-one correspondence between the power set $\mathbf{P}(S)$ of a set S and the set of functions 2^S , where 2 means the 2-element set $\{0, 1\}$; namely, the correspondence associating to each $T \in \mathbf{P}(S)$ its characteristic function (the function whose value is 1 on elements of T and 0 on elements of ${}^c T$). If we try to do arithmetic with these functions, we run into the difficulty that the sum of two $\{0, 1\}$ -valued functions is not generally $\{0, 1\}$ -valued. But if we identify $\{0, 1\}$ with the underlying set of the ring \mathbb{Z}_2 rather than treating it as a subset of \mathbb{Z} , this problem is circumvented: 2^S becomes the ring \mathbb{Z}_2^S – the direct product of an S -tuple of copies of \mathbb{Z}_2 . Moreover, it is possible to describe union, intersection, etc., of subsets of S in terms of the ring operations of \mathbb{Z}_2^S . Namely, writing \bar{a} for the characteristic function of $a \subseteq S$, we have

$$(3.14.2) \quad \overline{a \cap b} = \bar{a}\bar{b}, \quad \overline{a \cup b} = \bar{a} + \bar{b} + \bar{a}\bar{b}, \quad \overline{{}^c a} = 1 + \bar{a}, \quad \overline{\emptyset} = 0, \quad \overline{S} = 1.$$

Conversely, each ring operation of \mathbb{Z}_2^S , translated into an operation on subsets of S , can be expressed in terms of our set-theoretic Boolean algebra operations. The expressions for multiplication, for 0, and for 1 are clear from (3.14.2); additive inverse is the identity operation, and $+$ is described by

$$(3.14.3) \quad \bar{a} + \bar{b} = \overline{(a \cap {}^c b) \cup ({}^c a \cap b)}.$$

(The set $(a \cap {}^c b) \cup ({}^c a \cap b)$ is called the ‘‘symmetric difference’’ of the sets a and b .)

Now clearly the ring $B = \mathbb{Z}_2^S = (2^S, +, \cdot, -, 0, 1)$ will, like \mathbb{Z}_2 , satisfy

$$(3.14.4) \quad (\forall x \in |B|) \quad x^2 = x,$$

from which one easily deduces the further identities,

$$(3.14.5) \quad \begin{aligned} (\forall x, y \in |B|) \quad xy &= yx, \\ (\forall x \in |B|) \quad x + x &= 0 \quad (\text{equivalently: } 1 + 1 = 0 \text{ in } B). \end{aligned}$$

An associative ring satisfying (3.14.4) (and so also (3.14.5)) is called a *Boolean ring*. We shall see below (Exercise 3.14:2) that the identities defining a Boolean ring, i.e., the identities of associative rings together with (3.14.4), imply *all* identities satisfied by rings \mathbb{Z}_2^S . Hence we shall see that Boolean rings and Boolean algebras are essentially equivalent – we can turn one into the other using (3.14.2) and (3.14.3).

Exercise 3.14:1. The *free Boolean ring* $F(X)$ on any set X exists by the usual general arguments. Find a normal form for the elements of $F(X)$ when X is finite. To prove that distinct expressions in normal form represent distinct elements, you will need some kind of representation of $F(X)$; use a representation by subsets of a set S .

Exercise 3.14:2. Assume here the result implicit in the last sentence of the preceding exercise, that the free Boolean ring on any finite set X can be embedded in the Boolean ring of subsets of some set S .

- (i) Deduce that all identities satisfied by the rings \mathbb{Z}_2^S (S a set) follow from the identities by which we defined Boolean rings.
- (ii) Conclude that the free Boolean ring on an *arbitrary* set X can be embedded in the Boolean ring of $\{0, 1\}$ -valued functions on some set (if you did not already prove this as part of your proof of (i)).
- (iii) Deduce that there exists a finite list of identities for Boolean *algebras* which implies all identities holding for such structures (i.e., all identities holding in sets $\mathbf{P}(S)$ under $\cup, \cap, \overset{c}{}, 0$ and 1).

Exercise 3.14:3. An element a of a ring (or semigroup or monoid) is called *idempotent* if $a^2 = a$. If R is a *commutative* ring, let us define

$$\text{Idpt}(R) = (\{a \in R \mid a^2 = a\}, \dot{+}, \cdot, \dot{-}, 0, 1),$$

where $a \dot{+} b = a + b - 2ab$ and $\dot{-} a = a$.

