Cambridge studies in advanced mathematics 143

Basic Category Theory

TOM LEINSTER

CAMBRIDGE

Basic Category Theory

TOM LEINSTER
University of Edinburgh

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107044241

First published 2014

Printed in the United Kingdom by CPI Group Ltd, Croydon CRO 4YY

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-04424-1 Hardback

Contents

Note to the reader

This is not a sophisticated text. In writing it, I have assumed no more mathematical knowledge than might be acquired from an undergraduate degree at an ordinary British university, and I have not assumed that you are used to learning mathematics by reading a book rather than attending lectures. Furthermore, the list of topics covered is deliberately short, omitting all but the most fundamental parts of category theory. A 'further reading' section points to suitable follow-on texts.

There are two things that every reader should know about this book. One concerns the examples, and the other is about the exercises.

Each new concept is illustrated with a generous supply of examples, but it is not necessary to understand them all. In courses I have taught based on earlier versions of this text, probably no student has had the background to understand every example. All that matters is to understand enough examples that you can connect the new concepts with mathematics that you already know.

As for the exercises, I join every other textbook author in exhorting you to do them; but there is a further important point. In subjects such as number theory and combinatorics, some questions are simple to state but extremely hard to answer. Basic category theory is not like that. To understand the question is very nearly to know the answer. In most of the exercises, there is only one possible way to proceed. So, if you are stuck on an exercise, a likely remedy is to go back through each term in the question and make sure that you understand it *in full*. Take your time. Understanding, rather than problem solving, is the main challenge of learning category theory.

Citations such as Mac Lane (1971) refer to the sources listed in 'Further reading'.

This book developed out of master's-level courses taught several times at the University of Glasgow and, before that, at the University of Cambridge. In turn, the Cambridge version was based on Part III courses taught for many years by Martin Hyland and Peter Johnstone. Although this text is significantly different from any of their courses, I am conscious that certain exercises, lines of development and even turns of phrase have persisted through that long evolution. I would like to record my indebtedness to them, as well as my thanks to François Petit, my past students, the anonymous reviewers, and the staff of Cambridge University Press.

Introduction

Category theory takes a bird's eye view of mathematics. From high in the sky, details become invisible, but we can spot patterns that were impossible to detect from ground level. How is the lowest common multiple of two numbers like the direct sum of two vector spaces? What do discrete topological spaces, free groups, and fields of fractions have in common? We will discover answers to these and many similar questions, seeing patterns in mathematics that you may never have seen before.

The most important concept in this book is that of *universal property*. The further you go in mathematics, especially pure mathematics, the more universal properties you will meet. We will spend most of our time studying different manifestations of this concept.

Like all branches of mathematics, category theory has its own special vocabulary, which we will meet as we go along. But since the idea of universal property is so important, I will use this introduction to explain it with no jargon at all, by means of examples.

Our first example of a universal property is very simple.

Example 0.1 Let 1 denote a set with one element. (It does not matter what this element is called.) Then 1 has the following property:

for all sets X, there exists a unique map from X to 1.

(In this context, the words 'map', 'mapping' and 'function' all mean the same thing.)

Indeed, let X be a set. There *exists* a map $X \to 1$, because we can define $f\colon X \to 1$ by taking $f(x)$ to be the single element of 1 for each $x \in X$. This is the *unique* map $X \to 1$, because there is no choice in the matter: any map $X \to 1$ must send each element of X to the single element of 1.

Phrases of the form 'there exists a unique such-and-such satisfying some condition' are common in category theory. The phrase means that there is one and only one such-and-such satisfying the condition. To prove the existence part, we have to show that there is at least one. To prove the uniqueness part, we have to show that there is at most one; in other words, any two such-and-suches satisfying the condition are equal.

Properties such as this are called 'universal' because they state how the object being described (in this case, the set 1) relates to the entire universe in which it lives (in this case, the universe of sets). The property begins with the words '*for all* sets X', and therefore says something about the relationship between 1 and *every* set X: namely, that there is a unique map from X to 1.

Example 0.2 This example involves rings, which in this book are always taken to have a multiplicative identity, called 1. Similarly, homomorphisms of rings are understood to preserve multiplicative identities.

The ring $\mathbb{Z}$ has the following property: for all rings R, there exists a unique homomorphism $\mathbb{Z} \to R$.

To prove existence, let R be a ring. Define a function $\phi\colon \mathbb{Z} \to R$ by

$$\phi(n) = \begin{cases} \underbrace{1 + \cdots + 1}_{n} & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ -\phi(-n) & \text{if } n < 0 \end{cases}$$

($n \in \mathbb{Z}$). A series of elementary checks confirms that ϕ is a homomorphism.

To prove uniqueness, let R be a ring and let $\psi\colon \mathbb{Z} \to R$ be a homomorphism. We show that ψ is equal to the homomorphism ϕ just defined. Since homomorphisms preserve multiplicative identities, $\psi(1) = 1$. Since homomorphisms preserve addition,

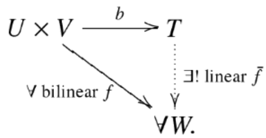
$$\psi(n) = \psi(\underbrace{1 + \cdots + 1}_{n}) = \underbrace{\psi(1) + \cdots + \psi(1)}_{n} = \underbrace{1 + \cdots + 1}_{n} = \phi(n)$$

for all $n > 0$. Since homomorphisms preserve zero, $\psi(0) = 0 = \phi(0)$. Finally, since homomorphisms preserve negatives, $\psi(n) = -\psi(-n) = -\phi(-n) = \phi(n)$ whenever $n < 0$.

