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!

1.1 Categories

Definition 1.1.1 A **category** $\mathscr{A}$ consists of:

- a collection $\mathrm{ob}(\mathscr{A})$ of **objects**;
- for each $A, B \in \mathrm{ob}(\mathscr{A})$, a collection $\mathscr{A}(A, B)$ of **maps** or **arrows** or **morphisms** from A to B;
- for each $A, B, C \in \mathrm{ob}(\mathscr{A})$, a function

$$\begin{array}{ccc} \mathscr{A}(B, C) \times \mathscr{A}(A, B) & \to & \mathscr{A}(A, C) \\ (g, f) & \mapsto & g \circ f, \end{array}$$

  called **composition**;
- for each $A \in \mathrm{ob}(\mathscr{A})$, an element 1_A of $\mathscr{A}(A, A)$, called the **identity** on A,

satisfying the following axioms:

- **associativity**: for each $f \in \mathscr{A}(A, B)$, $g \in \mathscr{A}(B, C)$ and $h \in \mathscr{A}(C, D)$, we have $(h \circ g) \circ f = h \circ (g \circ f)$;
- **identity laws**: for each $f \in \mathscr{A}(A, B)$, we have $f \circ 1_A = f = 1_B \circ f$.

Remarks 1.1.2 (a) We often write:

$$\begin{array}{rll} A \in \mathscr{A} & \text{to mean} & A \in \mathrm{ob}(\mathscr{A}); \\ f \colon A \to B \text{ or } A \xrightarrow{f} B & \text{to mean} & f \in \mathscr{A}(A, B); \\ gf & \text{to mean} & g \circ f. \end{array}$$

People also write $\mathscr{A}(A, B)$ as $\mathrm{Hom}_{\mathscr{A}}(A, B)$ or $\mathrm{Hom}(A, B)$. The notation 'Hom' stands for homomorphism, from one of the earliest examples of a category.

(b) The definition of category is set up so that in general, from each string

$$A_0 \xrightarrow{f_1} A_1 \xrightarrow{f_2} \cdots \xrightarrow{f_n} A_n$$

of maps in $\mathscr{A}$, it is possible to construct exactly one map

$$A_0 \to A_n$$

(namely, $f_n f_{n-1} \cdots f_1$). If we are given extra information then we may be able to construct other maps $A_0 \to A_n$; for instance, if we happen to know that $A_{n-1} = A_n$, then $f_{n-1} f_{n-2} \cdots f_1$ is another such map. But we are speaking here of the *general* situation, in the absence of extra information.

For example, a string like this with $n = 4$ gives rise to maps

$$\begin{tikzcd} A_0 \arrow[r, shift left, "((f_4 f_3) f_2) f_1"] \arrow[r, shift right, "(f_4(1_{A_3} f_3))((f_2 f_1) 1_{A_0})"'] & A_4, \end{tikzcd}$$

but the axioms imply that they are equal. It is safe to omit the brackets and write both as $f_4 f_3 f_2 f_1$.

Here it is intended that $n \geq 0$. In the case $n = 0$, the statement is that for each object A_0 of a category, it is possible to construct exactly one map $A_0 \to A_0$ (namely, the identity 1_{A_0}). An identity map can be thought of as a zero-fold composite, in much the same way that the number 1 can be thought of as the product of zero numbers.

(c) We often speak of **commutative diagrams**. For instance, given objects and maps

$$\begin{tikzcd}
A \arrow[rr, "f"] \arrow[d, "h"'] & & B \arrow[d, "g"] \\
C \arrow[r, "i"'] & D \arrow[r, "j"'] & E
\end{tikzcd}$$

in a category, we say that the diagram **commutes** if $gf = jih$. Generally, a diagram is said to commute if whenever there are two paths from an object X to an object Y, the map from X to Y obtained by composing along one path is equal to the map obtained by composing along the other.

(d) The slightly vague word 'collection' means *roughly* the same as 'set', although if you know about such things, it is better to interpret it as meaning 'class'. We come back to this in Chapter **??**.

(e) If $f \in \mathscr{A}(A, B)$, we call A the **domain** and B the **codomain** of f. Every map in every category has a definite domain and a definite codomain. (If you believe it makes sense to form the intersection of an arbitrary pair of abstract sets, you should add to the definition of category the condition that $\mathscr{A}(A, B) \cap \mathscr{A}(A', B') = \emptyset$ unless $A = A'$ and $B = B'$.)

Examples 1.1.3 (Categories of mathematical structures) (a) There is a category **Set** described as follows. Its objects are sets. Given sets A and B, a map from A to B in the category **Set** is exactly what is ordinarily called a map (or mapping, or function) from A to B. Composition in the category is ordinary composition of functions, and the identity maps are again what you would expect.

In situations such as this, we often do not bother to specify the composition and identities. We write 'the category of sets and functions', leaving the reader to guess the rest. In fact, we usually go further and call it just 'the category of sets'.

(b) There is a category **Grp** of groups, whose objects are groups and whose maps are group homomorphisms.

(c) Similarly, there is a category **Ring** of rings and ring homomorphisms.

(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 $f \colon A \to B$ in a category $\mathscr{A}$ is an **isomorphism** if there exists a map $g \colon B \to A$ 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 3.1.25.) 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:

$$\begin{tikzcd}
\bullet \arrow[r, shift left] \arrow[r, shift right] & \bullet
\end{tikzcd}
\qquad
\begin{tikzcd}
& B \arrow[dr, "g"] & \\
A \arrow[ur, "f"] \arrow[rr, "gf"'] & & C
\end{tikzcd}
\qquad
\begin{tikzcd}
& \bullet \arrow[r, "f"] \arrow[dl, "kj"'] \arrow[d, "j"] \arrow[dr, "hj=gf"] & \bullet \arrow[d, "g"] \\
\bullet & \bullet \arrow[l, "k"] \arrow[r, "h"'] & \bullet
\end{tikzcd}$$

(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

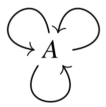
$$\circ\colon \mathscr{A}(A, A) \times \mathscr{A}(A, A) \to \mathscr{A}(A, A),$$

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 **??**).

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 $X \longrightarrow X$, and there is a monoid consisting of all functions $X \longrightarrow X$. 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 **??**.

(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 $A \longrightarrow B$ does exist. Let us write $A \leq B$ to mean that there exists a map $A \longrightarrow B$.

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 $A \longrightarrow B$ 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 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 $fg = 1_B$.

1.1.14 Let $\mathscr{A}$ and $\mathscr{B}$ be categories. Construction 1.1.11 defined the product category $\mathscr{A} \times \mathscr{B}$, except that the definitions of composition and identities in $\mathscr{A} \times \mathscr{B}$ were not given. There is only one sensible way to define them; write it down.

1.1.15 There is a category **Toph** whose objects are topological spaces and whose maps $X \to Y$ are homotopy classes of continuous maps from X to Y. What do you need to know about homotopy in order to prove that **Toph** is a category? What does it mean, in purely topological terms, for two objects of **Toph** to be isomorphic?

1.2 Functors

One of the lessons of category theory is that whenever we meet a new type of mathematical object, we should always ask whether there is a sensible notion of 'map' between such objects. We can ask this about categories themselves. The answer is yes, and a map between categories is called a functor.

Definition 1.2.1 Let $\mathscr{A}$ and $\mathscr{B}$ be categories. A **functor** $F\colon \mathscr{A} \to \mathscr{B}$ consists of:

- a function
$$\mathrm{ob}(\mathscr{A}) \to \mathrm{ob}(\mathscr{B}),$$
written as $A \mapsto F(A)$;
- for each $A, A' \in \mathscr{A}$, a function
$$\mathscr{A}(A, A') \to \mathscr{B}(F(A), F(A')),$$
written as $f \mapsto F(f)$,

satisfying the following axioms:

- $F(f' \circ f) = F(f') \circ F(f)$ whenever $A \xrightarrow{f} A' \xrightarrow{f'} A''$ in $\mathscr{A}$;
- $F(1_A) = 1_{F(A)}$ whenever $A \in \mathscr{A}$.

Remarks 1.2.2 (a) The definition of functor is set up so that from each string
$$A_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} A_n$$
of maps in $\mathscr{A}$ (with $n \geq 0$), it is possible to construct exactly one map
$$F(A_0) \to F(A_n)$$
in $\mathscr{B}$. For example, given maps
$$A_0 \xrightarrow{f_1} A_1 \xrightarrow{f_2} A_2 \xrightarrow{f_3} A_3 \xrightarrow{f_4} A_4$$

in $\mathscr{A}$, we can construct maps

$$\begin{tikzcd} F(A_0) \arrow[r, shift left, "F(f_4 f_3)F(f_2 f_1)"] \arrow[r, shift right, "F(1_{A_4})F(f_4)F(f_3 f_2)F(f_1)"'] & F(A_4) \end{tikzcd}$$

in $\mathscr{B}$, but the axioms imply that they are equal.

(b) We are familiar with the idea that structures and the structure-preserving maps between them form a category (such as **Grp**, **Ring**, etc.). In particular, this applies to categories and functors: there is a category **CAT** whose objects are categories and whose maps are functors.

One part of this statement is that functors can be composed. That is, given functors $\mathscr{A} \xrightarrow{F} \mathscr{B} \xrightarrow{G} \mathscr{C}$, there arises a new functor $\mathscr{A} \xrightarrow{G \circ F} \mathscr{C}$, defined in the obvious way. Another is that for every category $\mathscr{A}$, there is an identity functor $1_{\mathscr{A}} \colon \mathscr{A} \to \mathscr{A}$.

Examples 1.2.3 Perhaps the easiest examples of functors are the so-called **forgetful functors**. (This is an informal term, with no precise definition.) For instance:

(a) There is a functor $U \colon$ **Grp** $\to$ **Set** defined as follows: if G is a group then $U(G)$ is the underlying set of G (that is, its set of elements), and if $f \colon G \to H$ is a group homomorphism then $U(f)$ is the function f itself. So U forgets the group structure of groups and forgets that group homomorphisms are homomorphisms.

(b) Similarly, there is a functor **Ring** $\to$ **Set** forgetting the ring structure on rings, and (for any field k) there is a functor $\mathbf{Vect}_k \to$ **Set** forgetting the vector space structure on vector spaces.

(c) Forgetful functors do not have to forget *all* the structure. For example, let **Ab** be the category of abelian groups. There is a functor **Ring** $\to$ **Ab** that forgets the multiplicative structure, remembering just the underlying additive group. Or, let **Mon** be the category of monoids. There is a functor $U \colon$ **Ring** $\to$ **Mon** that forgets the additive structure, remembering just the underlying multiplicative monoid. (That is, if R is a ring then $U(R)$ is the set R made into a monoid via $\cdot$ and 1.)

(d) There is an inclusion functor $U \colon$ **Ab** $\to$ **Grp** defined by $U(A) = A$ for any abelian group A and $U(f) = f$ for any homomorphism f of abelian groups. It forgets that abelian groups are abelian.

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 $f : S \to S'$ gives rise to a homomorphism of groups $F(f) : F(S) \to F(S')$. For instance, take the map of sets

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

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

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

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 $\cdot$.) 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 **??** (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 **Top**$_*$ be the category of topological spaces equipped with a basepoint, together with the continuous basepoint-preserving maps. There is a functor $\pi_1 \colon \mathbf{Top}_* \to \mathbf{Grp}$ 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 gives rise to a functor **CRing** → **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 (**??**) and (**??**). 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 **CRing** → **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 $F \colon \mathscr{G} \to \mathscr{H}$ 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 $\mathscr{G} \to$ **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

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

The symbol '$-$' is a blank or placeholder, into which arguments can be inserted. Thus, the value of $\mathbf{Hom}(-, W)$ at V is $\mathbf{Hom}(V, W)$. Sometimes we use a blank space instead of $-$, as in $\mathbf{Hom}(\ \ , W)$.

An important special case is where W is k, seen as a one-dimensional vector space over itself. The vector space $\mathbf{Hom}(V, k)$ is called the **dual** of V, and is written as V^*. So there is a contravariant functor

$$(\ \ )^* = \mathbf{Hom}(-, k)\colon \mathbf{Vect}_k^{\mathrm{op}} \to \mathbf{Vect}_k$$

sending each vector space to its dual.

Example 1.2.13 For each $n \in \mathbb{N}$, there is a functor $H^n\colon \mathbf{Top}^{\mathrm{op}} \to \mathbf{Ab}$ assigning to a space its nth cohomology group.

Example 1.2.14 Let G be a monoid, regarded as a one-object category $\mathscr{G}$. A functor $\mathscr{G}^{\mathrm{op}} \to \mathbf{Set}$ is a *right* G-set, for essentially the same reasons as in Example 1.2.8.

That left actions are covariant functors and right actions are contravariant functors is a consequence of a basic notational choice: we write the value of a function f at an element x as $f(x)$, not $(x)f$.

Contravariant functors whose codomain is **Set** are important enough to have their own special name.

Definition 1.2.15 Let $\mathscr{A}$ be a category. A **presheaf** on $\mathscr{A}$ is a functor $\mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$.

The name comes from the following special case. Let X be a topological space. Write $\mathscr{O}(X)$ for the poset of open subsets of X, ordered by inclusion. View $\mathscr{O}(X)$ as a category, as in Example 1.1.8(e). Thus, the objects of $\mathscr{O}(X)$ are the open subsets of X, and for $U, U' \in \mathscr{O}(X)$, there is one map $U \to U'$ if $U \subseteq U'$, and there are none otherwise. A **presheaf** on the space X is a presheaf on the category $\mathscr{O}(X)$. For example, given any space X, there is a presheaf F on X defined by

$$F(U) = \{\text{continuous functions } U \to \mathbb{R}\}$$

$(U \in \mathscr{O}(X))$ and, whenever $U \subseteq U'$ are open subsets of X, by taking the map $F(U') \to F(U)$ to be restriction. Presheaves, and a certain class of presheaves called sheaves, play an important role in modern geometry.

We know very well that for *functions* between *sets*, it is sometimes useful to consider special kinds of function such as injections, surjections and bijections. We also know that the notions of injection and subset are related: for instance, whenever B is a subset of A, there is an injection $B \to A$ given by inclusion. In this section and the next, we introduce some similar notions for *functors* between *categories*, beginning with the following definitions.

Definition 1.2.16 A functor $F\colon \mathscr{A} \to \mathscr{B}$ is **faithful** (respectively, **full**) if for each $A, A' \in \mathscr{A}$, the function

$$\begin{array}{ccc} \mathscr{A}(A, A') & \to & \mathscr{B}(F(A), F(A')) \\ f & \mapsto & F(f) \end{array}$$

is injective (respectively, surjective).

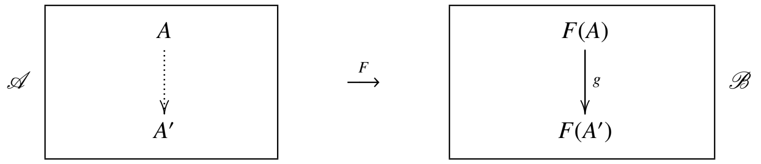


Figure 1.1 Fullness and faithfulness.