- (i) Verify that each of the above operations carries the set $|\text{Idpt}(R)|$ into itself.
- (ii) Show that if $a \in |\text{Idpt}(R)|$, then R can (up to isomorphism) be written $R_1 \times R_2$, in such a way that the element a has the form $(0, 1)$ in this direct product. Deduce that if $a_1, \dots, a_i \in |\text{Idpt}(R)|$, then R can be written as a finite direct product in such a way that each a_i has each coordinate 0 or 1 . This result can be used to get a proof of the next point that is conceptual, rather than purely computational:
- (iii) Show that for any commutative ring R , $\text{Idpt}(R)$ is a Boolean ring.
- (iv) Given any Boolean ring B , show that there is a universal pair (R, f) where R is a commutative ring, and $f: B \rightarrow \text{Idpt}(R)$ a homomorphism.
- (v) Investigate the structure of the R of the above construction in some simple cases, e.g., $B = \mathbb{Z}_2$, $B = \mathbb{Z}_2^2$, $B = \mathbb{Z}_2^X$.

(Students familiar with algebraic geometry will recognize that the idempotent elements of a commutative ring R correspond to the continuous $\{0, 1\}$ -valued functions on $\text{Spec}(R)$. Thus the Boolean rings $\text{Idpt}(R)$ of the above exercise have natural representations as Boolean rings of $\{0, 1\}$ -valued functions on sets.)

Exercise 3.14:4. (i) If $f: U \rightarrow V$ is a set map, what sort of homomorphism does it induce between the Boolean rings \mathbb{Z}_2^U and \mathbb{Z}_2^V ?

(ii) Let B be a Boolean ring. Formulate universal properties for a ‘‘universal representation of B by subsets of a set’’, in each of the following senses:

- (a) A universal pair (S, f) , where S is a set, and f a Boolean ring homomorphism $B \rightarrow \mathbb{Z}_2^S$.
- (b) A universal pair (T, g) , where T is a set, and g a Boolean ring homomorphism

$$\mathbb{Z}_2^T \rightarrow B.$$

(iii) Investigate whether such universal representations exist. If such representations are obtained, investigate whether the maps f, g will in general be one-to-one, or onto.

Exercise 3.14:5. (i) Show that every finite Boolean ring is isomorphic to one of the form 2^S for some finite set S .

(ii) For what finite sets S is the Boolean ring 2^S free? How is the number of free generators determined by the set S ?

Exercise 3.14:6. A subset T of a set S is said to be *cofinite* in S if cT (taken relative to S , i.e., $S - T$) is finite. Show that $\{T \subseteq \mathbb{Z} \mid T \text{ is finite or cofinite}\}$ is the underlying set of a Boolean subring of $2^{\mathbb{Z}}$, which is neither free, nor isomorphic to a Boolean ring 2^U for any set U .

Above, I have for purposes of exposition distinguished between the power set $\mathbf{P}(S)$ of a set S and the function-set 2^S . But these notations are often used interchangeably, and I may use them that way myself elsewhere in these notes.

3.15. Sets. The objects we have been studying have been sets with additional operations. Let us briefly note the forms that some of the constructions we have discovered take for plain sets.

Given a family of sets $(S_i)_{i \in I}$, the object with the universal property characterizing products is the usual direct product, $\prod_I S_i$, which may be described as the set of functions on I whose value at each element i belongs to the set S_i . The projection map p_i in the statement of the universal property takes each such function to its value at i . Note that the product of the vacuous family of sets (indexed by the empty set!) is a one-element set.

The coproduct of a family of sets $(S_i)_{i \in I}$ is their *disjoint union* $\sqcup_I S_i$, to which we referred in passing in §3.6. If the S_i are themselves disjoint, one can take for this set their ordinary union; the inclusions of the S_i in this union give the universal family of maps $q_j: S_j \rightarrow \sqcup_I S_i$ ($j \in I$). A construction that will work without any disjointness assumption is to take

$$(3.15.1) \quad \sqcup_I S_i = \{(i, s) \mid i \in I, s \in S_i\}$$

with universal maps given by

$$(3.15.2) \quad q_i(s) = (i, s) \quad (i \in I, s \in S_i).$$

A frequent practice in mathematical writing is to assume (“without loss of generality”) that a family of sets is disjoint, if this would be notationally convenient, and if there is nothing logically forcing them to have elements in common. When this disjointness condition holds one can, as noted, take the universal maps involved in the definition of a coproduct of sets to be inclusions. But in other cases, for instance if we want to consider a coproduct of a set with itself, or of a set and a subset, a construction like (3.15.1) is needed. Note that when a construction is described “in general” under such a disjointness assumption, and is later applied in a situation where one cannot make that assumption, one must be careful to insert q_i 's where appropriate.