Crucially, there can be essentially only *one* object satisfying a given universal property. The word 'essentially' means that two objects satisfying the same universal property need not literally be equal, but they are always isomorphic. For example:

Lemma 0.3 *Let A be a ring with the following property: for all rings R, there exists a unique homomorphism $A \to R$. Then $A \cong \mathbb{Z}$.*

map out of $U \times V$'. In other words, there exist a certain vector space T and a certain bilinear map $b\colon U \times V \to T$ with the following universal property:



(0.1)

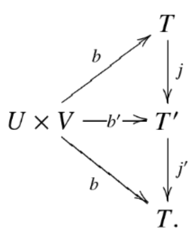
Roughly speaking, this property says that bilinear maps out of $U \times V$ correspond one-to-one with linear maps out of T.

Even without knowing that such a T and b exist, we can immediately prove that this universal property determines T and b uniquely up to isomorphism. The proof is essentially the same as that of Lemma 0.3, but looks more complicated because of the more complicated universal property.

Lemma 0.7 *Let U and V be vector spaces. Suppose that $b\colon U \times V \to T$ and $b'\colon U \times V \to T'$ are both universal bilinear maps out of $U \times V$. Then $T \cong T'$. More precisely, there exists a unique isomorphism $j\colon T \to T'$ such that $j \circ b = b'$.*

In the proof that follows, it does not actually matter what 'bilinear', 'linear' or even 'vector space' mean. The hard part is getting the logic straight. That done, you should be able to see that there is really only one possible proof. For instance, to use the universality of b, we will have to choose some bilinear map f out of $U \times V$. There are only two in sight, b and b', and we use each in the appropriate place.

Proof In diagram (0.1), take $\left(U \times V \xrightarrow{f} W\right)$ to be $\left(U \times V \xrightarrow{b'} T'\right)$. This gives a linear map $j\colon T \to T'$ satisfying $j \circ b = b'$. Similarly, using the universality of b', we obtain a linear map $j'\colon T' \to T$ satisfying $j' \circ b' = b$:



Now $j' \circ j\colon T \to T$ is a linear map satisfying $(j' \circ j) \circ b = b$; but also, the identity map $1_T\colon T \to T$ is linear and satisfies $1_T \circ b = b$. So by the uniqueness part of the universal property of b, we have $j' \circ j = 1_T$. (Here we took the 'f' of (0.1) to be b.) Similarly, $j \circ j' = 1_{T'}$. So j is an isomorphism. □

In Example 0.6, it was stated that given vector spaces U and V, there exists a pair (T, b) with the universal property of (0.1). We just proved that there is essentially only one such pair (T, b). The vector space T is called the **tensor product** of U and V, and is written as $U \otimes V$. Tensor products are very important in algebra. They reduce the study of bilinear maps to the study of linear maps, since a bilinear map out of $U \times V$ is really the same thing as a linear map out of $U \otimes V$.

However, tensor products will not be important in this book. The real lesson for us is that it is safe to speak of *the* tensor product, not just *a* tensor product, and the reason for that is Lemma 0.7. This is a general point that applies to anything satisfying a universal property.

Once you know a universal property of an object, it often does no harm to forget how it was constructed. For instance, if you look through a pile of algebra books, you will find several different ways of constructing the tensor product of two vector spaces. But once you have proved that the tensor product satisfies the universal property, you can forget the construction. The universal property tells you all you need to know, because it determines the object uniquely up to isomorphism.

Example 0.8 Let $\theta\colon G \to H$ be a homomorphism of groups. Associated with θ is a diagram

$$\ker(\theta) \overset{\iota}{\hookrightarrow} G \underset{\varepsilon}{\overset{\theta}{\rightrightarrows}} H, \tag{0.2}$$

where ι is the inclusion of $\ker(\theta)$ into G and ε is the trivial homomorphism. 'Inclusion' means that $\iota(x) = x$ for all $x \in \ker(\theta)$, and 'trivial' means that $\varepsilon(g) = 1$ for all $g \in G$. The symbol $\hookrightarrow$ is often used for inclusions; it is a combination of a subset symbol $\subset$ and an arrow.

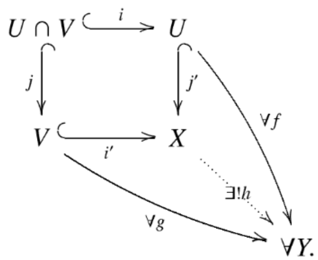
The map ι into G satisfies $\theta \circ \iota = \varepsilon \circ \iota$, and is universal as such. Exercise 0.11 asks you to make this precise.

Here is our final example of a universal property.