Warning 1.2.17 Note the roles of A and A' in the definition. Faithfulness does *not* say that if f_1 and f_2 are distinct maps in $\mathscr{A}$ then $F(f_1) \neq F(f_2)$ (Exercise 1.2.27). In the situation of Figure 1.1, F is faithful if for each A, A' and g as shown, there is at most one dotted arrow that F sends to g. It is full if for each such A, A' and g, there is at least one dotted arrow that F sends to g.

Definition 1.2.18 Let $\mathscr{A}$ be a category. A **subcategory** $\mathscr{S}$ of $\mathscr{A}$ consists of a subclass $\text{ob}(\mathscr{S})$ of $\text{ob}(\mathscr{A})$ together with, for each $S, S' \in \text{ob}(\mathscr{S})$, a subclass $\mathscr{S}(S, S')$ of $\mathscr{A}(S, S')$, such that $\mathscr{S}$ is closed under composition and identities. It is a **full** subcategory if $\mathscr{S}(S, S') = \mathscr{A}(S, S')$ for all $S, S' \in \text{ob}(\mathscr{S})$.

A full subcategory therefore consists of a selection of the objects, with all of the maps between them. So, a full subcategory can be specified simply by saying what its objects are. For example, **Ab** is the full subcategory of **Grp** consisting of the groups that are abelian.

Whenever $\mathscr{S}$ is a subcategory of a category $\mathscr{A}$, there is an inclusion functor $I : \mathscr{S} \to \mathscr{A}$ defined by $I(S) = S$ and $I(f) = f$. It is automatically faithful, and it is full if and only if $\mathscr{S}$ is a full subcategory.

Warning 1.2.19 The image of a functor need not be a subcategory. For example, consider the functor

$$\left(\begin{tikzcd} A \arrow[r, "f"] & B & B' \arrow[r, "g"] & C \end{tikzcd}\right) \xrightarrow{F} \left(\begin{tikzcd} & Y \arrow[dr, "q"] & \\ X \arrow[ur, "p"] \arrow[rr, "qp"'] & & Z \end{tikzcd}\right)$$

defined by $F(A) = X$, $F(B) = F(B') = Y$, $F(C) = Z$, $F(f) = p$, and $F(g) = q$. Then p and q are in the image of F, but qp is not.

Exercises

1.2.20 Find three examples of functors not mentioned above.

1.2.21 Show that functors preserve isomorphism. That is, prove that if $F\colon \mathscr{A} \to \mathscr{B}$ is a functor and $A, A' \in \mathscr{A}$ with $A \cong A'$, then $F(A) \cong F(A')$.

1.2.22 Prove the assertion made in Example 1.2.9. In other words, given ordered sets A and B, and denoting by $\mathscr{A}$ and $\mathscr{B}$ the corresponding categories, show that a functor $\mathscr{A} \to \mathscr{B}$ amounts to an order-preserving map $A \to B$.

1.2.23 Two categories $\mathscr{A}$ and $\mathscr{B}$ are **isomorphic**, written as $\mathscr{A} \cong \mathscr{B}$, if they are isomorphic as objects of **CAT**.

(a) Let G be a group, regarded as a one-object category all of whose maps are isomorphisms. Then its opposite G^{op} is also a one-object category all of whose maps are isomorphisms, and can therefore be regarded as a group too. What is G^{op}, in purely group-theoretic terms? Prove that G is isomorphic to G^{op}.

(b) Find a monoid not isomorphic to its opposite.

1.2.24 Is there a functor $Z\colon \mathbf{Grp} \to \mathbf{Grp}$ with the property that $Z(G)$ is the centre of G for all groups G?

1.2.25 Sometimes we meet functors whose domain is a product $\mathscr{A} \times \mathscr{B}$ of categories. Here you will show that such a functor can be regarded as an interlocking pair of families of functors, one defined on $\mathscr{A}$ and the other defined on $\mathscr{B}$. (This is very like the situation for bilinear and linear maps.)

(a) Let $F\colon \mathscr{A} \times \mathscr{B} \to \mathscr{C}$ be a functor. Prove that for each $A \in \mathscr{A}$, there is a functor $F^A\colon \mathscr{B} \to \mathscr{C}$ defined on objects $B \in \mathscr{B}$ by $F^A(B) = F(A, B)$ and on maps g in $\mathscr{B}$ by $F^A(g) = F(1_A, g)$. Prove that for each $B \in \mathscr{B}$, there is a functor $F_B\colon \mathscr{A} \to \mathscr{C}$ defined similarly.

(b) Let $F\colon \mathscr{A} \times \mathscr{B} \to \mathscr{C}$ be a functor. With notation as in (a), show that the families of functors $(F^A)_{A\in\mathscr{A}}$ and $(F_B)_{B\in\mathscr{B}}$ satisfy the following two conditions:

- if $A \in \mathscr{A}$ and $B \in \mathscr{B}$ then $F^A(B) = F_B(A)$;
- if $f\colon A \to A'$ in $\mathscr{A}$ and $g\colon B \to B'$ in $\mathscr{B}$ then $F^{A'}(g) \circ F_B(f) = F_{B'}(f) \circ F^A(g)$.

(c) Now take categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$, and take families of functors $(F^A)_{A\in\mathscr{A}}$ and $(F_B)_{B\in\mathscr{B}}$ satisfying the two conditions in (b). Prove that there is a unique functor $F\colon \mathscr{A} \times \mathscr{B} \to \mathscr{C}$ satisfying the equations in (a). (‘There is a unique functor’ means in particular that there *is* a functor, so you have to prove existence as well as uniqueness.)

1.2.26 Fill in the details of Example 1.2.11, thus constructing a functor $C\colon \mathbf{Top}^{\mathrm{op}} \to \mathbf{Ring}$.

1.2.27 Find an example of a functor $F\colon \mathscr{A} \to \mathscr{B}$ such that F is faithful but there exist distinct maps f_1 and f_2 in $\mathscr{A}$ with $F(f_1) = F(f_2)$.

1.2.28 (a) Of the examples of functors appearing in this section, which are faithful and which are full?

(b) Write down one example of a functor that is both full and faithful, one that is full but not faithful, one that is faithful but not full, and one that is neither.

1.2.29 (a) What are the subcategories of an ordered set? Which are full?

(b) What are the subcategories of a group? (Careful!) Which are full?

1.3 Natural transformations

We now know about categories. We also know about functors, which are maps between categories. Perhaps surprisingly, there is a further notion of 'map between functors'. Such maps are called natural transformations. This notion only applies when the functors have the same domain and codomain:

$$\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, shift right, "G"'] & \mathscr{B} \end{tikzcd}.$$

To see how this might work, let us consider a special case. Let $\mathscr{A}$ be the discrete category (Example 1.1.8(b)) whose objects are the natural numbers $0, 1, 2, \ldots$. A functor F from $\mathscr{A}$ to another category $\mathscr{B}$ is simply a sequence $(F_0, F_1, F_2, \ldots)$ of objects of $\mathscr{B}$. Let G be another functor from $\mathscr{A}$ to $\mathscr{B}$, consisting of another sequence $(G_0, G_1, G_2, \ldots)$ of objects of $\mathscr{B}$. It would be reasonable to define a 'map' from F to G to be a sequence

$$\Bigl(F_0 \xrightarrow{\alpha_0} G_0,\ F_1 \xrightarrow{\alpha_1} G_1,\ F_1 \xrightarrow{\alpha_2} G_2,\ \ldots\Bigr)$$

of maps in $\mathscr{B}$. The situation can be depicted as follows:

$$\mathscr{A} \quad \boxed{0 \quad 1 \quad 2 \quad \cdots} \qquad\qquad \boxed{\begin{tikzcd} F_0 \arrow[d, "\alpha_0"] & F_1 \arrow[d, "\alpha_1"] & F_2 \arrow[d, "\alpha_2"] & \cdots \\ G_0 & G_1 & G_2 & \end{tikzcd}} \quad \mathscr{B}$$

(The right-hand diagram should not be understood too literally. Some of the objects F_i or G_i might be equal, and there might be much else in $\mathscr{B}$ besides what is shown.)

This suggests that in the general case, a natural transformation between functors $\mathscr{A} \underset{G}{\overset{F}{\rightrightarrows}} \mathscr{B}$ should consist of maps $\alpha_A \colon F(A) \to G(A)$, one for each $A \in \mathscr{A}$. In the example above, the category $\mathscr{A}$ had the special property of not containing any nontrivial maps. In general, we demand some kind of compatibility between the maps in $\mathscr{A}$ and the maps α_A.

Definition 1.3.1 Let $\mathscr{A}$ and $\mathscr{B}$ be categories and let $\mathscr{A} \underset{G}{\overset{F}{\rightrightarrows}} \mathscr{B}$ be functors. A **natural transformation** $\alpha \colon F \to G$ is a family $\left(F(A) \xrightarrow{\alpha_A} G(A)\right)_{A \in \mathscr{A}}$ of maps in $\mathscr{B}$ such that for every map $A \xrightarrow{f} A'$ in $\mathscr{A}$, the square

$$\begin{tikzcd}
F(A) \arrow[r, "F(f)"] \arrow[d, "\alpha_A"'] & F(A') \arrow[d, "\alpha_{A'}"] \\
G(A) \arrow[r, "G(f)"'] & G(A')
\end{tikzcd} \qquad (1.1)$$

commutes. The maps α_A are called the **components** of α.

Remarks 1.3.2 (a) The definition of natural transformation is set up so that from each map $A \xrightarrow{f} A'$ in $\mathscr{A}$, it is possible to construct exactly one map $F(A) \to G(A')$ in $\mathscr{B}$. When $f = 1_A$, this map is α_A. For a general f, it is the diagonal of the square (1.1), and 'exactly one' implies that the square commutes.

(b) We write

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left, "F"{name=U}] \arrow[r, bend right, "G"'{name=D}] & \mathscr{B} \arrow[from=U, to=D, Rightarrow, "\alpha"]
\end{tikzcd}$$

to mean that α is a natural transformation from F to G.

Example 1.3.3 Let $\mathscr{A}$ be a discrete category, and let $F, G \colon \mathscr{A} \to \mathscr{B}$ be functors. Then F and G are just families $(F(A))_{A \in \mathscr{A}}$ and $(G(A))_{A \in \mathscr{A}}$ of objects of $\mathscr{B}$. A natural transformation $\alpha \colon F \to G$ is just a family $\left(F(A) \xrightarrow{\alpha_A} G(A)\right)_{A \in \mathscr{A}}$ of maps in $\mathscr{B}$, as claimed above in the case $\operatorname{ob} \mathscr{A} = \mathbb{N}$. In principle, this family must satisfy the naturality axiom (1.1) for every map f in $\mathscr{A}$; but the only maps in $\mathscr{A}$ are the identities, and when f is an identity, this axiom holds automatically.

Example 1.3.4 Recall from Examples 1.1.8 that a group (or more generally, a monoid) G can be regarded as a one-object category. Also recall from Example 1.2.8 that a functor from the category G to **Set** is nothing but a left G-set. (Previously we used $\mathscr{G}$ to denote the category corresponding to the group G;

from now on we use G to denote them both.) Take two G-sets, S and T. Since S and T can be regarded as functors $G \to \mathbf{Set}$, we can ask: what is a natural transformation

$$\begin{tikzcd} G \arrow[r, bend left, "S"{name=U}] \arrow[r, bend right, "T"'{name=D}] & \mathbf{Set}, \arrow[from=U, to=D, Rightarrow, "\alpha"] \end{tikzcd}$$

in concrete terms?

Such a natural transformation consists of a single map in **Set** (since G has just one object), satisfying some axioms. Precisely, it is a function $\alpha\colon S \to T$ such that $\alpha(g \cdot s) = g \cdot \alpha(s)$ for all $s \in S$ and $g \in G$. (Why?) In other words, it is just a map of G-sets, sometimes called a **G-equivariant** map.

Example 1.3.5 Fix a natural number n. In this example, we will see how 'determinant of an $n \times n$ matrix' can be understood as a natural transformation.

For any commutative ring R, the $n \times n$ matrices with entries in R form a monoid $M_n(R)$ under multiplication. Moreover, any ring homomorphism $R \to S$ induces a monoid homomorphism $M_n(R) \to M_n(S)$. This defines a functor $M_n\colon \mathbf{CRing} \to \mathbf{Mon}$ from the category of commutative rings to the category of monoids.

Also, the elements of any ring R form a monoid $U(R)$ under multiplication, giving another functor $U\colon \mathbf{CRing} \to \mathbf{Mon}$.

Now, every $n \times n$ matrix X over a commutative ring R has a determinant $\det_R(X)$, which is an element of R. Familiar properties of determinant –

$$\det_R(XY) = \det_R(X)\det_R(Y), \qquad \det_R(I) = 1$$

– tell us that for each R, the function $\det_R\colon M_n(R) \to U(R)$ is a monoid homomorphism. So, we have a family of maps

$$\Big(M_n(R) \xrightarrow{\det_R} U(R)\Big)_{R \in \mathbf{CRing}},$$

and it makes sense to ask whether they define a natural transformation

$$\begin{tikzcd} \mathbf{CRing} \arrow[r, bend left, "M_n"{name=U}] \arrow[r, bend right, "U"'{name=D}] & \mathbf{Mon}. \arrow[from=U, to=D, Rightarrow, "\det"] \end{tikzcd}$$

Indeed, they do. That the naturality squares commute (check!) reflects the fact that determinant is defined in the same way for all rings. We do not use one definition of determinant for one ring and a different definition for another ring. Generally speaking, the naturality axiom (1.1) is supposed to capture the idea that the family $(\alpha_A)_{A \in \mathscr{A}}$ is defined in a uniform way across all $A \in \mathscr{A}$.

Construction 1.3.6 Natural transformations are a kind of map, so we would expect to be able to compose them. We can. Given natural transformations

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=60, "F", ""{name=U, below}] \arrow[r, "G" description, ""{name=M}] \arrow[r, bend right=60, "H"', ""{name=D, above}] & \mathscr{B} ,
\arrow[Rightarrow, from=U, to=M, "\alpha"] \arrow[Rightarrow, from=M, to=D, "\beta"]
\end{tikzcd}$$

there is a composite natural transformation

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=40, "F", ""{name=U, below}] \arrow[r, bend right=40, "H"', ""{name=D, above}] & \mathscr{B}
\arrow[Rightarrow, from=U, to=D, "\beta\circ\alpha"]
\end{tikzcd}$$

defined by $(\beta \circ \alpha)_A = \beta_A \circ \alpha_A$ for all $A \in \mathscr{A}$. There is also an identity natural transformation

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=40, "F", ""{name=U, below}] \arrow[r, bend right=40, "F"', ""{name=D, above}] & \mathscr{B}
\arrow[Rightarrow, from=U, to=D, "1_F"]
\end{tikzcd}$$

on any functor F, defined by $(1_F)_A = 1_{F(A)}$. So for any two categories $\mathscr{A}$ and $\mathscr{B}$, there is a category whose objects are the functors from $\mathscr{A}$ to $\mathscr{B}$ and whose maps are the natural transformations between them. This is called the **functor category** from $\mathscr{A}$ to $\mathscr{B}$, and written as $[\mathscr{A}, \mathscr{B}]$ or $\mathscr{B}^{\mathscr{A}}$.