Exercise 3.15:1. Investigate laws such as associativity, distributivity, etc. which are satisfied up to natural isomorphism by the constructions of pairwise product and coproduct of sets.

Examine which of these laws are also satisfied by products and coproducts of groups, and which are not.

Sets can also be constructed by “generators and relations”. If X is a set, then relations are specified by a set R of ordered pairs of elements of X , which we want to make fall together.

The universal image of X under a map making the components of each of these pairs fall together is easily seen to be the quotient of X by the least equivalence relation containing R .

The constructions examined in this section – direct product of sets, disjoint union, and quotient by the equivalence relation generated by a given binary relation – were, of course, already used in earlier sections. So the point of this section is not to introduce the reader to those constructions, but to show their relation to the general patterns we have been noticing.

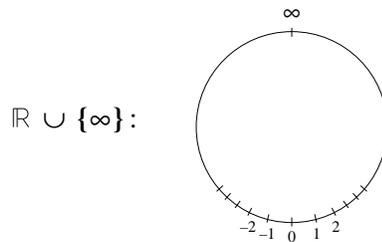
3.16. Some algebraic structures we have not looked at. ... lattices ([3]; cf. Chapter 5 below), modular lattices, distributive lattices; partially ordered sets (Chapter 4 below); cylindric algebras [74]; heaps (cf. Exercise 8.6:9 below); loops [6, p.52]; Lie algebras ([80], cf. §8.7 below), Jordan algebras [81], general nonassociative algebras; rings with polynomial identity [110], rings with involution, fields, division rings, Hopf algebras [117]; modules, bimodules (§§9.8-9.9 below); filtered groups, filtered rings, filtered modules; graded rings, graded modules; ordered groups, lattice-ordered groups [70], ...

We'll look at some of these in later chapters.

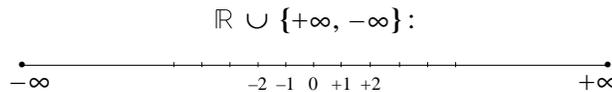
On the objects we *have* considered here, we have only looked at basic and familiar universal constructions. Once we develop a general theory of universal constructions, we shall see that they come in many more varied forms.

For diversity, I will end this chapter with two examples for those who know some general topology.

3.17. The Stone-Čech compactification of a topological space. As is well known, the real line \mathbb{R} is not compact. It is frequently convenient, when studying the limit-behavior of \mathbb{R} -valued functions or sequences, to adjoin to \mathbb{R} an additional point, “ ∞ ”. The resulting compact space, $\mathbb{R} \cup \{\infty\}$, is shown below.

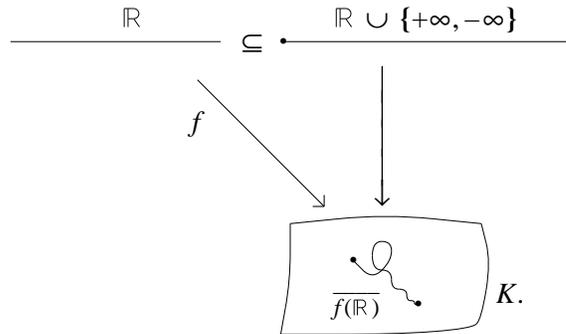


At other times, one adjoins to \mathbb{R} two points, $+\infty$ and $-\infty$, getting a space

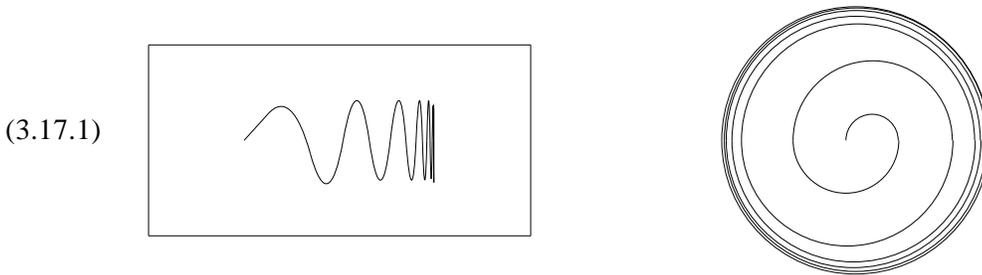


Note that $\mathbb{R} \cup \{\infty\}$ may be obtained from $\mathbb{R} \cup \{+\infty, -\infty\}$ by an *identification*. Hence $\mathbb{R} \cup \{+\infty, -\infty\}$ can be thought of as making “finer distinctions” in limiting behavior than $\mathbb{R} \cup \{\infty\}$.