Example 0.9 Take a topological space covered by two open subsets: $X = U \cup V$. The diagram

$$\begin{array}{ccc} U \cap V & \overset{i}{\hookrightarrow} & U \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle j'} \\ V & \underset{i'}{\hookrightarrow} & X \end{array}$$

of inclusion maps has a universal property in the world of topological spaces and continuous maps, as follows:



(0.3)

The diagram means that given Y, f and g such that $f \circ i = g \circ j$, there is exactly one continuous map $h\colon X \to Y$ such that $h \circ j' = f$ and $h \circ i' = g$.

Under favourable conditions, the induced diagram

$$\begin{array}{ccc} \pi_1(U \cap V) & \xrightarrow{i_*} & \pi_1(U) \\ {\scriptstyle j_*}\downarrow & & \downarrow{\scriptstyle j'_*} \\ \pi_1(V) & \xrightarrow[i'_*]{} & \pi_1(X) \end{array}$$

of fundamental groups has the same property in the world of groups and group homomorphisms. This is *van Kampen's theorem*. In fact, van Kampen stated his theorem in a much more complicated way. Stating it transparently requires some categorical language, but he was working in the 1930s, before category theory had been born.

You have now seen several examples of universal properties. As this book progresses, we will develop different ways of talking about them. Once we have set up the basic vocabulary of categories and functors, we will study *adjoint functors*, then *representable functors*, then *limits*. Each of these provides an approach to universal properties, and each places the idea in a different light. For instance, Examples 0.4 and 0.5 can most readily be described in terms of adjoint functors, Example 0.6 via representable functors, and Examples 0.1, 0.2, 0.8 and 0.9 in terms of limits.

Exercises

0.10 Let S be a set. The **indiscrete** topological space $I(S)$ is the space whose set of points is S and whose only open subsets are $\emptyset$ and S itself. Imitating Example 0.5, find a universal property satisfied by the space $I(S)$.

0.11 Fix a group homomorphism $\theta\colon G \to H$. Find a universal property satisfied by the pair $(\ker(\theta), \iota)$ of diagram (0.2). (This property can – indeed, must – make reference to θ.)

0.12 Verify the universal property shown in diagram (0.3).

0.13 Denote by $\mathbb{Z}[x]$ the polynomial ring over $\mathbb{Z}$ in one variable.

(a) Prove that for all rings R and all $r \in R$, there exists a unique ring homomorphism $\phi\colon \mathbb{Z}[x] \to R$ such that $\phi(x) = r$.
(b) Let A be a ring and $a \in A$. Suppose that for all rings R and all $r \in R$, there exists a unique ring homomorphism $\phi\colon A \to R$ such that $\phi(a) = r$. Prove that there is a unique isomorphism $\iota\colon \mathbb{Z}[x] \to A$ such that $\iota(x) = a$.

0.14 Let X and Y be vector spaces.

(a) For the purposes of this exercise only, a *cone* is a triple (V, f_1, f_2) consisting of a vector space V, a linear map $f_1\colon V \to X$, and a linear map $f_2\colon V \to Y$. Find a cone (P, p_1, p_2) with the following property: for all cones (V, f_1, f_2), there exists a unique linear map $f\colon V \to P$ such that $p_1 \circ f = f_1$ and $p_2 \circ f = f_2$.
(b) Prove that there is essentially only one cone with the property stated in (a). That is, prove that if (P, p_1, p_2) and (P', p'_1, p'_2) both have this property then there is an isomorphism $i\colon P \to P'$ such that $p'_1 \circ i = p_1$ and $p'_2 \circ i = p_2$.
(c) For the purposes of this exercise only, a *cocone* is a triple (V, f_1, f_2) consisting of a vector space V, a linear map $f_1\colon X \to V$, and a linear map $f_2\colon Y \to V$. Find a cocone (Q, q_1, q_2) with the following property: for all cocones (V, f_1, f_2), there exists a unique linear map $f\colon Q \to V$ such that $f \circ q_1 = f_1$ and $f \circ q_2 = f_2$.
(d) Prove that there is essentially only one cocone with the property stated in (c), in a sense that you should make precise.

1

Categories, functors and natural transformations

A category is a system of related objects. The objects do not live in isolation: there is some notion of map between objects, binding them together.

Typical examples of what ‘object’ might mean are ‘group’ and ‘topological space’, and typical examples of what ‘map’ might mean are ‘homomorphism’ and ‘continuous map’, respectively. We will see many examples, and we will also learn that some categories have a very different flavour from the two just mentioned. In fact, the ‘maps’ of category theory need not be anything like maps in the sense that you are most likely to be familiar with.

Categories are *themselves* mathematical objects, and with that in mind, it is unsurprising that there is a good notion of ‘map between categories’. Such maps are called functors. More surprising, perhaps, is the existence of a third level: we can talk about maps between *functors*, which are called natural transformations. These, then, are maps between maps between categories.

In fact, it was the desire to formalize the notion of natural transformation that led to the birth of category theory. By the early 1940s, researchers in algebraic topology had started to use the phrase ‘natural transformation’, but only in an informal way. Two mathematicians, Samuel Eilenberg and Saunders Mac Lane, saw that a precise definition was needed. But before they could define natural transformation, they had to define functor; and before they could define functor, they had to define category. And so the subject was born.