Example 1.3.7 Let 2 be the discrete category with two objects. A functor from 2 to a category $\mathscr{B}$ is a pair of objects of $\mathscr{B}$, and a natural transformation is a pair of maps. The functor category $[2, \mathscr{B}]$ is therefore isomorphic to the product category $\mathscr{B} \times \mathscr{B}$ (Construction 1.1.11). This fits well with the alternative notation $\mathscr{B}^2$ for the functor category.

Example 1.3.8 Let G be a monoid. Then $[G, \mathbf{Set}]$ is the category of left G-sets, and $[G^{\mathrm{op}}, \mathbf{Set}]$ is the category of right G-sets (Example 1.2.14).

Example 1.3.9 Take ordered sets A and B, viewed as categories (as in Example 1.1.8(e)). Given order-preserving maps $A \underset{g}{\overset{f}{\rightrightarrows}} B$, viewed as functors (as in Example 1.2.9), there is at most one natural transformation

$$\begin{tikzcd}
A \arrow[r, bend left=40, "f", ""{name=U, below}] \arrow[r, bend right=40, "g"', ""{name=D, above}] & B,
\arrow[Rightarrow, from=U, to=D]
\end{tikzcd}$$

and there is one if and only if $f(a) \leq g(a)$ for all $a \in A$. (The naturality axiom (1.1) holds automatically, because in an ordered set, all diagrams commute.) So $[A, B]$ is an ordered set too; its elements are the order-preserving maps from A to B, and $f \leq g$ if and only if $f(a) \leq g(a)$ for all $a \in A$.

Everyday phrases such as '*the* cyclic group of order 6' and '*the* product of two spaces' reflect the fact that given two isomorphic objects of a category, we usually neither know nor care whether they are actually equal. This is enormously important.

In particular, the lesson applies when the category concerned is a functor category. In other words, given two functors $F, G\colon \mathscr{A} \to \mathscr{B}$, we usually do not care whether they are literally equal. (Equality would imply that the objects $F(A)$ and $G(A)$ of $\mathscr{B}$ were equal for all $A \in \mathscr{A}$, a level of detail in which we have just declared ourselves to be uninterested.) What really matters is whether they are naturally isomorphic.

Definition 1.3.10 Let $\mathscr{A}$ and $\mathscr{B}$ be categories. A **natural isomorphism** between functors from $\mathscr{A}$ to $\mathscr{B}$ is an isomorphism in $[\mathscr{A}, \mathscr{B}]$.

An equivalent form of the definition is often useful:

Lemma 1.3.11 *Let*
$$\begin{tikzcd} \mathscr{A} \arrow[r, bend left, "F"{name=U}] \arrow[r, bend right, "G"'{name=D}] & \mathscr{B} \arrow[Rightarrow, from=U, to=D, "\alpha"] \end{tikzcd}$$
be a natural transformation. Then α is a natural isomorphism if and only if $\alpha_A\colon F(A) \to G(A)$ is an isomorphism for all $A \in \mathscr{A}$.

Proof Exercise 1.3.26. □

Of course, we say that functors F and G are **naturally isomorphic** if there exists a natural isomorphism from F to G. Since natural isomorphism is just isomorphism in a particular category (namely, $[\mathscr{A}, \mathscr{B}]$), we already have notation for this: $F \cong G$.

Definition 1.3.12 Given functors
$$\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, shift right, "G"'] & \mathscr{B} \end{tikzcd}$$
, we say that

$$F(A) \cong G(A) \textbf{ naturally in } A$$

if F and G are naturally isomorphic.

This alternative terminology can be understood as follows. If $F(A) \cong G(A)$ naturally in A then certainly $F(A) \cong G(A)$ for each individual A, but more is true: we can choose isomorphisms $\alpha_A\colon F(A) \to G(A)$ in such a way that the naturality axiom (1.1) is satisfied.

Example 1.3.13 Let $F, G\colon \mathscr{A} \to \mathscr{B}$ be functors from a discrete category $\mathscr{A}$ to a category $\mathscr{B}$. Then $F \cong G$ if and only if $F(A) \cong G(A)$ for all $A \in \mathscr{A}$.

So in *this* case, $F(A) \cong G(A)$ naturally in A if and only if $F(A) \cong G(A)$ for all A. But this is only true because $\mathscr{A}$ is discrete. In general, it is emphatically false. There are many examples of categories and functors $\mathscr{A} \overset{F}{\underset{G}{\rightrightarrows}} \mathscr{B}$ such that $F(A) \cong G(A)$ for all $A \in \mathscr{A}$, but not *naturally* in A. Exercise 1.3.31 gives an example from combinatorics.

Example 1.3.14 Let **FDVect** be the category of finite-dimensional vector spaces over some field k. The dual vector space construction defines a contravariant functor from **FDVect** to itself (Example 1.2.12), and the double dual construction therefore defines a covariant functor from **FDVect** to itself.

Moreover, we have for each $V \in$ **FDVect** a canonical isomorphism $\alpha_V\colon V \to V^{**}$. Given $v \in V$, the element $\alpha_V(v)$ of V^{**} is 'evaluation at v'; that is, $\alpha_V(v)\colon V^* \to k$ maps $\phi \in V^*$ to $\phi(v) \in k$. That α_V is an isomorphism is a standard result in the theory of finite-dimensional vector spaces.

This defines a natural transformation

$$\begin{tikzcd}
\mathbf{FDVect} \arrow[r, bend left, "1_{\mathbf{FDVect}}"] \arrow[r, bend right, "(\ )^{**}"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathbf{FDVect}
\end{tikzcd}$$

from the identity functor to the double dual functor. By Lemma 1.3.11, α is a natural isomorphism. So $1_{\mathbf{FDVect}} \cong (\ )^{**}$. Equivalently, in the language of Definition 1.3.12, $V \cong V^{**}$ naturally in V.

This is one of those occasions on which category theory makes an intuition precise. In some informal sense, evident before you learn anything about category theory, the isomorphism between a finite-dimensional vector space and its double dual is 'natural' or 'canonical': no arbitrary choices are needed in order to define it. In contrast, to specify an isomorphism between V and its single dual V^*, we need to make an arbitrary choice of basis, and the isomorphism really does depend on the basis that we choose.

In the example on vector spaces, the word **canonical** was used. It is an informal word, meaning something like 'God-given' or 'defined without making arbitrary choices'. For example, for any two sets A and B, there is a canonical bijection $A \times B \to B \times A$ defined by $(a, b) \mapsto (b, a)$, and there is a canonical function $A \times B \to A$ defined by $(a, b) \mapsto a$. But the function $B \to A$ defined by 'choose an element $a_0 \in A$ and send everything to a_0' is not canonical, because the choice of a_0 is arbitrary.

The concept of natural isomorphism leads unavoidably to another central concept: equivalence of categories.

Two elements of a set are either equal or not. Two objects of a category can be equal, not equal but isomorphic, or not even isomorphic. As explained before Definition 1.3.10, the notion of equality between two objects of a category is unreasonably strict; it is usually isomorphism that we care about. So:

- the right notion of sameness of two elements of a set is equality;
- the right notion of sameness of two objects of a category is isomorphism.

When applied to a functor category $[\mathscr{A}, \mathscr{B}]$, the second point tells us that:

- the right notion of sameness of two functors $\mathscr{A} \rightrightarrows \mathscr{B}$ is natural isomorphism.

But what is the right notion of sameness of two *categories*? Isomorphism is unreasonably strict, as if $\mathscr{A} \cong \mathscr{B}$ then there are functors

$$\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd} \tag{1.2}$$

such that

$$G \circ F = 1_{\mathscr{A}} \qquad \text{and} \qquad F \circ G = 1_{\mathscr{B}}, \tag{1.3}$$

and we have just seen that the notion of equality between functors is too strict. The most useful notion of sameness of categories, called 'equivalence', is looser than isomorphism. To obtain the definition, we simply replace the unreasonably strict equalities in (1.3) by isomorphisms. This gives

$$G \circ F \cong 1_{\mathscr{A}} \qquad \text{and} \qquad F \circ G \cong 1_{\mathscr{B}}.$$

Definition 1.3.15 An **equivalence** between categories $\mathscr{A}$ and $\mathscr{B}$ consists of a pair (1.2) of functors together with natural isomorphisms

$$\eta \colon 1_{\mathscr{A}} \to G \circ F, \qquad \varepsilon \colon F \circ G \to 1_{\mathscr{B}}.$$

If there exists an equivalence between $\mathscr{A}$ and $\mathscr{B}$, we say that $\mathscr{A}$ and $\mathscr{B}$ are **equivalent**, and write $\mathscr{A} \simeq \mathscr{B}$. We also say that the functors F and G are **equivalences**.

The directions of η and ε are not very important, since they are isomorphisms anyway. The reason for this particular choice will become apparent when we come to discuss adjunctions (Section 2.2).

Warning 1.3.16 The symbol $\cong$ is used for isomorphism of objects of a category, and in particular for isomorphism of categories (which are objects of **CAT**). The symbol $\simeq$ is used for equivalence of categories. At least, this is the convention used in this book and by most category theorists, although it is far from universal in mathematics at large.

There is a very useful alternative characterization of those functors that are equivalences. First, we need a definition.

Definition 1.3.17 A functor $F \colon \mathscr{A} \to \mathscr{B}$ is **essentially surjective on objects** if for all $B \in \mathscr{B}$, there exists $A \in \mathscr{A}$ such that $F(A) \cong B$.

Proposition 1.3.18 *A functor is an equivalence if and only if it is full, faithful and essentially surjective on objects.*

Proof Exercise 1.3.32. □

This result can be compared to the theorem that every bijective group homomorphism is an isomorphism (that is, its inverse is also a homomorphism), or that a natural transformation whose components are isomorphisms is itself an isomorphism (Lemma 1.3.11). Those two results are useful because they allow us to show that a map is an isomorphism without directly constructing an inverse. Proposition 1.3.18 provides a similar service, enabling us to prove that a functor F is an equivalence without actually constructing an 'inverse' G, or indeed an η or an ε (in the notation of Definition 1.3.15).

A corollary of Proposition 1.3.18 invites us to view full and faithful functors as, essentially, inclusions of full subcategories:

Corollary 1.3.19 *Let $F \colon \mathscr{C} \to \mathscr{D}$ be a full and faithful functor. Then $\mathscr{C}$ is equivalent to the full subcategory $\mathscr{C}'$ of $\mathscr{D}$ whose objects are those of the form $F(C)$ for some $C \in \mathscr{C}$.*

Proof The functor $F' \colon \mathscr{C} \to \mathscr{C}'$ defined by $F'(C) = F(C)$ is full and faithful (since F is) and essentially surjective on objects (by definition of $\mathscr{C}'$). □

This result is true, with the same proof, whether we interpret 'of the form $F(C)$' to mean 'equal to $F(C)$' or 'isomorphic to $F(C)$'.

Example 1.3.20 Let $\mathscr{A}$ be any category, and let $\mathscr{B}$ be any full subcategory containing at least one object from each isomorphism class of $\mathscr{A}$. Then the inclusion functor $\mathscr{B} \hookrightarrow \mathscr{A}$ is faithful (like any inclusion of subcategories), full, and essentially surjective on objects. Hence $\mathscr{B} \simeq \mathscr{A}$.

So if we take a category and remove some (but not all) of the objects in each isomorphism class, the slimmed-down version is equivalent to the original.

Conversely, if we take a category and throw in some more objects, each of them isomorphic to one of the existing objects, it makes no difference: the new, bigger, category is equivalent to the old one.

For example, let **FinSet** be the category of finite sets and functions between them. For each natural number n, choose a set $\mathbf{n}$ with n elements, and let $\mathscr{B}$ be the full subcategory of **FinSet** with objects $\mathbf{0}, \mathbf{1}, \ldots$. Then $\mathscr{B} \simeq \mathbf{FinSet}$, even though $\mathscr{B}$ is in some sense much smaller than **FinSet**.

Example 1.3.21 In Example 1.1.8(d), we saw that monoids are essentially the same thing as one-object categories. With the definition of equivalence in hand, we are nearly ready to make this statement precise. We are missing some set-theoretic language, and we will return to this result once we have that language (Example **??**), but the essential point can be stated now.

Let $\mathscr{C}$ be the full subcategory of **CAT** whose objects are the one-object categories. Let **Mon** be the category of monoids. Then $\mathscr{C} \simeq \mathbf{Mon}$. To see this, first note that given any object A of any category, the maps $A \to A$ form a monoid under composition (at least, subject to some set-theoretic restrictions). There is, therefore, a canonical functor $F : \mathscr{C} \to \mathbf{Mon}$ sending a one-object category to the monoid of maps from the single object to itself. This functor F is full and faithful (by Example 1.2.7) and essentially surjective on objects. Hence F is an equivalence.

Example 1.3.22 An equivalence of the form $\mathscr{A}^{\mathrm{op}} \simeq \mathscr{B}$ is sometimes called a **duality** between $\mathscr{A}$ and $\mathscr{B}$. One says that $\mathscr{A}$ is **dual** to $\mathscr{B}$. There are many famous dualities in which $\mathscr{A}$ is a category of algebras and $\mathscr{B}$ is a category of spaces; recall the slogan ‘algebra is dual to geometry’ from Example 1.2.11.

Here are some quite advanced examples, well beyond the scope of this book.

- Stone duality: the category of Boolean algebras is dual to the category of totally disconnected compact Hausdorff spaces.
- Gelfand–Naimark duality: the category of commutative unital C^*-algebras is dual to the category of compact Hausdorff spaces. (C^*-algebras are certain algebraic structures important in functional analysis.)
- Algebraic geometers have several notions of ‘space’, one of which is ‘affine variety’. Let k be an algebraically closed field. Then the category of affine varieties over k is dual to the category of finitely generated k-algebras with no nontrivial nilpotents.
- Pontryagin duality: the category of locally compact abelian topological groups is dual to itself. As the words ‘topological group’ suggest, both sides of the duality are algebraic *and* geometric. Pontryagin duality is an abstraction of the properties of the Fourier transform.

Example 1.3.23 It is rarely useful to consider a category of structured objects in which the maps do not respect that structure. For instance, let $\mathscr{A}$ be the category whose objects are groups and whose maps are *all* functions between them, not necessarily homomorphisms. Let $\mathbf{Set}_{\neq\emptyset}$ be the category of nonempty sets. The forgetful functor $U\colon \mathscr{A} \to \mathbf{Set}_{\neq\emptyset}$ is full and faithful. It is a (not profound) fact that every nonempty set can be given at least one group structure, so U is essentially surjective on objects. Hence U is an equivalence. This implies that the category $\mathscr{A}$, although defined in terms of groups, is really just the category of nonempty sets.