One might imagine that $\mathbb{R} \cup \{+\infty, -\infty\}$ makes “the finest possible distinctions”. A precise formulation of this would be a conjecture that for any continuous map f of \mathbb{R} into a compact Hausdorff space K , the closure of the image of \mathbb{R} should be an image of $\mathbb{R} \cup \{+\infty, -\infty\}$; i.e., that the map f should factor through the inclusion $\mathbb{R} \subseteq \mathbb{R} \cup \{+\infty, -\infty\}$. Here is a picture of an example where this is true:

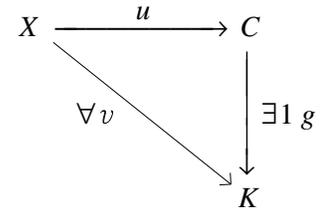


But by thinking about either of the following pictures, you can see that the above conjecture is not true in general:



But we can still ask whether there is *some* compactification of \mathbb{R} which makes “the most possible distinctions”. Let us raise the same question with \mathbb{R} replaced by a general topological space X , and give the desired object a name.

Definition 3.17.2. Let X be a topological space. A Stone-Čech compactification of X will mean a pair (C, u) , where C is a compact Hausdorff space and u a continuous map $X \rightarrow C$, universal among all continuous maps of X into compact Hausdorff spaces K (diagram at right).



Exercise 3.17:1. Show that if a pair (C, u) as in the above definition exists, then $u(X)$ is dense in C . In fact, show that if (C, u) has the indicated universal property but without the condition of uniqueness of factoring maps g (see above diagram), then

- (i) uniqueness of such maps holds if and only if $u(X)$ is dense in C ; and
- (ii) if C' is the closure of $u(X)$ in C , the pair (C', u) has the full universal property.

We want to know whether such compactifications always exist.

The analog of our construction of free groups from terms as in §2.2 would be to adjoin to X some kinds of “formal limit points”. But limit points of what? Not every sequence in a compact Hausdorff space K converges, nor need every point of the closure of a subset $J \subseteq K$ be the limit of a sequence of points of J (unless K is first countable); so adjoining limits of *sequences* would not do. The approach of adjoining limit points *can* in fact be made to work, but it requires considerable study of how such points may be described; the end result is a construction of the Stone-Čech compactification of X in terms of *ultrafilters*. We shall not pursue that approach here; it is used in [119, Theorem 17.17 et seq.]. (NB: The compactification constructed there may not be

Hausdorff when X is “bad”, so in such cases it will not satisfy our definition.)

The “big direct product” approach is more easily adapted. If $v_1: X \rightarrow K_1$ and $v_2: X \rightarrow K_2$ are two continuous maps of X into compact Hausdorff spaces, then the induced map $(v_1, v_2): X \rightarrow K_1 \times K_2$ will “make all the distinctions among limit points made by either v_1 or v_2 ”, since the maps v_1 and v_2 can each be factored through it; further, if we let K' denote the closure of the image of X in $K_1 \times K_2$ under that map, and $v': X \rightarrow K'$ the induced map, then all these distinctions are still made in K' , and the image of X is dense in this space. We can do the same with an arbitrary family of maps $v_i: X \rightarrow K_i$ ($i \in I$), since Tychonoff’s Theorem tells us that the product space $\prod_I K_i$ is again compact.

As in the construction of free groups, to obtain our Stone-Čech compactification by this approach we have to find some set of pairs (K_i, v_i) which are “as good as” the class of all maps v of X into all compact Hausdorff spaces K . For this purpose, we want a bound on the cardinalities of the closures of all images of X under maps into compact Hausdorff spaces K . To get this, we would like to say that every point of the closure of the image of X somehow “depends” on the images of elements of X , in such a fashion that different points “depend” on these differently; and then bound the number of kinds of “dependence” there can be, in terms of the cardinality of X . The next lemma establishes the “different points depend on X in different ways” idea, and the corollary that follows gives the desired bound.

Lemma 3.17.3. *Let K be a Hausdorff topological space, and for any $k \in |K|$, let $N(k)$ denote the set of all open neighborhoods of k (open sets in K containing k). Then for any map v from a set X into K , and any two points $k_1 \neq k_2$ of the closure of $v(X)$ in K , one has $v^{-1}(N(k_1)) \neq v^{-1}(N(k_2))$ (where by $v^{-1}(N(k))$ we mean $\{v^{-1}(U) \mid U \in N(k)\}$, a subset of $\mathbf{P}(X)$).*