Nowadays, the uses of category theory have spread far beyond algebraic topology. Its tentacles extend into most parts of pure mathematics. They also reach some parts of applied mathematics; perhaps most notably, category theory has become a standard tool in certain parts of computer science. Applied mathematics is more than just applied differential equations!

(d) For each field k, there is a category $\mathbf{Vect}_k$ of vector spaces over k and linear maps between them.
(e) There is a category **Top** of topological spaces and continuous maps.

This chapter is mostly about the interaction *between* categories, rather than what goes on *inside* them. We will, however, need the following definition.

Definition 1.1.4 A map $f: A \to B$ in a category $\mathscr{A}$ is an **isomorphism** if there exists a map $g: B \to A$ in $\mathscr{A}$ such that $gf = 1_A$ and $fg = 1_B$.

In the situation of Definition 1.1.4, we call g the **inverse** of f and write $g = f^{-1}$. (The word 'the' is justified by Exercise 1.1.13.) If there exists an isomorphism from A to B, we say that A and B are **isomorphic** and write $A \cong B$.

Example 1.1.5 The isomorphisms in **Set** are exactly the bijections. This statement is not quite a logical triviality. It amounts to the assertion that a function has a two-sided inverse if and only if it is injective and surjective.

Example 1.1.6 The isomorphisms in **Grp** are exactly the isomorphisms of groups. Again, this is not quite trivial, at least if you were taught that the definition of group isomorphism is 'bijective homomorphism'. In order to show that this is equivalent to being an isomorphism in **Grp**, you have to prove that the inverse of a bijective homomorphism is also a homomorphism.

Similarly, the isomorphisms in **Ring** are exactly the isomorphisms of rings.

Example 1.1.7 The isomorphisms in **Top** are exactly the homeomorphisms. Note that, in contrast to the situation in **Grp** and **Ring**, a bijective map in **Top** is not necessarily an isomorphism. A classic example is the map

$$\begin{array}{ccc} [0,1) & \to & \{z \in \mathbb{C} \mid |z| = 1\} \\ t & \mapsto & e^{2\pi i t}, \end{array}$$

which is a continuous bijection but not a homeomorphism.

The examples of categories mentioned so far are important, but could give a false impression. In each of them, the objects of the category are sets with structure (such as a group structure, a topology, or, in the case of **Set**, no structure at all). The maps are the functions preserving the structure, in the appropriate sense. And in each of them, there is a clear sense of what the elements of a given object are.

However, not all categories are like this. In general, the objects of a category are not 'sets equipped with extra stuff'. Thus, in a general category, it does not make sense to talk about the 'elements' of an object. (At least, it does not make sense in an immediately obvious way; we return to this in Definition 4.1.25.) Similarly, in a general category, the maps need not be mappings or functions in the usual sense. So:

> *The objects of a category need not be remotely like sets.*
>
> *The maps in a category need not be remotely like functions.*

The next few examples illustrate these points. They also show that, contrary to the impression that might have been given so far, categories need not be enormous. Some categories are small, manageable structures in their own right, as we now see.

Examples 1.1.8 (Categories as mathematical structures) (a) A category can be specified by saying directly what its objects, maps, composition and identities are. For example, there is a category $\emptyset$ with no objects or maps at all. There is a category $\mathbf{1}$ with one object and only the identity map. It can be drawn like this:

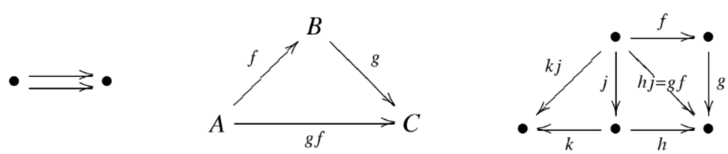
$$\bullet$$

(Since every object is required to have an identity map on it, we usually do not bother to draw the identities.) There is another category that can be drawn as

$$\bullet \to \bullet \qquad \text{or} \qquad A \xrightarrow{f} B,$$

with two objects and one non-identity map, from the first object to the second. (Composition is defined in the only possible way.) To reiterate the points made above, it is not obvious what an 'element' of A or B would be, or how one could regard f as a 'function' of any sort.

It is easy to make up more complicated examples. For instance, here are three more categories:



(b) Some categories contain no maps at all apart from identities (which, as categories, they are obliged to have). These are called **discrete** categories. A discrete category amounts to just a class of objects. More poetically, a category is a collection of objects related to one another to a greater or lesser degree; a discrete category is the extreme case in which each object is totally isolated from its companions.

(c) A group is essentially the same thing as a category that has only one object and in which all the maps are isomorphisms.

To understand this, first consider a category $\mathscr{A}$ with just one object. It is not important what letter or symbol we use to denote the object; let us call it A. Then $\mathscr{A}$ consists of a set (or class) $\mathscr{A}(A,A)$, an associative composition function

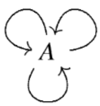
$$\circ\colon \mathscr{A}(A,A) \times \mathscr{A}(A,A) \to \mathscr{A}(A,A),$$

and a two-sided unit $1_A \in \mathscr{A}(A,A)$. This would make $\mathscr{A}(A,A)$ into a group, except that we have not mentioned inverses. However, to say that every map in $\mathscr{A}$ is an isomorphism is exactly to say that every element of $\mathscr{A}(A,A)$ has an inverse with respect to $\circ$.