Remarks 1.3.24 Here is a kind of review of the chapter so far. We have defined:

- categories (Section 1.1);
- functors between categories (Section 1.2);
- natural transformations between functors (Section 1.3);
- composition of functors

$$\begin{tikzcd} \cdot \arrow[r] & \cdot \arrow[r] & \cdot \end{tikzcd}$$

  and the identity functor on any category (Remark 1.2.2(b));
- composition of natural transformations

$$\begin{tikzcd} \cdot \arrow[r, bend left=60, ""{name=U, below}] \arrow[r, ""{name=M}] \arrow[r, bend right=60, ""{name=D, above}] \arrow[Rightarrow, from=U, to=M] \arrow[Rightarrow, from=M, to=D] & \cdot \end{tikzcd}$$

  and the identity natural transformation on any functor (Construction 1.3.6).

This composition of natural transformations is sometimes called **vertical composition**. There is also **horizontal composition**, which takes natural transformations

$$\begin{tikzcd} \mathscr{A} \arrow[r, bend left, "F", ""{name=A, below}] \arrow[r, bend right, "G"', ""{name=B, above}] \arrow[Rightarrow, from=A, to=B, "\alpha"] & \mathscr{A}' \arrow[r, bend left, "F'", ""{name=C, below}] \arrow[r, bend right, "G'"', ""{name=D, above}] \arrow[Rightarrow, from=C, to=D, "\alpha'"] & \mathscr{A}'' \end{tikzcd}$$

and produces a natural transformation

$$\begin{tikzcd} \mathscr{A} \arrow[r, bend left, "F'\circ F", ""{name=A, below}] \arrow[r, bend right, "G'\circ G"', ""{name=B, above}] \arrow[Rightarrow, from=A, to=B] & \mathscr{A}'', \end{tikzcd}$$

traditionally written as $\alpha' * \alpha$. The component of $\alpha' * \alpha$ at $A \in \mathscr{A}$ is defined to be the diagonal of the naturality square

$$\begin{tikzcd}
F'(F(A)) \arrow[r, "F'(\alpha_A)"] \arrow[d, "\alpha'_{F(A)}"'] & F'(G(A)) \arrow[d, "\alpha'_{G(A)}"] \\
G'(F(A)) \arrow[r, "G'(\alpha_A)"'] & G'(G(A)).
\end{tikzcd}$$

In other words, $(\alpha' * \alpha)_A$ can be defined as either $\alpha'_{G(A)} \circ F'(\alpha_A)$ or $G'(\alpha_A) \circ \alpha'_{F(A)}$; it makes no difference which, since they are equal.

The special cases of horizontal composition where either α or α' is an identity are especially important, and have their own notation. Thus,

$$\begin{tikzcd}
\mathscr{A} \arrow[r, "F"] & \mathscr{A}' \arrow[r, bend left=40, "F'"{name=U}] \arrow[r, bend right=40, "G'"'{name=D}] \arrow[Rightarrow, from=U, to=D, "\alpha'"] & \mathscr{A}''
\end{tikzcd}
\quad \text{gives rise to} \quad
\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=40, "F' \circ F"{name=U}] \arrow[r, bend right=40, "G' \circ F"'{name=D}] \arrow[Rightarrow, from=U, to=D, "\alpha' F"] & \mathscr{A}''
\end{tikzcd}$$

where $(\alpha' F)_A = \alpha'_{F(A)}$, and

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=40, "F"{name=U}] \arrow[r, bend right=40, "G"'{name=D}] \arrow[Rightarrow, from=U, to=D, "\alpha"] & \mathscr{A}' \arrow[r, "F'"] & \mathscr{A}''
\end{tikzcd}
\quad \text{gives rise to} \quad
\begin{tikzcd}
\mathscr{A} \arrow[r, bend left=40, "F' \circ F"{name=U}] \arrow[r, bend right=40, "F' \circ G"'{name=D}] \arrow[Rightarrow, from=U, to=D, "F' \alpha"] & \mathscr{A}''
\end{tikzcd}$$

where $(F'\alpha)_A = F'(\alpha_A)$.

Vertical and horizontal composition interact well: natural transformations

$$\begin{tikzcd}[column sep=large]
\mathscr{A} \arrow[r, bend left=70, "F"{name=F}] \arrow[r, "G"{name=G, description}] \arrow[r, bend right=70, "H"'{name=H}] \arrow[Rightarrow, from=F, to=G, "\alpha"] \arrow[Rightarrow, from=G, to=H, "\beta"] & \mathscr{A}' \arrow[r, bend left=70, "F'"{name=F2}] \arrow[r, "G'"{name=G2, description}] \arrow[r, bend right=70, "H'"'{name=H2}] \arrow[Rightarrow, from=F2, to=G2, "\alpha'"] \arrow[Rightarrow, from=G2, to=H2, "\beta'"] & \mathscr{A}''
\end{tikzcd}$$

obey the **interchange law**,

$$(\beta' \circ \alpha') * (\beta \circ \alpha) = (\beta' * \beta) \circ (\alpha' * \alpha) \colon F' \circ F \to H' \circ H.$$

As usual, a statement on composition is accompanied by a statement on identities: $1_{F'} * 1_F = 1_{F' \circ F}$ too.

All of this enables us to construct, for any categories $\mathscr{A}$, $\mathscr{A}'$ and $\mathscr{A}''$, a functor

$$[\mathscr{A}', \mathscr{A}''] \times [\mathscr{A}, \mathscr{A}'] \to [\mathscr{A}, \mathscr{A}''],$$

given on objects by $(F', F) \mapsto F' \circ F$ and on maps by $(\alpha', \alpha) \mapsto \alpha' * \alpha$. In particular, if $F' \cong G'$ and $F \cong G$ then $F' \circ F \cong G' \circ G$, since functors preserve isomorphism (Exercise 1.2.21).

(The existence of this functor is similar to the fact that *inside* a category $\mathscr{C}$, we have, for any objects A, A' and A'', a funct*ion*

$$\mathscr{C}(A', A'') \times \mathscr{C}(A, A') \to \mathscr{C}(A, A''),$$

given by $(f', f) \mapsto f' \circ f$.)

The diagrams above contain not only objects (0-dimensional) and arrows $\to$ (1-dimensional), but also double arrows $\Rightarrow$ sweeping out 2-dimensional regions between arrows. What we are implicitly doing is called 2-category theory. There is a 2-category of categories, functors and natural transformations, whose anatomy we have just been describing. If we are really serious about categories, we have to get serious about 2-categories. And if we are really serious about 2-categories, we have to get serious about 3-categories... and before we know it, we are studying ∞-categories. But in this book, we climb no higher than the first rung or two of this infinite ladder.

Exercises

1.3.25 Find three examples of natural transformations not mentioned above.

1.3.26 Prove Lemma 1.3.11.

1.3.27 Let $\mathscr{A}$ and $\mathscr{B}$ be categories. Prove that $[\mathscr{A}^{\mathrm{op}}, \mathscr{B}^{\mathrm{op}}] \cong [\mathscr{A}, \mathscr{B}]^{\mathrm{op}}$.

1.3.28 Let A and B be sets, and denote by B^A the set of functions from A to B. Write down:

(a) a canonical function $A \times B^A \to B$;
(b) a canonical function $A \to B^{(B^A)}$.

(Although in principle there could be many such canonical functions, in both these cases there is only one.)

1.3.29 Here we consider natural transformations between functors whose domain is a product category $\mathscr{A} \times \mathscr{B}$. Your task is to show that naturality in two variables simultaneously is equivalent to naturality in each variable separately.

Take functors $F, G\colon \mathscr{A} \times \mathscr{B} \to \mathscr{C}$. For each $A \in \mathscr{A}$, there are functors $F^A, G^A\colon \mathscr{B} \to \mathscr{C}$, as in Exercise 1.2.25. Similarly, for each $B \in \mathscr{B}$, there are functors $F_B, G_B\colon \mathscr{A} \to \mathscr{C}$.

Let $(\alpha_{A,B}\colon F(A, B) \to G(A, B))_{A \in \mathscr{A}, B \in \mathscr{B}}$ be a family of maps. Show that this family is a natural transformation $F \to G$ if and only if it satisfies the following two conditions:

- for each $A \in \mathscr{A}$, the family $(\alpha_{A,B}\colon F^A(B) \to G^A(B))_{B\in\mathscr{B}}$ is a natural transformation $F^A \to G^A$;
- for each $B \in \mathscr{B}$, the family $(\alpha_{A,B}\colon F_B(A) \to G_B(A))_{A\in\mathscr{A}}$ is a natural transformation $F_B \to G_B$.

1.3.30 Let G be a group. For each $g \in G$, there is a unique homomorphism $\phi\colon \mathbb{Z} \to G$ satisfying $\phi(1) = g$. Thus, elements of G are essentially the same thing as homomorphisms $\mathbb{Z} \to G$. When groups are regarded as one-object categories, homomorphisms $\mathbb{Z} \to G$ are in turn the same as functors $\mathbb{Z} \to G$. Natural isomorphism defines an equivalence relation on the set of functors $\mathbb{Z} \to G$, and, therefore, an equivalence relation on G itself. What is this equivalence relation, in purely group-theoretic terms?

(First have a guess. For a general group G, what equivalence relations on G can you think of?)

1.3.31 A **permutation** of a set X is a bijection $X \to X$. Write $\mathbf{Sym}(X)$ for the set of permutations of X. A **total order** on a set X is an order $\leq$ such that for all $x, y \in X$, either $x \leq y$ or $y \leq x$; so a total order on a finite set amounts to a way of placing its elements in sequence. Write $\mathbf{Ord}(X)$ for the set of total orders on X.

Let $\mathscr{B}$ denote the category of finite sets and bijections.

(a) Give a definition of **Sym** on maps in $\mathscr{B}$ in such a way that **Sym** becomes a functor $\mathscr{B} \to \mathbf{Set}$. Do the same for **Ord**. Both your definitions should be canonical (no arbitrary choices).
(b) Show that there is no natural transformation $\mathbf{Sym} \to \mathbf{Ord}$. (Hint: consider identity permutations.)
(c) For an n-element set X, how many elements do the sets $\mathbf{Sym}(X)$ and $\mathbf{Ord}(X)$ have?

Conclude that $\mathbf{Sym}(X) \cong \mathbf{Ord}(X)$ for all $X \in \mathscr{B}$, but not *naturally* in $X \in \mathscr{B}$. (The moral is that for each finite set X, there are exactly as many permutations of X as there are total orders on X, but there is no natural way of matching them up.)

1.3.32 In this exercise, you will prove Proposition 1.3.18. Let $F\colon \mathscr{A} \to \mathscr{B}$ be a functor.

(a) Suppose that F is an equivalence. Prove that F is full, faithful and essentially surjective on objects. (Hint: prove faithfulness before fullness.)
(b) Now suppose instead that F is full, faithful and essentially surjective on objects. For each $B \in \mathscr{B}$, choose an object $G(B)$ of $\mathscr{A}$ and an isomorphism $\varepsilon_B\colon F(G(B)) \to B$. Prove that G extends to a functor in such a way that

$(\varepsilon_B)_{B\in\mathscr{B}}$ is a natural isomorphism $FG \to 1_{\mathscr{B}}$. Then construct a natural isomorphism $1_{\mathscr{A}} \to GF$, thus proving that F is an equivalence.

1.3.33 This exercise makes precise the idea that linear algebra can equivalently be done with matrices or with linear maps.

Fix a field k. Let **Mat** be the category whose objects are the natural numbers and with

$$\mathbf{Mat}(m, n) = \{n \times m \text{ matrices over } k\}.$$

Prove that **Mat** is equivalent to **FDVect**, the category of finite-dimensional vector spaces over k. Does your equivalence involve a *canonical* functor from **Mat** to **FDVect**, or from **FDVect** to **Mat**?

(Part of the exercise is to work out what composition in the category **Mat** is supposed to be; there is only one sensible possibility. Proposition 1.3.18 makes the exercise easier.)

1.3.34 Show that equivalence of categories is an equivalence relation. (Not as obvious as it looks.)

2

Adjoints

The slogan of Saunders Mac Lane's book *Categories for the Working Mathematician* is:

> *Adjoint functors arise everywhere.*

We will see the truth of this, meeting examples of adjoint functors from diverse parts of mathematics. To complement the understanding provided by examples, we will approach the theory of adjoints from three different directions, each of which carries its own intuition. Then we will prove that the three approaches are equivalent.

Understanding adjointness gives you a valuable addition to your mathematical toolkit. Most professional pure mathematicians know what categories and functors are, but far fewer know about adjoints. More should: adjoint functors are both common and easy, and knowing about adjoints helps you to spot patterns in the mathematical landscape.

2.1 Definition and examples

Consider a pair of functors in opposite directions, $F\colon \mathscr{A} \to \mathscr{B}$ and $G\colon \mathscr{B} \to \mathscr{A}$. Roughly speaking, F is said to be left adjoint to G if, whenever $A \in \mathscr{A}$ and $B \in \mathscr{B}$, maps $F(A) \to B$ are essentially the same thing as maps $A \to G(B)$.

Definition 2.1.1 Let $\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd}$ be categories and functors. We say that F is **left adjoint** to G, and G is **right adjoint** to F, and write $F \dashv G$, if

$$\mathscr{B}(F(A), B) \cong \mathscr{A}(A, G(B)) \tag{2.1}$$

naturally in $A \in \mathscr{A}$ and $B \in \mathscr{B}$. The meaning of 'naturally' is defined below. An **adjunction** between F and G is a choice of natural isomorphism (2.1).

'Naturally in $A \in \mathscr{A}$ and $B \in \mathscr{B}$' means that there is a specified bijection (2.1) for each $A \in \mathscr{A}$ and $B \in \mathscr{B}$, and that it satisfies a naturality axiom. To state it, we need some notation. Given objects $A \in \mathscr{A}$ and $B \in \mathscr{B}$, the correspondence (2.1) between maps $F(A) \to B$ and $A \to G(B)$ is denoted by a horizontal bar, in both directions:

$$\begin{aligned}\bigl(F(A) \xrightarrow{g} B\bigr) &\quad\mapsto\quad \bigl(A \xrightarrow{\bar{g}} G(B)\bigr),\\ \bigl(F(A) \xrightarrow{\bar{f}} B\bigr) &\quad\mapsfrom\quad \bigl(A \xrightarrow{f} G(B)\bigr).\end{aligned}$$

So $\bar{\bar{f}} = f$ and $\bar{\bar{g}} = g$. We call $\bar{f}$ the **transpose** of f, and similarly for g. The naturality axiom has two parts:

$$\overline{\bigl(F(A) \xrightarrow{g} B \xrightarrow{q} B'\bigr)} \quad=\quad \bigl(A \xrightarrow{\bar{g}} G(B) \xrightarrow{G(q)} G(B')\bigr) \tag{2.2}$$

(that is, $\overline{q \circ g} = G(q) \circ \bar{g}$) for all g and q, and

$$\overline{\bigl(A' \xrightarrow{p} A \xrightarrow{f} G(B)\bigr)} \quad=\quad \bigl(F(A') \xrightarrow{F(p)} F(A) \xrightarrow{\bar{f}} B\bigr) \tag{2.3}$$

for all p and f. It makes no difference whether we put the long bar over the left or the right of these equations, since bar is self-inverse.