Proof. Since k_i ($i = 1, 2$) is in the closure of $v(X)$, every neighborhood of k_i in K has nonempty intersection with $v(X)$, i.e., every member of $v^{-1}(N(k_i))$ is nonempty. Since $N(k_i)$ is closed under pairwise intersections, so is $v^{-1}(N(k_i))$. But since K is Hausdorff and $k_1 \neq k_2$, these two points possess disjoint neighborhoods, whose inverse images in X will have empty intersection. If the sets $v^{-1}(N(k_1))$ and $v^{-1}(N(k_2))$ were the same, this would give a contradiction to the above nonemptiness observation. \square

Thus, we can associate to distinct points of the closure of $v(X)$ distinct sets of subsets of X . Hence,

Corollary 3.17.4. *In the situation of the above lemma, the cardinality of the closure of $v(X)$ in K is $\leq 2^{2^{\text{card } X}}$. \square*

So now, given any topological space X , let us choose a set S of cardinality $2^{2^{\text{card } |X|}}$, and let A denote the set of all pairs $a = (K_a, u_a)$ such that K_a is a compact Hausdorff topological space with underlying set $|K_a| \subseteq S$, and u_a is a continuous map $X \rightarrow K_a$. (We no longer need to keep track of cardinalities, but if we want to, $\text{card } A \leq 2^{2^{2^{\text{card } |X|}}}$, assuming X infinite. The two additional exponentials come in when we estimate the number of topologies on a set of $\leq 2^{2^{\text{card } |X|}}$ elements.) Thus, if v is any continuous map of X into a compact Hausdorff space K , and we write K' for the closure of $v(X)$ in K , then the pair (K', v) will be “isomorphic” to some pair $(K_a, u_a) \in A$, in the sense that there exists a homeomorphism between K' and K_a

making the diagram below commute.

$$\begin{array}{ccc} & & K' \subseteq K \\ & \nearrow v & \\ X & & \\ & \searrow u_a & \\ & & K_a \end{array} \quad \begin{array}{c} \\ \\ \parallel \\ \parallel \\ \} \text{ homeomorphism} \end{array}$$

We now form the compact Hausdorff space $P = \prod_{a \in A} K_a$, and the map $u: X \rightarrow P$ induced by the u_a 's, and let $C \subseteq P$ be the closure of $u(X)$. It is easy to show, as we did for groups in §2.3, that the pair (C, u) satisfies the universal property of Definition 3.17.2. Thus:

Theorem 3.17.5. *Every topological space X has a Stone-Čech compactification (C, u) in the sense of Definition 3.17.2. \square*

Exercise 3.17:2. Show that in the above construction, $u(X)$ will be homeomorphic to X under u if and only if X can be *embedded* in a compact Hausdorff space K (where an “embedding” means a continuous map $f: X \rightarrow K$ inducing a homeomorphism between X and $f(X)$, the latter set being given the topology induced by that of K). Examine conditions on X under which these equivalent statements will hold. Show that for any topological space X , there exists a universal map into a space Y embeddable in a compact Hausdorff space, and that this map is always onto, but that it may not be one-to-one. Can it be one-to-one and onto but not a homeomorphism?

Note: Most authors use the term “compactification” to mean a dense *embedding* in a compact space. Hence, they only consider a space X to have a Stone-Čech compactification if the map u that we have constructed *is* an embedding.

Exercise 3.17:3. Suppose we leave off the condition “Hausdorff” – does a space X always have a universal map into a *compact* space C ? A compact T_1 space C ? ...

Exercise 3.17:4. Let C be the Stone-Čech compactification of the real line \mathbb{R} , and regard \mathbb{R} as a subspace of C .

- (i) Show that $C - \mathbb{R}$ has exactly two connected components.
(The above shows that there was a grain of truth in the naive idea that $\mathbb{R} \cup \{+\infty, -\infty\}$ was the universal compactification of \mathbb{R} . Exercise 3.17:5 will also be relevant to that idea.)
- (ii) What can you say about path-connected components of $C - \mathbb{R}$?
- (iii) Show that no *sequence* in \mathbb{R} converges to a point of $C - \mathbb{R}$.

A continuous map of \mathbb{R} into a topological space K may be thought of as an open *curve* in K . If K is a *metric space* one can define the *length* (possibly infinite) of this curve.

Exercise 3.17:5. Show that if $v: \mathbb{R} \rightarrow K$ is a curve of *finite length* in a compact (or more generally, a complete) metric space K , then v factors through the inclusion of \mathbb{R} in $\mathbb{R} \cup \{+\infty, -\infty\}$.

Is the converse true? I.e., must every map $\mathbb{R} \rightarrow K$ which factors through the inclusion of \mathbb{R} in $\mathbb{R} \cup \{+\infty, -\infty\}$ have finite length?