If we write G for the group $\mathscr{A}(A,A)$, then the situation is this:

category $\mathscr{A}$ with single object A	*corresponding group G*
maps in $\mathscr{A}$	elements of G
$\circ$ in $\mathscr{A}$	$\cdot$ in G
1_A	$1 \in G$

The category $\mathscr{A}$ looks something like this:



The arrows represent different maps $A \to A$, that is, different elements of the group G.

What the object of $\mathscr{A}$ is called makes no difference. It matters exactly as much as whether we choose x or y or t to denote some variable in an algebra problem, which is to say, not at all. Later we will define 'equivalence' of categories, which will enable us to make a precise statement: the category of groups is equivalent to the category of (small) one-object categories in which every map is an isomorphism (Example 3.2.11).

The first time one meets the idea that a group is a kind of category, it is tempting to dismiss it as a coincidence or a trick. But it is not; there is real content.

To see this, suppose that your education had been shuffled and that you already knew about categories before being taught about groups. In your first group theory class, the lecturer declares that a group is supposed to be the system of all symmetries of an object. A symmetry of an object X, she says, is a way of mapping X to itself in a reversible or invertible manner. At this point, you realize that she is talking about a very special type of category. In general, a category is a system consisting of *all* the mappings (not usually just the invertible ones) between *many* objects (not usually just one). So a group is just a category with the special properties that all the maps are invertible and there is only one object.

(d) The inverses played no essential part in the previous example, suggesting that it is worth thinking about 'groups without inverses'. These are called monoids.

Formally, a **monoid** is a set equipped with an associative binary operation and a two-sided unit element. Groups describe the reversible transformations, or symmetries, that can be applied to an object; monoids describe the not-necessarily-reversible transformations. For instance, given any set X, there is a group consisting of all bijections $X \to X$, and there is a monoid consisting of all functions $X \to X$. In both cases, the binary operation is composition and the unit is the identity function on X. Another example of a monoid is the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of natural numbers, with $+$ as the operation and 0 as the unit. Alternatively, we could take the set $\mathbb{N}$ with $\cdot$ as the operation and 1 as the unit.

A category with one object is essentially the same thing as a monoid, by the same argument as for groups. This is stated formally in Example 3.2.11.

(e) A **preorder** is a reflexive transitive binary relation. A **preordered set** $(S, \leq)$ is a set S together with a preorder $\leq$ on it. Examples: $S = \mathbb{R}$ and $\leq$ has its usual meaning; S is the set of subsets of $\{1, \ldots, 10\}$ and $\leq$ is $\subseteq$ (inclusion); $S = \mathbb{Z}$ and $a \leq b$ means that a divides b.

A preordered set can be regarded as a category $\mathscr{A}$ in which, for each $A, B \in \mathscr{A}$, there is at most one map from A to B. To see this, consider a category $\mathscr{A}$ with this property. It is not important what letter we use to denote the unique map from an object A to an object B; all we need to record is which pairs (A, B) of objects have the property that a map $A \to B$ does exist. Let us write $A \leq B$ to mean that there exists a map $A \to B$.

Since $\mathscr{A}$ is a category, and categories have composition, if $A \leq B \leq C$ then $A \leq C$. Since categories also have identities, $A \leq A$ for all A. The associativity and identity axioms are automatic. So, $\mathscr{A}$ amounts to a collection of objects equipped with a transitive reflexive binary relation, that is, a preorder. One can think of the unique map $A \to B$ as the statement or assertion that $A \leq B$.

An **order** on a set is a preorder $\leq$ with the property that if $A \leq B$ and $B \leq A$ then $A = B$. (Equivalently, if $A \cong B$ in the corresponding category then $A = B$.) Ordered sets are also called **partially ordered sets** or **posets**.

An example of a preorder that is not an order is the divisibility relation $|$ on $\mathbb{Z}$: for there we have $2 \mid -2$ and $-2 \mid 2$ but $2 \neq -2$.

Here are two ways of constructing new categories from old.

Construction 1.1.9 Every category $\mathscr{A}$ has an **opposite** or **dual** category $\mathscr{A}^{\mathrm{op}}$, defined by reversing the arrows. Formally, $\mathrm{ob}(\mathscr{A}^{\mathrm{op}}) = \mathrm{ob}(\mathscr{A})$ and $\mathscr{A}^{\mathrm{op}}(B, A) = \mathscr{A}(A, B)$ for all objects A and B. Identities in $\mathscr{A}^{\mathrm{op}}$ are the same as in $\mathscr{A}$. Composition in $\mathscr{A}^{\mathrm{op}}$ is the same as in $\mathscr{A}$, but with the arguments reversed. To spell this out: if $A \xrightarrow{f} B \xrightarrow{g} C$ are maps in $\mathscr{A}^{\mathrm{op}}$ then $A \xleftarrow{f} B \xleftarrow{g} C$ are maps in $\mathscr{A}$; these give rise to a map $A \xleftarrow{f \circ g} C$ in $\mathscr{A}$, and the composite of the original pair of maps is the corresponding map $A \to C$ in $\mathscr{A}^{\mathrm{op}}$.