Remarks 2.1.2 (a) The naturality axiom might seem ad hoc, but we will see in Chapter 3 that it simply says that two particular functors are naturally isomorphic. In this section, we ignore the naturality axiom altogether, trusting that it embodies our usual intuitive idea of naturality: something defined without making any arbitrary choices.

(b) The naturality axiom implies that from each array of maps

$$A_0 \to \cdots \to A_n, \quad F(A_n) \to B_0, \quad B_0 \to \cdots \to B_m,$$

it is possible to construct exactly one map

$$A_0 \to G(B_m).$$

Compare the comments on the definitions of category, functor and natural transformation (Remarks 1.1.2(b), 1.2.2(a), and 1.3.2(a)).

(c) Not only do adjoint functors arise everywhere; better, whenever you see a pair of functors $\mathscr{A} \rightleftarrows \mathscr{B}$, there is an excellent chance that they are adjoint (one way round or the other).

For example, suppose you get talking to a mathematician who tells you that her work involves Lie algebras and associative algebras. You try to object that you don't know what either of those things is, but she carries on talking anyway, explaining that there's a way of turning any Lie algebra into an associative algebra, and also a way of turning any associative algebra into a Lie algebra. At this point, even without knowing what she's talking about, you should bet her that one process is adjoint to the other. This almost always works.

(d) A given functor G may or may not have a left adjoint, but if it does, it is unique up to isomorphism, so we may speak of '*the* left adjoint of G'. The same goes for right adjoints. We prove this later (Example 3.3.13).

You might ask 'what do we gain from knowing that two functors are adjoint?' The uniqueness is a crucial part of the answer. Let us return to the example of (c). It would take you only a few minutes to learn what Lie algebras are, what associative algebras are, and what the standard functor G is that turns an associative algebra into a Lie algebra. What about the functor F in the opposite direction? The description of F that you will find in most algebra books (under 'universal enveloping algebra') takes much longer to understand. However, you can bypass that process completely, just by knowing that F is the left adjoint of G. Since G can have only *one* left adjoint, this characterizes F completely. In a sense, it tells you all you need to know.

Examples 2.1.3 (Algebra: free ⊣ forgetful) Forgetful functors between categories of algebraic structures usually have left adjoints. For instance:

(a) Let k be a field. There is an adjunction

$$\begin{tikzcd}
\mathbf{Vect}_k \arrow[d, shift left=1.5ex, "U"] \arrow[d, phantom, "\dashv"] \\
\mathbf{Set}, \arrow[u, shift left=1.5ex, "F"]
\end{tikzcd}$$

where U is the forgetful functor of Example 1.2.3(b) and F is the free functor of Example 1.2.4(c). Adjointness says that given a set S and a vector space V, a linear map $F(S) \to V$ is essentially the same thing as a function $S \to U(V)$.

We saw this in Example **??**, but let us now check it in detail.

Fix a set S and a vector space V. Given a linear map $g\colon F(S) \to V$, we may define a map of sets $\bar{g}\colon S \to U(V)$ by $\bar{g}(s) = g(s)$ for all $s \in S$. This gives a function

$$\begin{array}{ccc}
\mathbf{Vect}_k(F(S), V) & \to & \mathbf{Set}(S, U(V)) \\
g & \mapsto & \bar{g}.
\end{array}$$

In the other direction, given a map of sets $f\colon S \to U(V)$, we may define a linear map $\bar{f}\colon F(S) \to V$ by $\bar{f}(\sum_{s\in S} \lambda_s s) = \sum_{s\in S} \lambda_s f(s)$ for all formal linear combinations $\sum \lambda_s s \in F(S)$. This gives a function

$$\begin{array}{ccc} \mathbf{Set}(S, U(V)) & \to & \mathbf{Vect}_k(F(S), V) \\ f & \mapsto & \bar{f}. \end{array}$$

These two functions 'bar' are mutually inverse: for any linear map $g \colon F(S) \to V$, we have

$$\bar{\bar{g}}\Bigl(\sum_{s \in S} \lambda_s s\Bigr) = \sum_{s \in S} \lambda_s \bar{g}(s) = \sum_{s \in S} \lambda_s g(s) = g\Bigl(\sum_{s \in S} \lambda_s s\Bigr)$$

for all $\sum \lambda_s s \in F(S)$, so $\bar{\bar{g}} = g$, and for any map of sets $f \colon S \to U(V)$, we have

$$\bar{\bar{f}}(s) = \bar{f}(s) = f(s)$$

for all $s \in S$, so $\bar{\bar{f}} = f$. We therefore have a canonical bijection between $\mathbf{Vect}_k(F(S), V)$ and $\mathbf{Set}(S, U(V))$ for each $S \in \mathbf{Set}$ and $V \in \mathbf{Vect}_k$, as required.

Here we have been careful to distinguish between the vector space V and its underlying set $U(V)$. Very often, though, in category theory as in mathematics at large, the symbol for a forgetful functor is omitted. In this example, that would mean dropping the U and leaving the reader to figure out whether each occurrence of V is intended to denote the vector space itself or its underlying set. We will soon start using such notational shortcuts ourselves.

(b) In the same way, there is an adjunction

$$\begin{tikzcd} \mathbf{Grp} \arrow[d, shift left=2, "U"] \arrow[d, phantom, "\dashv"] \\ \mathbf{Set} \arrow[u, shift left=2, "F"] \end{tikzcd}$$

where F and U are the free and forgetful functors of Examples 1.2.3(a) and 1.2.4(a).

The free group functor is tricky to construct explicitly. In Chapter **??**, we will prove a result (the general adjoint functor theorem) guaranteeing that U and many functors like it all have left adjoints. To some extent, this removes the need to construct F explicitly, as observed in Remark 2.1.2(d). The point can be overstated: for a group theorist, the more descriptions of free groups that are available, the better. Explicit constructions really can be useful. But it is an important general principle that forgetful functors of this type always have left adjoints.

(c) There is an adjunction

$$\begin{tikzcd}\mathbf{Ab} \arrow[d, shift left, "U"] \\ \mathbf{Grp} \arrow[u, shift left, "F"] \arrow[u, phantom, "\dashv"]\end{tikzcd}$$

where U is the inclusion functor of Example 1.2.3(d). If G is a group then $F(G)$ is the **abelianization** G_{ab} of G. This is an abelian quotient group of G, with the property that every map from G to an abelian group factorizes uniquely through G_{ab}:

$$\begin{tikzcd}G \arrow[r, "\eta"] \arrow[dr, "\forall\phi"'] & G_{\text{ab}} \arrow[d, dotted, "\exists!\bar{\phi}"] \\ & \forall A.\end{tikzcd}$$

Here η is the natural map from G to its quotient G_{ab}, and A is any abelian group. (We have adopted the abuse of notation advertised in example (a), omitting the symbol U at several places in this diagram.) The bijection

$$\mathbf{Ab}(G_{\text{ab}}, A) \cong \mathbf{Grp}(G, U(A))$$

is given in the left-to-right direction by $\psi \mapsto \psi \circ \eta$, and in the right-to-left direction by $\phi \mapsto \bar{\phi}$.

(To construct G_{ab}, let G' be the smallest normal subgroup of G containing $xyx^{-1}y^{-1}$ for all $x, y \in G$, and put $G_{\text{ab}} = G/G'$. The kernel of any homomorphism from G to an abelian group contains G', and the universal property follows.)

(d) There are adjunctions

$$\begin{tikzcd}\mathbf{Grp} \arrow[d, "U"] \\ \mathbf{Mon} \arrow[u, shift left=3, "F"] \arrow[u, shift right=3, "R"'] \arrow[u, phantom, shift left=1.5, "\dashv"] \arrow[u, phantom, shift right=1.5, "\dashv"]\end{tikzcd}$$

between the categories of groups and monoids. The middle functor U is inclusion. The left adjoint F is, again, tricky to describe explicitly. Informally, $F(M)$ is obtained from M by throwing in an inverse to every element. (For example, if M is the additive monoid of natural numbers then $F(M)$ is the group of integers.) Again, the general adjoint functor theorem (Theorem **??**) guarantees the existence of this adjoint.

This example is unusual in that forgetful functors do not usually have *right* adjoints. Here, given a monoid M, the group $R(M)$ is the submonoid of M consisting of all the invertible elements.

The category **Grp** is both a **reflective** and a **coreflective** subcategory of **Mon**. This means, by definition, that the inclusion functor **Grp** $\hookrightarrow$ **Mon** has both a left and a right adjoint. The previous example tells us that **Ab** is a reflective subcategory of **Grp**.

(e) Let **Field** be the category of fields, with ring homomorphisms as the maps. The forgetful functor **Field** $\rightarrow$ **Set** does *not* have a left adjoint. (For a proof, see Example **??**.) The theory of fields is unlike the theories of groups, rings, and so on, because the operation $x \mapsto x^{-1}$ is not defined for *all* x (only for $x \neq 0$).

Remark 2.1.4 At several points in this book, we make contact with the idea of an **algebraic theory**. You already know several examples: the theory of groups is an algebraic theory, as are the theory of rings, the theory of vector spaces over $\mathbb{R}$, the theory of vector spaces over $\mathbb{C}$, the theory of monoids, and (rather trivially) the theory of sets. After reading the description below, you might conclude that the word 'theory' is overly grand, and that 'definition' would be more appropriate. Nevertheless, this is the established usage.

We will not need to define 'algebraic theory' formally, but it will be important to have the general idea. Let us begin by considering the theory of groups.

A group can be defined as a set X equipped with a function $\cdot \colon X \times X \rightarrow X$ (multiplication), another function $(\ )^{-1} \colon X \rightarrow X$ (inverse), and an element $e \in X$ (the identity), satisfying a familiar list of equations. More systematically, the three pieces of structure on X can be seen as maps of sets

$$\cdot \colon X^2 \rightarrow X, \qquad (\ )^{-1} \colon X^1 \rightarrow X, \qquad e \colon X^0 \rightarrow X,$$

where in the last case, X^0 is the one-element set 1 and we are using the observation that a map $1 \rightarrow X$ of sets is essentially the same thing as an element of X.

(You may be more familiar with a definition of group in which only the multiplication and perhaps the identity are specified as pieces of *structure*, with the existence of inverses required as a *property*. In that approach, the definition is swiftly followed by a lemma on uniqueness of inverses, guaranteeing that it makes sense to speak of *the* inverse of an element. The two approaches are equivalent, but for many purposes, it is better to frame the definition in the way described in the previous paragraph.)

An algebraic theory consists of two things: first, a collection of operations, each with a specified arity (number of inputs), and second, a collection of equations. For example, the theory of groups has one operation of arity 2, one of arity 1, and one of arity 0. An **algebra** or **model** for an algebraic theory consists of a set X together with a specified map $X^n \rightarrow X$ for each operation of arity n, such that the equations hold everywhere. For example, an algebra for the theory of groups is exactly a group.

A more subtle example is the theory of vector spaces over $\mathbb{R}$. This is an algebraic theory with, among other things, an infinite number of operations of arity 1: for each $\lambda \in \mathbb{R}$, we have the operation $\lambda \cdot - \colon X \to X$ of scalar multiplication by λ (for any vector space X). There is nothing special about the field $\mathbb{R}$ here; the only point is that it was chosen in advance. The theory of vector spaces over $\mathbb{R}$ is different from the theory of vector spaces over $\mathbb{C}$, because they have different operations of arity 1.

In a nutshell, the main property of algebras for an algebraic theory is that the operations are defined everywhere on the set, and the equations hold everywhere too. For example, *every* element of a group has a specified inverse, and *every* element x satisfies the equation $x \cdot x^{-1} = 1$. This is why the theories of groups, rings, and so on, are algebraic theories, but the theory of fields is not.

Example 2.1.5 There are adjunctions

$$\begin{tikzcd}
\mathbf{Top} \arrow[d, "{\dashv}\ U\ {\dashv}" description] \\
\mathbf{Set} \arrow[u, shift left=6, "D"] \arrow[u, shift right=6, "I"']
\end{tikzcd}$$

where U sends a space to its set of points, D equips a set with the discrete topology, and I equips a set with the indiscrete topology.

Example 2.1.6 Given sets A and B, we can form their (cartesian) product $A \times B$. We can also form the set B^A of functions from A to B. This is the same as the set $\mathbf{Set}(A, B)$, but we tend to use the notation B^A when we want to emphasize that it is an object of the same category as A and B.

Now fix a set B. Taking the product with B defines a functor

$$\begin{array}{rccc} - \times B\colon & \mathbf{Set} & \to & \mathbf{Set} \\ & A & \mapsto & A \times B. \end{array}$$

(Here we are using the blank notation introduced in Example 1.2.12.) There is also a functor

$$\begin{array}{rccc} (-)^B\colon & \mathbf{Set} & \to & \mathbf{Set} \\ & C & \mapsto & C^B. \end{array}$$

Moreover, there is a canonical bijection

$$\mathbf{Set}(A \times B, C) \cong \mathbf{Set}(A, C^B)$$

for any sets A and C. It is defined by simply changing the punctuation: given a

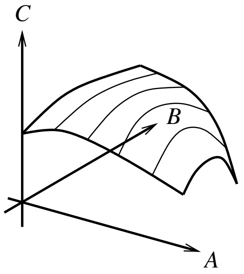


Figure 2.1 In **Set**, a map $A \times B \to C$ can be seen as a way of assigning to each element of A a map $B \to C$.

map $g\colon A \times B \to C$, define $\bar{g}\colon A \to C^B$ by

$$(\bar{g}(a))(b) = g(a, b)$$

($a \in A$, $b \in B$), and in the other direction, given $f\colon A \to C^B$, define $\bar{f}\colon A \times B \to C$ by

$$\bar{f}(a, b) = (f(a))(b)$$

($a \in A$, $b \in B$). Figure 2.1 shows an example with $A = B = C = \mathbb{R}$. By slicing up the surface as shown, a map $\mathbb{R}^2 \to \mathbb{R}$ can be seen as a map from $\mathbb{R}$ to {maps $\mathbb{R} \to \mathbb{R}$}.

Putting all this together, we obtain an adjunction

$$\begin{tikzcd} \mathbf{Set} \arrow[d, shift left=2, "(-)^B"] \arrow[d, phantom, "\dashv"] \\ \mathbf{Set} \arrow[u, shift left=2, "-\times B"] \end{tikzcd}$$

for every set B.

Definition 2.1.7 Let $\mathscr{A}$ be a category. An object $I \in \mathscr{A}$ is **initial** if for every $A \in \mathscr{A}$, there is exactly one map $I \to A$. An object $T \in \mathscr{A}$ is **terminal** if for every $A \in \mathscr{A}$, there is exactly one map $A \to T$.