Exercise 3.17:6. (Exploring possible variants of Exercises 3.17:4-5.) It would be nice to get a result like the first assertion of the preceding exercise with a purely topological hypothesis on the map v , rather than a condition involving a metric on K . Consider, for instance, the following condition on a map v of the real line into a compact Hausdorff space K :

- (3.17.6) For every closed set $V \subseteq K$, and open set $U \supseteq V$, the set $v^{-1}(U) \subseteq \mathbb{R}$ has only finitely many connected components that contain points of $v^{-1}(V)$.

(You should convince yourself that this fails for the two cases shown in (3.17.1).)

- (i) Can we replace the assumptions in Exercise 3.17:5 that K is a metric space and v has finite length by (3.17.6) or some similar condition?
- (ii) Let X be the open unit disc, C the closed unit disc, and $u: X \rightarrow C$ the inclusion map. Does the pair (C, u) have any universal property with respect to X , like that indicated for $\mathbb{R} \cup \{+\infty, -\infty\}$ with respect to \mathbb{R} in the preceding exercise?
- (iii) Does the open disc have a universal path-connected compactification?
- (iv) In general, if C is the Stone-Čech compactification of a “nice” space X , what can be said about connected components, path components, homotopy, cohomotopy, etc. of $C - X$?

Having seen that the Stone-Čech compactification of the topological space \mathbb{R} is enormous, one may wonder whether any noncompact Hausdorff space can have a Stone-Čech compactification that is more modest. Can it add only one point to the space, for instance? The next exercise finds conditions for this to happen. We shall see in Exercise 4.5:16 how to get a space X that satisfies these conditions.

Exercise 3.17:7. Let X be a noncompact topological space which can be embedded in a compact Hausdorff space. Show that the following conditions are equivalent.

- (a) The Stone-Čech compactification of X has the form $u(X) \cup \{y\}$, where u is the universal map of X into that compactification, and y is a single point not in $u(X)$.
- (b) Of any two disjoint closed subsets $F, G \subseteq X$, at least one is compact.
- (c) Every continuous function $X \rightarrow [0, 1]$ is constant on the complement of some compact subset of X .

One can also consider universal constructions which mix topological and algebraic structure:

Exercise 3.17:8. Let G be any *topological group* (a group given with a Hausdorff topology on its underlying set, such that the group operations are continuous). Show that there exists a universal pair (C, h) , where C is a *compact* topological group, and $h: G \rightarrow C$ a continuous group homomorphism. This is called the *Bohr compactification* of G .

Show that $h(G)$ is dense in C . Is h generally one-to-one? A topological embedding? What will be the relation between C and the Stone-Čech compactification of the underlying topological space of G ?

If it helps, you might consider some of these questions in the particular case where G is the additive group of the real line.

In §2.4 we saw that we could improve on the construction of the free group on X from “terms” by noting that a certain subset of the terms would make do for all of them. For the Stone-Čech compactification, the “big direct product” construction is subject to a similar simplification. In that construction, we made use of all maps (up to homeomorphism) of X into compact Hausdorff spaces of reasonable size. I claim that we can in fact make all the “distinctions” we need using maps into the closed unit interval, $[0, 1]$! The key fact is that any two points of a compact Hausdorff space K can be separated by a continuous map into $[0, 1]$ (Urysohn’s Lemma). I will sketch how this is used.

Let X be any topological space, let W denote the set of continuous maps $w: X \rightarrow [0, 1]$, let $u: X \rightarrow [0, 1]^W$ be the map induced by $(w)_{w \in W}$, and let $C \subseteq [0, 1]^W$ be the closure of $u(X)$. It is immediate that C has the property

- (3.17.7) Every continuous function of X into $[0, 1]$ is the composite of u with a unique continuous function $C \rightarrow [0, 1]$ (namely, the restriction to C of one of the projections $[0, 1]^W \rightarrow [0, 1]$).

To show that C has the universal property of the Stone-Ćech compactification of X , let K be a compact Hausdorff space. We can separate points of K by some set S of continuous maps $s: K \rightarrow [0, 1]$, hence we can embed K in a ‘‘cube’’ $[0, 1]^S$. (The map $K \rightarrow [0, 1]^S$ given by our separating family of functions is one-to-one; hence as K is Hausdorff it will be a topological embedding [84, Theorem 5.8, p.141].) Let us therefore assume, without loss of generality, that K is a subspace of $[0, 1]^S$. Now given any map $v: X \rightarrow K$, we regard it as a map into the overspace $[0, 1]^S$, and get a factorization $v = gu$ for a unique map $g: C \rightarrow [0, 1]^S$ by applying (3.17.7) to each coordinate. Because K is compact, it is closed in $[0, 1]^S$, so g will take C , the closure of $u(X)$, into K , establishing the universal property of C . Cf. [84, pp. 152-153].