So, arrows $A \to B$ in $\mathscr{A}$ correspond to arrows $B \to A$ in $\mathscr{A}^{\mathrm{op}}$. According to the definition above, if $f\colon A \to B$ is an arrow in $\mathscr{A}$ then the corresponding arrow $B \to A$ in $\mathscr{A}^{\mathrm{op}}$ is also called f. Some people prefer to give it a different name, such as f^{op}.

Remark 1.1.10 The **principle of duality** is fundamental to category theory. Informally, it states that every categorical definition, theorem and proof has a **dual**, obtained by reversing all the arrows. Invoking the principle of duality can save work: given any theorem, reversing the arrows throughout its statement and proof produces a dual theorem. Numerous examples of duality appear throughout this book.

Construction 1.1.11 Given categories $\mathscr{A}$ and $\mathscr{B}$, there is a **product category** $\mathscr{A} \times \mathscr{B}$, in which

$$\mathrm{ob}(\mathscr{A} \times \mathscr{B}) = \mathrm{ob}(\mathscr{A}) \times \mathrm{ob}(\mathscr{B}),$$
$$(\mathscr{A} \times \mathscr{B})((A, B), (A', B')) = \mathscr{A}(A, A') \times \mathscr{B}(B, B').$$

Put another way, an object of the product category $\mathscr{A} \times \mathscr{B}$ is a pair (A, B) where $A \in \mathscr{A}$ and $B \in \mathscr{B}$. A map $(A, B) \to (A', B')$ in $\mathscr{A} \times \mathscr{B}$ is a pair (f, g) where $f\colon A \to A'$ in $\mathscr{A}$ and $g\colon B \to B'$ in $\mathscr{B}$. For the definitions of composition and identities in $\mathscr{A} \times \mathscr{B}$, see Exercise 1.1.14.

Exercises

1.1.12 Find three examples of categories not mentioned above.

1.1.13 Show that a map in a category can have at most one inverse. That is, given a map $f\colon A \to B$, show that there is at most one map $g\colon B \to A$ such that $gf = 1_A$ and $fg = 1_B$.

The forgetful functors in examples (a)–(c) forget *structure* on the objects, but that of example (d) forgets a *property*. Nevertheless, it turns out to be convenient to use the same word, 'forgetful', in both situations.

Although forgetting is a trivial operation, there are situations in which it is powerful. For example, it is a theorem that the order of any finite field is a prime power. An important step in the proof is to simply forget that the field is a field, remembering only that it is a vector space over its subfield $\{0, 1, 1+1, 1+1+1, \ldots\}$.

Examples 1.2.4 **Free functors** are in some sense dual to forgetful functors (as we will see in the next chapter), although they are less elementary. Again, 'free functor' is an informal but useful term.

(a) Given any set S, one can build the **free group** $F(S)$ on S. This is a group containing S as a subset and with no further properties other than those it is forced to have, in a sense made precise in Section 2.1. Intuitively, the group $F(S)$ is obtained from the set S by adding just enough new elements that it becomes a group, but without imposing any equations other than those forced by the definition of group.

A little more precisely, the elements of $F(S)$ are formal expressions or **words** such as $x^{-4}yx^2zy^{-3}$ (where $x, y, z \in S$). Two such words are seen as equal if one can be obtained from the other by the usual cancellation rules, so that, for example, x^3xy, x^4y, and $x^2y^{-1}yx^2y$ all represent the same element of $F(S)$. To multiply two words, just write one followed by the other; for instance, $x^{-4}yx$ times xzy^{-3} is $x^{-4}yx^2zy^{-3}$.

This construction assigns to each set S a group $F(S)$. In fact, F is a functor: any map of sets $f\colon S \to S'$ gives rise to a homomorphism of groups $F(f)\colon F(S) \to F(S')$. For instance, take the map of sets

$$f\colon \{w, x, y, z\} \to \{u, v\}$$

defined by $f(w) = f(x) = f(y) = u$ and $f(z) = v$. This gives rise to a homomorphism

$$F(f)\colon F(\{w, x, y, z\}) \to F(\{u, v\}),$$

which maps $x^{-4}yx^2zy^{-3} \in F(\{w, x, y, z\})$ to

$$u^{-4}uu^2vu^{-3} = u^{-1}vu^{-3} \in F(\{u, v\}).$$

(b) Similarly, we can construct the free commutative ring $F(S)$ on a set S, giving a functor F from **Set** to the category **CRing** of commutative rings. In fact, $F(S)$ is something familiar, namely, the ring of polynomials over $\mathbb{Z}$ in commuting variables x_s ($s \in S$). (A polynomial is, after all, just a formal expression built from the variables using the ring operations +, − and ·.) For example, if S is a two-element set then $F(S) \cong \mathbb{Z}[x, y]$.