For example, the empty set is initial in **Set**, the trivial group is initial in **Grp**, and $\mathbb{Z}$ is initial in **Ring** (Example **??**). The one-element set is terminal in **Set**, the trivial group is terminal (as well as initial) in **Grp**, and the trivial (one-element) ring is terminal in **Ring**. The terminal object of **CAT** is the category **1** containing just one object and one map (necessarily the identity on that object).

A category need not have an initial object, but if it does have one, it is unique up to isomorphism. Indeed, it is unique up to *unique* isomorphism, as follows.

Lemma 2.1.8 *Let I and I' be initial objects of a category. Then there is a unique isomorphism $I \to I'$. In particular, $I \cong I'$.*

Proof Since I is initial, there is a unique map $f \colon I \to I'$. Since I' is initial, there is a unique map $f' \colon I' \to I$. Now $f' \circ f$ and 1_I are both maps $I \to I$, and I is initial, so $f' \circ f = 1_I$. Similarly, $f \circ f' = 1_{I'}$. Hence f is an isomorphism, as required. □

Example 2.1.9 Initial and terminal objects can be described as adjoints. Let $\mathscr{A}$ be a category. There is precisely one functor $\mathscr{A} \to \mathbf{1}$. Also, a functor $\mathbf{1} \to \mathscr{A}$ is essentially just an object of $\mathscr{A}$ (namely, the object to which the unique object of $\mathbf{1}$ is mapped). Viewing functors $\mathbf{1} \to \mathscr{A}$ as objects of $\mathscr{A}$, a left adjoint to $\mathscr{A} \to \mathbf{1}$ is exactly an initial object of $\mathscr{A}$.

Similarly, a right adjoint to the unique functor $\mathscr{A} \to \mathbf{1}$ is exactly a terminal object of $\mathscr{A}$.

Remark 2.1.10 In the language introduced in Remark 1.1.10, the concept of terminal object is dual to the concept of initial object. (More generally, the concepts of left and right adjoint are dual to one another.) Since any two initial objects of a category are uniquely isomorphic, the principle of duality implies that the same is true of terminal objects.

Remark 2.1.11 Adjunctions can be composed. Take adjunctions

$$\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{A}' \arrow[l, shift left, "G"] \arrow[r, shift left, "F'"] \arrow[r, phantom, "\perp"] & \mathscr{A}'' \arrow[l, shift left, "G'"] \end{tikzcd}$$

where the $\perp$ symbol is a rotated $\dashv$ (thus, $F \dashv G$ and $F' \dashv G'$). Then we obtain an adjunction

$$\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F' \circ F"] \arrow[r, phantom, "\perp"] & \mathscr{A}'' \arrow[l, shift left, "G \circ G'"] \end{tikzcd},$$

since for $A \in \mathscr{A}$ and $A'' \in \mathscr{A}''$,

$$\mathscr{A}''(F'(F(A)), A'') \cong \mathscr{A}'(F(A), G'(A'')) \cong \mathscr{A}(A, G(G'(A'')))$$

naturally in A and A''.

Exercises

2.1.12 Find three examples of adjoint functors not mentioned above. Do the same for initial and terminal objects.

2.1.13 What can be said about adjunctions between discrete categories?

2.1.14 Show that the naturality equations (2.2) and (2.3) can equivalently be replaced by the single equation

$$\overline{\left(A' \xrightarrow{p} A \xrightarrow{f} G(B) \xrightarrow{G(q)} G(B')\right)} \quad = \quad \left(F(A') \xrightarrow{F(p)} F(A) \xrightarrow{\bar{f}} B \xrightarrow{q} B'\right)$$

for all p, f and q.

2.1.15 Show that left adjoints preserve initial objects: that is, if $\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd}$ and I is an initial object of $\mathscr{A}$, then $F(I)$ is an initial object of $\mathscr{B}$. Dually, show that right adjoints preserve terminal objects.

(In Section **??**, we will see this as part of a bigger picture: right adjoints preserve limits and left adjoints preserve colimits.)

2.1.16 Let G be a group.

(a) What interesting functors are there (in either direction) between **Set** and the category $[G, \mathbf{Set}]$ of left G-sets? Which of those functors are adjoint to which?

(b) Similarly, what interesting functors are there between $\mathbf{Vect}_k$ and the category $[G, \mathbf{Vect}_k]$ of k-linear representations of G, and what adjunctions are there between those functors?

2.1.17 Fix a topological space X, and write $\mathscr{O}(X)$ for the poset of open subsets of X, ordered by inclusion. Let

$$\Delta \colon \mathbf{Set} \to [\mathscr{O}(X)^{\mathrm{op}}, \mathbf{Set}]$$

be the functor assigning to a set A the presheaf ΔA with constant value A. Exhibit a chain of adjoint functors

$$\Lambda \dashv \Pi \dashv \Delta \dashv \Gamma \dashv \nabla.$$

2.2 Adjunctions via units and counits

In the previous section, we met the definition of adjunction. In this section and the next, we meet two ways of rephrasing the definition. The one in this section is most useful for theoretical purposes, while the one in the next fits well with many examples.

To start building the theory of adjoint functors, we have to take seriously the naturality requirement (equations (2.2) and (2.3)), which has so far been ignored. Take an adjunction $\begin{tikzcd}\mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, shift left, "G"]\end{tikzcd}$. Intuitively, naturality says that as A varies in $\mathscr{A}$ and B varies in $\mathscr{B}$, the isomorphism between $\mathscr{B}(F(A), B)$ and $\mathscr{A}(A, G(B))$ varies in a way that is compatible with all the structure already in place. In other words, it is compatible with composition in the categories $\mathscr{A}$ and $\mathscr{B}$ and the action of the functors F and G.

But what does 'compatible' mean? Suppose, for example, that we have maps

$$F(A) \xrightarrow{g} B \xrightarrow{q} B'$$

in $\mathscr{B}$. There are two things we can do with this data: either compose then take the transpose, which produces a map $\overline{q \circ g}\colon A \to G(B')$, or take the transpose of g then compose it with $G(q)$, which produces a potentially different map $G(q) \circ \bar{g}\colon A \to G(B')$. Compatibility means that they are equal; and that is the first naturality equation (2.2). The second is its dual, and can be explained in a similar way.

For each $A \in \mathscr{A}$, we have a map

$$\left(A \xrightarrow{\eta_A} GF(A)\right) = \overline{\left(F(A) \xrightarrow{1} F(A)\right)}.$$

Dually, for each $B \in \mathscr{B}$, we have a map

$$\left(FG(B) \xrightarrow{\varepsilon_B} B\right) = \overline{\left(G(B) \xrightarrow{1} G(B)\right)}.$$

(We have begun to omit brackets, writing $GF(A)$ instead of $G(F(A))$, etc.) These define natural transformations

$$\eta\colon 1_{\mathscr{A}} \to G \circ F, \qquad \varepsilon\colon F \circ G \to 1_{\mathscr{B}},$$

called the **unit** and **counit** of the adjunction, respectively.

Example 2.2.1 Take the usual adjunction $\begin{tikzcd}\mathbf{Vect}_k \arrow[r, shift left, "U"] \arrow[r, phantom, "\top"] & \mathbf{Set} \arrow[l, shift left, "F"]\end{tikzcd}$. Its unit $\eta\colon 1_{\mathbf{Set}} \to U \circ F$ has components

$$\begin{array}{rccl} \eta_S\colon & S & \to & UF(S) = \{\text{formal } k\text{-linear sums } \textstyle\sum_{s \in S} \lambda_s s\} \\ & s & \mapsto & s \end{array}$$

($S \in \mathbf{Set}$). The component of the counit ε at a vector space V is the linear map

$$\varepsilon_V\colon FU(V) \to V$$

that sends a *formal* linear sum $\sum_{v \in V} \lambda_v v$ to its *actual* value in V.

The vector space $FU(V)$ is enormous. For instance, if $k = \mathbb{R}$ and V is the vector space $\mathbb{R}^2$, then $U(V)$ is the set $\mathbb{R}^2$ and $FU(V)$ is a vector space with one basis element for every element of $\mathbb{R}^2$; thus, it is uncountably infinite-dimensional. Then ε_V is a map from this infinite-dimensional space to the 2-dimensional space V.

Lemma 2.2.2 *Given an adjunction $F \dashv G$ with unit η and counit ε, the triangles*

$$\begin{tikzcd}
F \arrow[r, "F\eta"] \arrow[dr, "1_F"'] & FGF \arrow[d, "\varepsilon F"] \\
& F
\end{tikzcd}
\qquad
\begin{tikzcd}
G \arrow[r, "\eta G"] \arrow[dr, "1_G"'] & GFG \arrow[d, "G\varepsilon"] \\
& G
\end{tikzcd}$$

commute.

Remark 2.2.3 These are called the **triangle identities**. They are commutative diagrams in the functor categories $[\mathscr{A}, \mathscr{B}]$ and $[\mathscr{B}, \mathscr{A}]$, respectively. For an explanation of the notation, see Remarks 1.3.24 (particularly the special cases mentioned on page 29). An equivalent statement is that the triangles

$$\begin{tikzcd}
F(A) \arrow[r, "F(\eta_A)"] \arrow[dr, "1_{F(A)}"'] & FGF(A) \arrow[d, "\varepsilon_{F(A)}"] \\
& F(A)
\end{tikzcd}
\qquad
\begin{tikzcd}
G(B) \arrow[r, "\eta_{G(B)}"] \arrow[dr, "1_{G(B)}"'] & GFG(B) \arrow[d, "G(\varepsilon_B)"] \\
& G(B)
\end{tikzcd}
\tag{2.4}$$

commute for all $A \in \mathscr{A}$ and $B \in \mathscr{B}$.

Proof of Lemma 2.2.2 We prove that the triangles (2.4) commute. Let $A \in \mathscr{A}$. Since $\overline{1_{GF(A)}} = \varepsilon_{F(A)}$, equation (2.3) gives

$$\overline{\left(A \xrightarrow{\eta_A} GF(A) \xrightarrow{1} GF(A)\right)} = \left(F(A) \xrightarrow{F(\eta_A)} FGF(A) \xrightarrow{\varepsilon_{F(A)}} F(A)\right).$$

But the left-hand side is $\overline{\eta_A} = \overline{\overline{1_{F(A)}}} = 1_{F(A)}$, proving the first identity. The second follows by duality. □

Amazingly, the unit and counit determine the whole adjunction, even though they appear to know only the transposes *of identities*. This is the main content of the following pair of results.

Lemma 2.2.4 *Let* $\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd}$ *be an adjunction, with unit η and counit ε. Then*

$$\bar{g} = G(g) \circ \eta_A$$

for any $g\colon F(A) \to B$, *and*

$$\bar{f} = \varepsilon_B \circ F(f)$$

for any $f\colon A \to G(B)$.

Proof For any map $g\colon F(A) \to B$, we have by equation (2.2), giving the first statement. The second follows by duality. □

Theorem 2.2.5 *Take categories and functors* $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$. *There is a one-to-one correspondence between:*

(a) *adjunctions between F and G (with F on the left and G on the right);*
(b) *pairs* $\left(1_{\mathscr{A}} \xrightarrow{\eta} GF,\ FG \xrightarrow{\varepsilon} 1_{\mathscr{B}}\right)$ *of natural transformations satisfying the triangle identities.*

(Recall that by definition, an adjunction between F and G is a choice of isomorphism (2.1) for each A and B, satisfying the naturality equations (2.2) and (2.3).)

Proof We have shown that every adjunction between F and G gives rise to a pair (η, ε) satisfying the triangle identities. We now have to show that this process is bijective. So, take a pair (η, ε) of natural transformations satisfying the triangle identities. We must show that there is a unique adjunction between F and G with unit η and counit ε.

Uniqueness follows from Lemma 2.2.4. For existence, take natural transformations η and ε as in (b). For each A and B, define functions

$$\mathscr{B}(F(A), B) \rightleftarrows \mathscr{A}(A, G(B)), \tag{2.5}$$

both denoted by a bar, as follows. Given $g \in \mathscr{B}(F(A), B)$, put $\bar{g} = G(g) \circ \eta_A \in \mathscr{A}(A, G(B))$. Similarly, in the opposite direction, put $\bar{f} = \varepsilon_B \circ F(f)$.

I claim that for each A and B, the two functions $g \mapsto \bar{g}$ and $f \mapsto \bar{f}$ are mutually inverse. Indeed, given a map $g\colon F(A) \to B$ in $\mathscr{B}$, we have a commutative diagram

$$\begin{tikzcd}
F(A) \arrow[r, "F(\eta_A)"] \arrow[dr, "1"'] & FGF(A) \arrow[r, "FG(g)"] \arrow[d, "\varepsilon_{F(A)}"] & FG(B) \arrow[d, "\varepsilon_B"] \\
& F(A) \arrow[r, "g"'] & B.
\end{tikzcd}$$

The composite map from $F(A)$ to B by one route around the outside of the diagram is

$$\varepsilon_B \circ FG(g) \circ F(\eta_A) = \varepsilon_B \circ F(\bar{g}) = \bar{\bar{g}},$$

and by the other is $g \circ 1 = g$, so $\bar{\bar{g}} = g$. Dually, $\bar{\bar{f}} = f$ for any map $f : A \to G(B)$ in $\mathscr{A}$. This proves the claim.

It is straightforward to check the naturality equations (2.2) and (2.3). The functions (2.5) therefore define an adjunction. Finally, its unit and counit are η and ε, since the component of the unit at A is

$$\overline{1_{F(A)}} = G(1_{F(A)}) \circ \eta_A = 1 \circ \eta_A = \eta_A,$$

and dually for the counit. □

Corollary 2.2.6 *Take categories and functors* $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$. *Then* $F \dashv G$ *if and only if there exist natural transformations* $1 \xrightarrow{\eta} GF$ *and* $FG \xrightarrow{\varepsilon} 1$ *satisfying the triangle identities.* □

Example 2.2.7 An adjunction between ordered sets consists of order-preserving maps $A \underset{g}{\overset{f}{\rightleftarrows}} B$ such that

$$\forall a \in A, \forall b \in B, \qquad f(a) \leq b \iff a \leq g(b). \tag{2.6}$$

This is because both sides of the isomorphism (2.1) in the definition of adjunction are sets with at most one element, so they are isomorphic if and only if they are both empty or both nonempty. The naturality requirements (2.2) and (2.3) hold automatically, since in an ordered set, any two maps with the same domain and codomain are equal.

Recall from Example 1.3.9 that if $C \underset{q}{\overset{p}{\rightrightarrows}} D$ are order-preserving maps of ordered sets then there is at most one natural transformation from p to q, and there is one if and only if $p(c) \leq q(c)$ for all $c \in C$. The unit of the adjunction above is the statement that $a \leq gf(a)$ for all $a \in A$, and the counit is the statement that $fg(b) \leq b$ for all $b \in B$. The triangle identities say nothing, since they assert the equality of two maps in an ordered set with the same domain and codomain.

In the case of ordered sets, Corollary 2.2.6 states that condition (2.6) is equivalent to:

$$\forall a \in A,\ a \leq gf(a) \qquad \text{and} \qquad \forall b \in B,\ fg(b) \leq b.$$

This equivalence can also be proved directly (Exercise 2.2.10).