Another twist: Following the idea of Exercise 2.3:6, we may regard a point c of the Stone-Ćech compactification of a space X as determining a function \tilde{c} which associates to every continuous map v of X into a compact Hausdorff space K a point $\tilde{c}(v) \in K$ – namely, the image of c under the unique extension of v to C . This map \tilde{c} will be ‘‘functorial’’, i.e., will respect continuous maps $f: K_1 \rightarrow K_2$, in the sense indicated in the diagram below.

$$\begin{array}{ccc} & v & \longrightarrow K_1 \ni \tilde{c}(v) \\ X & \searrow & \downarrow f \\ & fv & \longrightarrow K_2 \ni \tilde{c}(fv) \end{array}$$

From Urysohn’s Lemma one can deduce that \tilde{c} is determined by its behavior on maps $w: X \rightarrow [0, 1]$, hence, more generally, by its behavior on maps w of X into closed intervals $[a, b] \subseteq \mathbb{R}$. We carry this observation further in

Exercise 3.17:9. A *bounded* real-valued continuous function on X can be regarded as a continuous map from X into a compact subset of \mathbb{R} , and our \tilde{c} can be applied to this map.

- (i) Show that in this way one may obtain from \tilde{c} a function from the set $B(X)$ of all bounded real-valued continuous functions on X to the real numbers \mathbb{R} . (To prove this function well-defined, i.e., that the result of applying \tilde{c} to a bounded function is independent of our choice of compact subset of \mathbb{R} containing the range of this function, use the functoriality property of \tilde{c} .)
- (ii) Show that this map is a ring homomorphism $B(X) \rightarrow \mathbb{R}$ (with respect to the obvious ring structure on $B(X)$).

One can show, further, that every ring homomorphism $B(X) \rightarrow \mathbb{R}$ is continuous, and deduce that each such homomorphism is induced by a point of C . So one gets another description of the Stone-Ćech compactification C of X , as the space of homomorphisms into \mathbb{R} of the ring $B(X)$ of bounded continuous real-valued functions on X . The topology of C is the function topology on maps of $B(X)$ into \mathbb{R} .

Perhaps I have made this approach sound too esoteric. A simpler way of putting it is to note that every bounded continuous real function on X (i.e., every continuous function which has range in a compact subset of \mathbb{R}) extends uniquely to a bounded continuous real function on its Stone-Ćech compactification C , so $B(X) \cong B(C)$; and then to recall that for any compact Hausdorff space C , the homomorphisms from the function-ring $B(C)$ into \mathbb{R} are just the evaluation functions at points of C .

One can use this approach to get another proof of the existence of the Stone-Ćech compactification of a topological space [68, Chapter 6]. This homomorphism space can also be identified with the space of all *maximal ideals* of $B(X)$, equivalently, of all *prime ideals* that are closed in the topology given by the sup norm.

Exercise 3.17:10. Suppose B' is any \mathbb{R} -subalgebra of $B(X)$. Let C' denote the set of all maximal ideals of B' . Show that there is a natural map $m: C \rightarrow C'$. Show by examples that this map can fail to be one-to-one (even if B' separates points of X), or to be onto. Try to find conditions for it to be one or the other.

In [93, §41], the Bohr compactification (Exercise 3.17:8 above) of a topological group G is obtained similarly as the maximal ideal space of a subring of $B(G)$, the subring of “almost periodic” functions.

Most often, complex- rather than real-valued functions are used in these constructions.

3.18. Universal covering spaces. Let X be a pathwise connected topological space with a basepoint (distinguished point) x_0 . (Formally, this would be defined as a 3-tuple $(|X|, T, x_0)$, where $|X|$ is a set, T is a pathwise connected topology on $|X|$, and x_0 is an element of $|X|$.)

A *covering space* of X means a pair (Y, c) , where Y is a pathwise connected space with a basepoint y_0 , and c is a continuous basepoint-preserving map $Y \rightarrow X$, such that every $x \in X$ has a neighborhood V such that $c^{-1}(V)$ is homeomorphic, as a space mapped to V , to a direct product of V with a discrete space. (Draw a picture!) Such a c will have the *unique path-lifting property*: Given any continuous map $p: [0, 1] \rightarrow X$ taking 0 to x_0 , there will exist a unique continuous map $\tilde{p}: [0, 1] \rightarrow Y$ taking 0 to y_0 such that $p = c\tilde{p}$; and further, \tilde{p} depends continuously on p in the appropriate function-space topology.