(c) We can also construct the free vector space on a set. Fix a field k. The free functor $F\colon \mathbf{Set} \to \mathbf{Vect}_k$ is defined on objects by taking $F(S)$ to be a vector space with basis S. Any two such vector spaces are isomorphic; but it is perhaps not obvious that there is any such vector space at all, so we have to construct one. Loosely, $F(S)$ is the set of all formal k-linear combinations of elements of S, that is, expressions

$$\sum_{s \in S} \lambda_s s$$

where each λ_s is a scalar and there are only finitely many values of s such that $\lambda_s \neq 0$. (This restriction is imposed because one can only take *finite* sums in a vector space.) Elements of $F(S)$ can be added:

$$\sum_{s \in S} \lambda_s s + \sum_{s \in S} \mu_s s = \sum_{s \in S} (\lambda_s + \mu_s) s.$$

There is also a scalar multiplication on $F(S)$:

$$c \cdot \sum_{s \in S} \lambda_s s = \sum_{s \in S} (c\lambda_s) s$$

($c \in k$). In this way, $F(S)$ becomes a vector space.

To be completely precise and avoid talking about 'expressions', we can define $F(S)$ to be the set of all functions $\lambda\colon S \to k$ such that $\{s \in S \mid \lambda(s) \neq 0\}$ is finite. (Think of such a function λ as corresponding to the expression $\sum_{s \in S} \lambda(s) s$.) To define addition on $F(S)$, we must define for each $\lambda, \mu \in F(S)$ a sum $\lambda + \mu \in F(S)$; it is given by

$$(\lambda + \mu)(s) = \lambda(s) + \mu(s)$$

($s \in S$). Similarly, the scalar multiplication is given by $(c \cdot \lambda)(s) = c \cdot \lambda(s)$ ($c \in k$, $\lambda \in F(S)$, $s \in S$).

Rings and vector spaces have the special property that it is relatively easy to write down an explicit formula for the free functor. The case of groups is much more typical. For most types of algebraic structure, describing the free functor requires as much fussy work as it does for groups. We return to this point in Example 2.1.3 and Example 6.3.11 (where we see how to avoid the fussy work entirely).

Examples 1.2.5 (Functors in algebraic topology) Historically, some of the first examples of functors arose in algebraic topology. There, the strategy is to learn about a space by extracting data from it in some clever way, assembling that data into an algebraic structure, then studying the algebraic structure instead of the original space. Algebraic topology therefore involves many functors from categories of spaces to categories of algebras.

(a) Let $\mathbf{Top}_*$ be the category of topological spaces equipped with a basepoint, together with the continuous basepoint-preserving maps. There is a functor $\pi_1 : \mathbf{Top}_* \to \mathbf{Grp}$ assigning to each space X with basepoint x the fundamental group $\pi_1(X, x)$ of X at x. (Some texts use the simpler notation $\pi_1(X)$, ignoring the choice of basepoint. This is more or less safe if X is path-connected, but strictly speaking, the basepoint should always be specified.)

    That π_1 is a functor means that it not only assigns to each space-with-basepoint (X, x) a group $\pi_1(X, x)$, but also assigns to each basepoint-preserving continuous map

$$f\colon (X, x) \to (Y, y)$$

    a homomorphism

$$\pi_1(f)\colon \pi_1(X, x) \to \pi_1(Y, y).$$

    Usually $\pi_1(f)$ is written as f_*. The functoriality axioms say that $(g \circ f)_* = g_* \circ f_*$ and $(1_{(X,x)})_* = 1_{\pi_1(X,x)}$.

(b) For each $n \in \mathbb{N}$, there is a functor $H_n\colon \mathbf{Top} \to \mathbf{Ab}$ assigning to a space its nth homology group (in any of several possible senses).

Example 1.2.6 Any system of polynomial equations such as

$$2x^2 + y^2 - 3z^2 = 1 \tag{1.1}$$

$$x^3 + x = y^2 \tag{1.2}$$

gives rise to a functor $\mathbf{CRing} \to \mathbf{Set}$. Indeed, for each commutative ring A, let $F(A)$ be the set of triples $(x, y, z) \in A \times A \times A$ satisfying equations (1.1) and (1.2). Whenever $f\colon A \to B$ is a ring homomorphism and $(x, y, z) \in F(A)$, we have $(f(x), f(y), f(z)) \in F(B)$; so the map of rings $f\colon A \to B$ induces a map of sets $F(f)\colon F(A) \to F(B)$. This defines a functor $F\colon \mathbf{CRing} \to \mathbf{Set}$.

In algebraic geometry, a **scheme** is a functor $\mathbf{CRing} \to \mathbf{Set}$ with certain properties. (This is not the most common way of phrasing the definition, but it is equivalent.) The functor F above is a simple example.

Example 1.2.7 Let G and H be monoids (or groups, if you prefer), regarded as one-object categories $\mathscr{G}$ and $\mathscr{H}$. A functor $F : \mathscr{G} \to \mathscr{H}$ must send the unique object of $\mathscr{G}$ to the unique object of $\mathscr{H}$, so it is determined by its effect on maps. Hence, the functor $F\colon \mathscr{G} \to \mathscr{H}$ amounts to a function $F\colon G \to H$ such that $F(g'g) = F(g')F(g)$ for all $g', g \in G$, and $F(1) = 1$. In other words, a functor $\mathscr{G} \to \mathscr{H}$ is just a homomorphism $G \to H$.