For instance, let X be a topological space. Take the set $\mathscr{C}(X)$ of closed subsets of X and the set $\mathscr{P}(X)$ of all subsets of X, both ordered by $\subseteq$. There are order-preserving maps

$$\begin{tikzcd}\mathscr{P}(X) \arrow[r, shift left, "\mathrm{Cl}"] & \mathscr{C}(X) \arrow[l, shift left, "i"]\end{tikzcd}$$

where i is the inclusion map and Cl is closure. This is an adjunction, with Cl left adjoint to i, as witnessed by the fact that

$$\mathrm{Cl}(A) \subseteq B \iff A \subseteq B$$

for all $A \subseteq X$ and closed $B \subseteq X$. An equivalent statement is that $A \subseteq \mathrm{Cl}(A)$ for all $A \subseteq X$ and $\mathrm{Cl}(B) \subseteq B$ for all closed $B \subseteq X$. Either way, we see that the topological operation of closure arises as an adjoint functor.

Remark 2.2.8 Theorem 2.2.5 states that an adjunction may be regarded as a quadruple $(F, G, \eta, \varepsilon)$ of functors and natural transformations satisfying the triangle identities. An equivalence $(F, G, \eta, \varepsilon)$ of categories (as in Definition 1.3.15) is not necessarily an adjunction. It *is* true that F is left adjoint to G (Exercise 2.3.10), but η and ε are not necessarily the unit and counit (because there is no reason why they should satisfy the triangle identities).

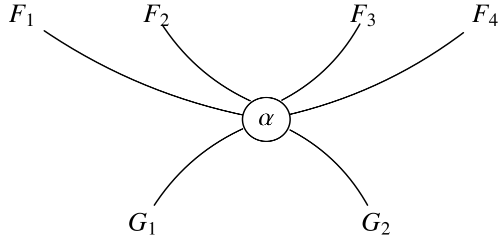
Remark 2.2.9 There is a way of drawing natural transformations that makes the triangle identities intuitively plausible. Suppose, for instance, that we have categories and functors

$$\mathscr{A} \xrightarrow{F_1} \mathscr{C}_1 \xrightarrow{F_2} \mathscr{C}_2 \xrightarrow{F_3} \mathscr{C}_3 \xrightarrow{F_4} \mathscr{B}, \qquad \mathscr{A} \xrightarrow{G_1} \mathscr{D}_1 \xrightarrow{G_2} \mathscr{B}$$

and a natural transformation $\alpha \colon F_4F_3F_2F_1 \to G_2G_1$. We usually draw α like this:

$$\begin{tikzcd}
 & \mathscr{C}_1 \arrow[r, "F_2"] & \mathscr{C}_2 \arrow[r, "F_3"] \arrow[d, phantom, "\Downarrow \alpha"] & \mathscr{C}_3 \arrow[dr, "F_4"] & \\
\mathscr{A} \arrow[ur, "F_1"] \arrow[rr, "G_1"'] & & \mathscr{D}_1 \arrow[rr, "G_2"'] & & \mathscr{B}
\end{tikzcd}$$

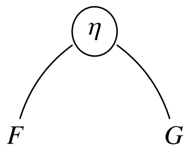
However, we can also draw α as a **string diagram**:



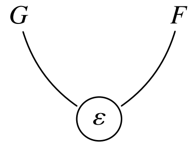
There is nothing special about 4 and 2; we could replace them by any natural numbers m and n. If $m = 0$ then $\mathscr{A} = \mathscr{B}$ and the domain of α is $1_{\mathscr{A}}$ (keeping in mind the last paragraph of Remark 1.1.2(b)). In that case, the disk labelled α has no strings coming into the top. Similarly, if $n = 0$ then there are no strings coming out of the bottom.

Vertical composition of natural transformations corresponds to joining string diagrams together vertically, and horizontal composition corresponds to putting them side by side. The identity on a functor F is drawn as a simple string,

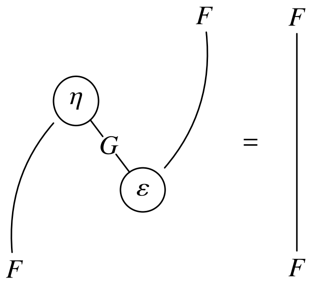
Now let us apply this notation to adjunctions. The unit and counit are drawn as



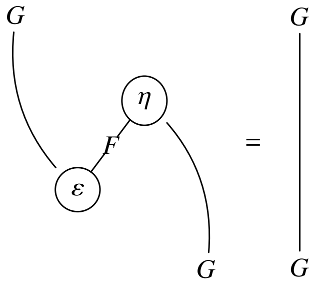
and



The triangle identities now become the topologically plausible equations



and



In both equations, the right-hand side is obtained from the left by simply pulling the string straight.

Exercises

2.2.10 Let $A \underset{g}{\overset{f}{\rightleftarrows}} B$ be order-preserving maps between ordered sets. Prove *directly* that the following conditions are equivalent:

(a) for all $a \in A$ and $b \in B$,

$$f(a) \leq b \iff a \leq g(b);$$

(b) $a \leq g(f(a))$ for all $a \in A$ and $f(g(b)) \leq b$ for all $b \in B$.

(Both conditions state that $f \dashv g$; see Example 2.2.7.)

2.2.11 (a) Let $\begin{tikzcd}\mathscr{A} \arrow[r, bend left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, bend left, "G"]\end{tikzcd}$ be an adjunction with unit η and counit ε. Write $\mathbf{Fix}(GF)$ for the full subcategory of $\mathscr{A}$ whose objects are those $A \in \mathscr{A}$ such that η_A is an isomorphism, and dually $\mathbf{Fix}(FG) \subseteq \mathscr{B}$. Prove that the adjunction $(F, G, \eta, \varepsilon)$ restricts to an equivalence $(F', G', \eta', \varepsilon')$ between $\mathbf{Fix}(GF)$ and $\mathbf{Fix}(FG)$.

(b) Part (a) shows that every adjunction restricts to an equivalence between full subcategories in a canonical way. Take some examples of adjunctions and work out what this equivalence is.

2.2.12 (a) Show that for any adjunction, the right adjoint is full and faithful if and only if the counit is an isomorphism.

(b) An adjunction satisfying the equivalent conditions of part (a) is called a **reflection**. (Compare Example 2.1.3(d).) Of the examples of adjunctions given in this chapter, which are reflections?

2.2.13 (a) Let $f \colon K \to L$ be a map of sets, and denote by $f^* \colon \mathscr{P}(L) \to \mathscr{P}(K)$ the map sending a subset S of L to its inverse image $f^{-1}S \subseteq K$. Then f^* is order-preserving with respect to the inclusion orderings on $\mathscr{P}(K)$ and $\mathscr{P}(L)$, and so can be seen as a functor. Find left and right adjoints to f^*.

(b) Now let X and Y be sets, and write $p \colon X \times Y \to X$ for first projection. Regard a subset S of X as a predicate $S(x)$ in one variable $x \in X$, and similarly a subset R of $X \times Y$ as a predicate $R(x, y)$ in two variables. What, in terms of predicates, are the left and right adjoints to p^*? For each of the adjunctions, interpret the unit and counit as logical implications. (Hint: the left adjoint to p^* is often written as $\exists_Y$, and the right adjoint as $\forall_Y$.)

2.2.14 Given a functor $F \colon \mathscr{A} \to \mathscr{B}$ and a category $\mathscr{S}$, there is a functor $F^* \colon [\mathscr{B}, \mathscr{S}] \to [\mathscr{A}, \mathscr{S}]$ defined on objects $Y \in [\mathscr{B}, \mathscr{S}]$ by $F^*(Y) = Y \circ F$ and on maps α by $F^*(\alpha) = \alpha F$. Show that any adjunction $\begin{tikzcd}\mathscr{A} \arrow[r, bend left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, bend left, "G"]\end{tikzcd}$ and category $\mathscr{S}$ give rise to an adjunction

$$\begin{tikzcd}
{[\mathscr{A}, \mathscr{S}]} \arrow[r, bend left, "G^*"] \arrow[r, phantom, "\perp"] & {[\mathscr{B}, \mathscr{S}]}. \arrow[l, bend left, "F^*"]
\end{tikzcd}$$

(Hint: use Theorem 2.2.5.)

2.3 Adjunctions via initial objects

We now come to the third formulation of adjointness, which is the one you will probably see most often in everyday mathematics.

Consider once more the adjunction

$$\begin{tikzcd}\mathbf{Vect}_k \arrow[d, shift left, "U"] \arrow[d, phantom, "\dashv"] \\ \mathbf{Set}. \arrow[u, shift left, "F"]\end{tikzcd}$$

Let S be a set. The universal property of $F(S)$, the vector space whose basis is S, is most commonly stated like this:

> given a vector space V, any function $f\colon S \to V$ extends uniquely to a linear map $\bar{f}\colon F(S) \to V$.

As remarked in Example 2.1.3(a), forgetful functors are often forgotten: in this statement, '$f\colon S \to V$' should strictly speaking be '$f\colon S \to U(V)$'. Also, the word 'extends' refers implicitly to the embedding

$$\begin{array}{cccc} \eta_S\colon & S & \to & UF(S) \\ & s & \mapsto & s. \end{array}$$

So in precise language, the statement reads:

> for any $V \in \mathbf{Vect}_k$ and $f \in \mathbf{Set}(S, U(V))$, there is a unique $\bar{f} \in \mathbf{Vect}_k(F(S), V)$ such that the diagram
>
> $$\begin{tikzcd} S \arrow[r, "\eta_S"] \arrow[dr, "f"'] & U(F(S)) \arrow[d, "U(\bar{f})"] \\ & U(V) \end{tikzcd} \tag{2.7}$$
>
> commutes.

(Compare Example **??**.) In this section, we show that this statement is equivalent to the statement that F is left adjoint to U with unit η.

To do this, we need a definition.

Definition 2.3.1 Given categories and functors

$$\begin{tikzcd} & \mathscr{B} \arrow[d, "Q"] \\ \mathscr{A} \arrow[r, "P"'] & \mathscr{C}, \end{tikzcd}$$

the **comma category** $(P \Rightarrow Q)$ (often written as $(P \downarrow Q)$) is the category defined as follows:

- objects are triples (A, h, B) with $A \in \mathscr{A}$, $B \in \mathscr{B}$, and $h\colon P(A) \to Q(B)$ in $\mathscr{C}$;
- maps $(A, h, B) \to (A', h', B')$ are pairs $(f\colon A \to A', g\colon B \to B')$ of maps such that the square

$$\begin{tikzcd}
P(A) \arrow[r, "P(f)"] \arrow[d, "h"'] & P(A') \arrow[d, "h'"] \\
Q(B) \arrow[r, "Q(g)"'] & Q(B')
\end{tikzcd}$$

commutes.

Remark 2.3.2 Given $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, P and Q as above, there are canonical functors and a canonical natural transformation as shown:

$$\begin{tikzcd}
(P \Rightarrow Q) \arrow[r] \arrow[d] \arrow[dr, phantom, "\nearrow"] & \mathscr{B} \arrow[d, "Q"] \\
\mathscr{A} \arrow[r, "P"'] & \mathscr{C}
\end{tikzcd}$$

In a suitable 2-categorical sense, $(P \Rightarrow Q)$ is universal with this property.

Example 2.3.3 Let $\mathscr{A}$ be a category and $A \in \mathscr{A}$. The **slice category** of $\mathscr{A}$ over A, denoted by $\mathscr{A}/A$, is the category whose objects are maps into A and whose maps are commutative triangles. More precisely, an object is a pair (X, h) with $X \in \mathscr{A}$ and $h\colon X \to A$ in $\mathscr{A}$, and a map $(X, h) \to (X', h')$ in $\mathscr{A}/A$ is a map $f\colon X \to X'$ in $\mathscr{A}$ making the triangle

$$\begin{tikzcd}
X \arrow[rr, "f"] \arrow[dr, "h"'] & & X' \arrow[dl, "h'"] \\
& A &
\end{tikzcd}$$

commute.

Slice categories are a special case of comma categories. Recall from Example 2.1.9 that functors $\mathbf{1} \to \mathscr{A}$ are just objects of $\mathscr{A}$. Now, given an object A of $\mathscr{A}$, consider the comma category $(1_{\mathscr{A}} \Rightarrow A)$, as in the diagram

$$\begin{tikzcd}
& \mathbf{1} \arrow[d, "A"] \\
\mathscr{A} \arrow[r, "1_{\mathscr{A}}"'] & \mathscr{A}.
\end{tikzcd}$$

An object of $(1_{\mathscr{A}} \Rightarrow A)$ is in principle a triple (X, h, B) with $X \in \mathscr{A}$, $B \in \mathbf{1}$, and $h \colon X \to A$ in $\mathscr{A}$; but $\mathbf{1}$ has only one object, so it is essentially just a pair (X, h). Hence the comma category $(1_{\mathscr{A}} \Rightarrow A)$ has the same objects as the slice category $\mathscr{A}/A$. One can check that it has the same maps too, so that $\mathscr{A}/A \cong (1_{\mathscr{A}} \Rightarrow A)$.

Dually (reversing all the arrows), there is a **coslice category** $A/\mathscr{A} \cong (A \Rightarrow 1_{\mathscr{A}})$, whose objects are the maps out of A.

Example 2.3.4 Let $G \colon \mathscr{B} \to \mathscr{A}$ be a functor and let $A \in \mathscr{A}$. We can form the comma category $(A \Rightarrow G)$, as in the diagram

$$\begin{tikzcd}
 & \mathscr{B} \arrow[d, "G"] \\
\mathbf{1} \arrow[r, "A"'] & \mathscr{A}.
\end{tikzcd}$$

Its objects are pairs $(B \in \mathscr{B}, f \colon A \to G(B))$. A map $(B, f) \to (B', f')$ in $(A \Rightarrow G)$ is a map $q \colon B \to B'$ in $\mathscr{B}$ making the triangle

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[dr, "f'"'] & G(B) \arrow[d, "G(q)"] \\
 & G(B')
\end{tikzcd}$$

commute.

Notice how this diagram resembles the diagram (2.7) in the vector space example. We will use comma categories $(A \Rightarrow G)$ to capture the kind of universal property discussed there.

Speaking casually, we say that $f \colon A \to G(B)$ is an object of $(A \Rightarrow G)$, when what we should really say is that the pair (B, f) is an object of $(A \Rightarrow G)$. There is potential for confusion here, since there may be different objects B, B' of $\mathscr{B}$ with $G(B) = G(B')$. Nevertheless, we will often use this convention.

We now make the connection between comma categories and adjunctions.

Lemma 2.3.5 *Take an adjunction* $\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd}$ *and an object* $A \in \mathscr{A}$. *Then the unit map* $\eta_A \colon A \to GF(A)$ *is an initial object of* $(A \Rightarrow G)$.

Proof Let $(B, f \colon A \to G(B))$ be an object of $(A \Rightarrow G)$. We have to show that there is exactly one map from $(F(A), \eta_A)$ to (B, f).