Given X , consider any covering space (Y, c) of X , and let us ask what points of Y we can “describe” in a well-defined manner.

Of course, we have the basepoint, y_0 . Further, for any path p in X starting at the basepoint x_0 , we know there will be a unique lifting of p to a path \tilde{p} in Y starting from y_0 ; so Y also has all points of this lifted path. It is enough, however, to note that we have the endpoint $\tilde{p}(1)$ of each such lifted path, since all the other points of \tilde{p} can be described as endpoints of liftings of “subpaths” of p . In fact, every $y \in Y$ will be the endpoint $\tilde{p}(1)$ of a lifted path in X . For Y was assumed pathwise connected, hence for any $y \in Y$ we can find a path q in Y with $q(0) = y_0$, $q(1) = y$. Letting $p = cq$, a path in X , we see that $q = \tilde{p}$, so $y = \tilde{p}(1)$.

Suppose p and p' are two paths in X ; when will $\tilde{p}(1)$ and $\tilde{p}'(1)$ be the same point of Y ? Clearly, a necessary condition is that these two points have the same image x in X : $p(1) = p'(1) = x$. Assuming this condition, note that if p and p' are homotopic in the class of paths in X from x_0 to x , then as one continuously deforms p to p' in this class, the lifted path in Y will vary continuously, hence its endpoint in $c^{-1}(x)$ will vary continuously. But $c^{-1}(x)$ is discrete, so the endpoint must remain constant. Thus, p 's being homotopic to p' in the class of paths with these specified endpoints implies $\tilde{p}(1) = \tilde{p}'(1)$.

So in general, we get a point of Y for every homotopy class $[p]$ of paths in X with initial point x_0 and common final point. In a particular covering space Y , there may or may not be further equalities among these points of Y ; but we can ask whether, if we write U for the set of such homotopy classes of paths, and u for the map from U to X defined by $u([p]) = p(1)$, we can make U a topological space in such a way that the pair (U, u) is a covering space for X . Under appropriate assumptions on the topology of X (the hypotheses used in [77] are that X is connected, locally pathwise connected, and semi-locally simply connected), this can indeed be done. The resulting covering space U has a unique continuous map onto each covering space Y of X , which respects basepoints and respects the maps into X . Hence (U, u) is called the *universal covering space* of X .

The universal covering space is a versatile animal – like the direct product of groups, it has, in

addition to the above left universal property, a right universal one:

It is not hard to show that U is simply connected. Consider, now, pairs (S, c) , where S is a simply connected pathwise connected topological space with basepoint s_0 , and $c: S \rightarrow X$ a basepoint-respecting continuous map. Let us ask, for such a space S , the question that we noted in §3.8 as leading to *right universal* constructions: If s is an arbitrary point of S , what data will it determine that can be formulated in terms of the given space X ? Well, obviously s determines the point $c(s) \in X$. To get more information, note that since S is pathwise connected, there will be some path q in S connecting s_0 to s ; and since S is *simply* connected, all such paths q are homotopic. Applying c to these paths, we see that s determines a *homotopy class* of paths in X from x_0 to $c(s)$. But as we have just noted, the set of homotopy classes of paths from x_0 to points of X can (under appropriate conditions) itself be made into a simply connected space, the universal covering space of X . One deduces that this space U is right universal among simply connected spaces with basepoint, given with maps into X (diagram below).

$$\begin{array}{ccc}
 U & \xrightarrow{u} & X \\
 \uparrow \exists! d & & \nearrow \forall c \\
 S & &
 \end{array}$$

More generally, one can show that the universal covering space of the pathwise connected space X (when it exists) is right universal among pathwise connected spaces S with basepoint, given with basepoint-preserving maps c into X such that the group homomorphism $\pi_1(c): \pi_1(S) \rightarrow \pi_1(X)$ is trivial. It is easy to give examples showing that such a space S need not itself be simply connected.

We could also look for a *right* universal covering space for X , or a simply connected space with basepoint having a *left* universal map into X . But these turn out to be uninteresting: They are X itself, and the one-point space.

There are many other occurrences of universal constructions in topology. Some, like the constructions considered in this and the preceding section, can be approached in the same way as universal constructions in algebra. Others, used in algebraic topology, are different in that one is interested, not in maps being equal, unique, etc., but *homotopic*, unique *up to homotopy*, etc.. These conditions can be brought into the same framework as our other universal properties via the formalism of *category theory* (Chapters 6 and 7 below), but the tasks of constructing and studying the objects these conditions characterize require different approaches, which we will not treat in this course.