Example 1.2.8 Let G be a monoid, regarded as a one-object category $\mathscr{G}$. A functor $F\colon \mathscr{G} \to \mathbf{Set}$ consists of a set S (the value of F at the unique object of $\mathscr{G}$) together with, for each $g \in G$, a function $F(g)\colon S \to S$, satisfying the functoriality axioms. Writing $(F(g))(s) = g \cdot s$, we see that the functor F amounts to a set S together with a function

$$\begin{array}{ccc} G \times S & \to & S \\ (g, s) & \mapsto & g \cdot s \end{array}$$

satisfying $(g'g) \cdot s = g' \cdot (g \cdot s)$ and $1 \cdot s = s$ for all $g, g' \in G$ and $s \in S$. In other words, a functor $\mathscr{G} \to \mathbf{Set}$ is a set equipped with a left action by G: a **left G-set**, for short.

Similarly, a functor $\mathscr{G} \to \mathbf{Vect}_k$ is exactly a k-linear representation of G, in the sense of representation theory. This can reasonably be taken as the *definition* of representation.

Example 1.2.9 When A and B are (pre)ordered sets, a functor between the corresponding categories is exactly an **order-preserving map**, that is, a function $f\colon A \to B$ such that $a \leq a' \implies f(a) \leq f(a')$. Exercise 1.2.22 asks you to verify this.

Sometimes we meet functor-like operations that reverse the arrows, with a map $A \to A'$ in $\mathscr{A}$ giving rise to a map $F(A) \leftarrow F(A')$ in $\mathscr{B}$. Such operations are called contravariant functors.

Definition 1.2.10 Let $\mathscr{A}$ and $\mathscr{B}$ be categories. A **contravariant functor** from $\mathscr{A}$ to $\mathscr{B}$ is a functor $\mathscr{A}^{\mathrm{op}} \to \mathscr{B}$.

To avoid confusion, we write 'a contravariant functor from $\mathscr{A}$ to $\mathscr{B}$' rather than 'a contravariant functor $\mathscr{A} \to \mathscr{B}$'.

Functors $\mathscr{C} \to \mathscr{D}$ correspond one-to-one with functors $\mathscr{C}^{\mathrm{op}} \to \mathscr{D}^{\mathrm{op}}$, and $(\mathscr{A}^{\mathrm{op}})^{\mathrm{op}} = \mathscr{A}$, so a contravariant functor from $\mathscr{A}$ to $\mathscr{B}$ can also be described as a functor $\mathscr{A} \to \mathscr{B}^{\mathrm{op}}$. Which description we use is not enormously important, but in the long run, the convention in Definition 1.2.10 makes life easier.

An ordinary functor $\mathscr{A} \to \mathscr{B}$ is sometimes called a **covariant functor** from $\mathscr{A}$ to $\mathscr{B}$, for emphasis.

Example 1.2.11 We can tell a lot about a space by examining the functions on it. The importance of this principle in twentieth- and twenty-first-century mathematics can hardly be exaggerated.

For example, given a topological space X, let $C(X)$ be the ring of continuous real-valued functions on X. The ring operations are defined ‘pointwise’: for instance, if $p_1, p_2 \colon X \to \mathbb{R}$ are continuous maps then the map $p_1 + p_2 \colon X \to \mathbb{R}$ is defined by

$$(p_1 + p_2)(x) = p_1(x) + p_2(x)$$

($x \in X$). A continuous map $f \colon X \to Y$ induces a ring homomorphism $C(f)\colon C(Y) \to C(X)$, defined at $q \in C(Y)$ by taking $(C(f))(q)$ to be the composite map

$$X \xrightarrow{f} Y \xrightarrow{q} \mathbb{R}.$$

Note that $C(f)$ goes in the opposite direction from f. After checking some axioms (Exercise 1.2.26), we conclude that C is a contravariant functor from **Top** to **Ring**.

While this particular example will not play a large part in this text, it is worth close attention. It illustrates the important idea of a structure whose elements are maps (in this case, a ring whose elements are continuous functions). The way in which C becomes a functor, via composition, is also important. Similar constructions will be crucial in later chapters.

For certain classes of space, the passage from X to $C(X)$ loses no information: there is a way of reconstructing the space X from the ring $C(X)$. For this and related reasons, it is sometimes said that ‘algebra is dual to geometry’.

Example 1.2.12 Let k be a field. For any two vector spaces V and W over k, there is a vector space

$$\mathbf{Hom}(V, W) = \{\text{linear maps } V \to W\}.$$

The elements of this vector space are themselves maps, and the vector space operations (addition and scalar multiplication) are defined pointwise, as in the last example.

Now fix a vector space W. Any linear map $f \colon V \to V'$ induces a linear map

$$f^* \colon \mathbf{Hom}(V', W) \to \mathbf{Hom}(V, W),$$

defined at $q \in \mathbf{Hom}(V', W)$ by taking $f^*(q)$ to be the composite map

$$V \xrightarrow{f} V' \xrightarrow{q} W.$$

This defines a functor

$$\mathbf{Hom}(-, W) \colon \mathbf{Vect}_k^{\mathrm{op}} \to \mathbf{Vect}_k.$$