A map $(F(A), \eta_A) \to (B, f)$ in $(A \Rightarrow G)$ is a map $q \colon F(A) \to B$ in $\mathscr{B}$ such that

$$\begin{tikzcd}
A \arrow[r, "\eta_A"] \arrow[rd, "f"'] & GF(A) \arrow[d, "G(q)"] \\
& G(B)
\end{tikzcd} \tag{2.8}$$

commutes. But $G(q) \circ \eta_A = \bar{q}$ by Lemma 2.2.4, so (2.8) commutes if and only if $f = \bar{q}$, if and only if $q = \bar{f}$. Hence $\bar{f}$ is the unique map $(F(A), \eta_A) \to (B, f)$ in $(A \Rightarrow G)$. □

We now meet our third and final formulation of adjointness.

Theorem 2.3.6 *Take categories and functors* $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$. *There is a one-to-one correspondence between:*

(a) *adjunctions between F and G (with F on the left and G on the right);*
(b) *natural transformations* $\eta\colon 1_{\mathscr{A}} \to GF$ *such that* $\eta_A\colon A \to GF(A)$ *is initial in* $(A \Rightarrow G)$ *for every* $A \in \mathscr{A}$.

Proof We have just shown that every adjunction between F and G gives rise to a natural transformation η with the property stated in (b). To prove the theorem, we have to show that every η with the property in (b) is the unit of exactly one adjunction between F and G.

By Theorem 2.2.5, an adjunction between F and G amounts to a pair (η, ε) of natural transformations satisfying the triangle identities. So it is enough to prove that for every η with the property in (b), there exists a unique natural transformation $\varepsilon\colon FG \to 1_{\mathscr{B}}$ such that the pair (η, ε) satisfies the triangle identities.

Let $\eta\colon 1_{\mathscr{A}} \to GF$ be a natural transformation with the property in (b).

Uniqueness Suppose that $\varepsilon, \varepsilon'\colon FG \to 1_{\mathscr{B}}$ are natural transformations such that both (η, ε) and (η, ε') satisfy the triangle identities. One of the triangle identities states that for all $B \in \mathscr{B}$, the triangle

$$\begin{tikzcd}
G(B) \arrow[r, "\eta_{G(B)}"] \arrow[rd, "1"'] & G(FG(B)) \arrow[d, "G(\varepsilon_B)"] \\
& G(B)
\end{tikzcd} \tag{2.9}$$

commutes. Thus, ε_B is a map

$$\Bigl(FG(B),\ G(B) \xrightarrow{\eta_{G(B)}} G(FG(B))\Bigr) \quad \longrightarrow \quad \Bigl(B,\ G(B) \xrightarrow{1} G(B)\Bigr)$$

in $(G(B) \Rightarrow G)$. The same is true of ε'_B. But $\eta_{G(B)}$ is initial, so there is only one such map, so $\varepsilon_B = \varepsilon'_B$. This holds for all B, so $\varepsilon = \varepsilon'$.

Existence For $B \in \mathscr{B}$, define $\varepsilon_B \colon FG(B) \to B$ to be the unique map

$$(FG(B), \eta_{G(B)}) \to (B, 1_{G(B)})$$

in $(G(B) \Rightarrow G)$. (So by definition of ε_B, triangle (2.9) commutes.) We show that $(\varepsilon_B)_{B\in\mathscr{B}}$ is a natural transformation $FG \to 1$ such that η and ε satisfy the triangle identities.

To prove naturality, take $B \xrightarrow{q} B'$ in $\mathscr{B}$. We have commutative diagrams

$$\begin{tikzcd}
G(B) \arrow[r, "\eta_{G(B)}"] \arrow[dr, "1"] \arrow[ddr, "G(q)" swap] & GFG(B) \arrow[d, "G(\varepsilon_B)"] \\
& G(B) \arrow[d, "G(q)"] \\
& G(B')
\end{tikzcd}
\qquad
\begin{tikzcd}
G(B) \arrow[r, "\eta_{G(B)}"] \arrow[d, "G(q)"] \arrow[ddr, bend right=60, "G(q)" swap] & GFG(B) \arrow[d, "GFG(q)"] \\
G(B') \arrow[r, "\eta_{G(B')}"] \arrow[dr, "1"] & GFG(B') \arrow[d, "G(\varepsilon_{B'})"] \\
& G(B').
\end{tikzcd}$$

So $q \circ \varepsilon_B$ and $\varepsilon_{B'} \circ FG(q)$ are both maps $\eta_{G(B)} \to G(q)$ in $(G(B) \Rightarrow G)$, and since $\eta_{G(B)}$ is initial, they must be equal. This proves naturality of ε with respect to q. Hence ε is a natural transformation.

We have already observed that one of the triangle identities, equation (2.9), holds. The other states that for $A \in \mathscr{A}$,

$$\begin{tikzcd}
F(A) \arrow[r, "F(\eta_A)"] \arrow[dr, "1_{F(A)}" swap] & FGF(A) \arrow[d, "\varepsilon_{F(A)}"] \\
& F(A)
\end{tikzcd}$$

commutes. To prove it, we repeat our previous technique: there are commutative diagrams

$$\begin{tikzcd}
A \arrow[r, "\eta_A"] \arrow[ddr, "\eta_A" swap] & GF(A) \arrow[dd, "G(1_{F(A)})"] \\
& \\
& GF(A)
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "\eta_A"] \arrow[d, "\eta_A"] \arrow[ddr, bend right=60, "\eta_A" swap] & GF(A) \arrow[d, "GF(\eta_A)"] \\
GF(A) \arrow[r, "\eta_{GF(A)}"] \arrow[dr, "1"] & GFGF(A) \arrow[d, "G(\varepsilon_{F(A)})"] \\
& GF(A),
\end{tikzcd}$$

so by initiality of η_A, we have $\varepsilon_{F(A)} \circ F(\eta_A) = 1_{F(A)}$, as required. □

In Section **??** we will meet the adjoint functor theorems, which state conditions under which a functor is guaranteed to have a left adjoint. The following corollary is the starting point for their proofs.

Corollary 2.3.7 *Let $G\colon \mathscr{B} \to \mathscr{A}$ be a functor. Then G has a left adjoint if and only if for each $A \in \mathscr{A}$, the category $(A \Rightarrow G)$ has an initial object.*

Proof Lemma 2.3.5 proves 'only if'. To prove 'if', let us choose for each $A \in \mathscr{A}$ an initial object of $(A \Rightarrow G)$ and call it $(F(A), \eta_A\colon A \to GF(A))$. (Here $F(A)$ and η_A are just the names we choose to use.) For each map $f\colon A \to A'$ in $\mathscr{A}$, let $F(f)\colon F(A) \to F(A')$ be the unique map such that

$$\begin{tikzcd}
A \arrow[rr, "\eta_A"] \arrow[dr, "f"'] & & G(F(A)) \arrow[dd, "G(F(f))"] \\
& A' \arrow[dr, "\eta_{A'}"'] & \\
& & G(F(A'))
\end{tikzcd}$$

commutes (in other words, the unique map $\eta_A \to \eta_{A'} \circ f$ in $(A \Rightarrow G)$). It is easily checked that F is a functor $\mathscr{A} \to \mathscr{B}$, and the diagram tells us that η is a natural transformation $1 \to GF$. So by Theorem 2.3.6, F is left adjoint to G. □

This corollary justifies the claim made at the beginning of the section: that given functors F and G, to have an adjunction $F \dashv G$ amounts to having maps $\eta_A\colon A \to GF(A)$ with the universal property stated there.

Exercises

2.3.8 What can be said about adjunctions between groups (regarded as one-object categories)?

2.3.9 State the dual of Corollary 2.3.7. How would you prove your dual statement?

2.3.10 Let $(F, G, \eta, \varepsilon)$ be an equivalence of categories, as in Definition 1.3.15. Prove that F is left adjoint to G (heeding the warning in Remark 2.2.8).

2.3.11 Let $\mathscr{A} \underset{F}{\overset{U}{\rightleftarrows}} \mathbf{Set}$ (with $\top$) be an adjunction. Suppose that for at least one $A \in \mathscr{A}$, the set $U(A)$ has at least two elements. Prove that for each set S, the unit map $\eta_S\colon S \to UF(S)$ is injective. What does this mean in the case of the usual adjunction between **Grp** and **Set**?

2.3.12 Given sets A and B, a **partial function** from A to B is a pair (S, f) consisting of a subset $S \subseteq A$ and a function $S \to B$. (Think of it as like a function from A to B, but undefined at certain elements of A.) Let **Par** be the category of sets and partial functions.

Show that **Par** is equivalent to $\mathbf{Set}_*$, the category of sets equipped with a distinguished element and functions preserving distinguished elements. Show also that $\mathbf{Set}_*$ can be described as a coslice category in a simple way.

3

Representables

A category is a world of objects, all looking at one another. Each sees the world from a different viewpoint.

Consider, for instance, the category of topological spaces, and let us ask how it looks when viewed from the one-point space 1. A map from 1 to a space X is essentially the same thing as a point of X, so we might say that 1 'sees points'. Similarly, a map from $\mathbb{R}$ to a space X could reasonably be called a curve in X, and in this sense, $\mathbb{R}$ sees curves.

Now consider the category of groups. A map from the infinite cyclic group $\mathbb{Z}$ to a group G amounts to an element of G. (For given $g \in G$, there is a unique homomorphism $\phi\colon \mathbb{Z} \to G$ such that $\phi(1) = g$.) So, $\mathbb{Z}$ sees elements. Similarly, if p is a prime number then the cyclic group $\mathbb{Z}/p\mathbb{Z}$ sees elements of order 1 or p.

Any ring homomorphism between fields is injective, so in the category of fields, a map $K \to L$ is a way of realizing L as an extension of K. Hence each field K sees the extensions of itself. If K and L are fields of different characteristic then there are no homomorphisms between K and L, so the category of fields is the union of disjoint subcategories $\mathbf{Field}_0$, $\mathbf{Field}_2$, $\mathbf{Field}_3$, $\mathbf{Field}_5$, … consisting of the fields of characteristics $0, 2, 3, 5, \ldots$. Each field is blind to the fields of different characteristic.

In the ordered set $(\mathbb{R}, \leq)$, the object 0 sees whether a number is nonnegative. In other words, if x is nonnegative then there is one map $0 \to x$, and if not, there are none.

We can also ask the dual question: fixing an object of a category, what are the maps *into* it? Let S be the two-element set, for instance. For an arbitrary set X, the maps from X to S correspond to the subsets of X (as we saw in Section **??**). Now give S the topology in which one of the singleton subsets is open but the other is not. For any topological space X, the continuous maps from X into S correspond to the *open* subsets of X.

This chapter explores the theme of how each object sees and is seen by the category in which it lives. We are naturally led to the notion of representable functor, which (after adjunctions) provides our second approach to the idea of universal property.

3.1 Definitions and examples

Fix an object A of a category $\mathscr{A}$. We will consider the totality of maps out of A. To each $B \in \mathscr{A}$, there is assigned the set (or class) $\mathscr{A}(A, B)$ of maps from A to B. The content of the following definition is that this assignation is functorial in B: any map $B \to B'$ induces a function $\mathscr{A}(A, B) \to \mathscr{A}(A, B')$.

Definition 3.1.1 Let $\mathscr{A}$ be a locally small category and $A \in \mathscr{A}$. We define a functor

$$H^A = \mathscr{A}(A, -) \colon \mathscr{A} \to \mathbf{Set}$$

as follows:

- for objects $B \in \mathscr{A}$, put $H^A(B) = \mathscr{A}(A, B)$;
- for maps $B \xrightarrow{g} B'$ in $\mathscr{A}$, define

  $$H^A(g) = \mathscr{A}(A, g) \colon \mathscr{A}(A, B) \to \mathscr{A}(A, B')$$

  by

  $$p \mapsto g \circ p$$

  for all $p \colon A \to B$.

Remarks 3.1.2 (a) Recall that 'locally small' means that each class $\mathscr{A}(A, B)$ is in fact a set. This hypothesis is clearly necessary in order for the definition to make sense.

(b) Sometimes $H^A(g)$ is written as $g \circ -$ or g_*. All three forms, as well as $\mathscr{A}(A, g)$, are in use.

Definition 3.1.3 Let $\mathscr{A}$ be a locally small category. A functor $X \colon \mathscr{A} \to \mathbf{Set}$ is **representable** if $X \cong H^A$ for some $A \in \mathscr{A}$. A **representation** of X is a choice of an object $A \in \mathscr{A}$ and an isomorphism between H^A and X.

Representable functors are sometimes just called 'representables'. Only set-valued functors (that is, functors with codomain **Set**) can be representable.

Example 3.1.4 Consider $H^1 \colon \mathbf{Set} \to \mathbf{Set}$, where 1 is the one-element set. Since a map from 1 to a set B amounts to an element of B, we have

$$H^1(B) \cong B$$

for each $B \in \mathbf{Set}$. It is easily verified that this isomorphism is natural in B, so H^1 is isomorphic to the identity functor $1_{\mathbf{Set}}$. Hence $1_{\mathbf{Set}}$ is representable.

Example 3.1.5 All of the 'seeing' functors in the introduction to this chapter are representable. The forgetful functor $\mathbf{Top} \to \mathbf{Set}$ is isomorphic to $H^1 = \mathbf{Top}(1, -)$, and the forgetful functor $\mathbf{Grp} \to \mathbf{Set}$ is isomorphic to $\mathbf{Grp}(\mathbb{Z}, -)$. For each prime p, there is a functor $U_p \colon \mathbf{Grp} \to \mathbf{Set}$ defined on objects by

$$U_p(G) = \{\text{elements of } G \text{ of order 1 or } p\},$$

and as claimed above, $U_p \cong \mathbf{Grp}(\mathbb{Z}/p\mathbb{Z}, -)$ (Exercise 3.1.28). Hence U_p is representable.

Example 3.1.6 There is a functor ob : $\mathbf{Cat} \to \mathbf{Set}$ sending a small category to its set of objects. (The category **Cat** was introduced in Definition **??**.) It is representable. Indeed, consider the terminal category **1** (with one object and only the identity map). A functor from **1** to a category $\mathscr{B}$ simply picks out an object of $\mathscr{B}$. Thus,

$$H^{\mathbf{1}}(\mathscr{B}) \cong \operatorname{ob}\mathscr{B}.$$

Again, it is easily verified that this isomorphism is natural in $\mathscr{B}$; hence ob $\cong \mathbf{Cat}(\mathbf{1}, -)$. It can be shown similarly that the functor $\mathbf{Cat} \to \mathbf{Set}$ sending a small category to its set of maps is representable (Exercise 3.1.31).

Example 3.1.7 Let M be a monoid, regarded as a one-object category. Recall from Example 1.2.8 that a set-valued functor on M is just an M-set. Since the category M has only one object, there is only one representable functor on it (up to isomorphism). As an M-set, the unique representable is the so-called **left regular representation** of M, that is, the underlying set of M acted on by multiplication on the left.

Example 3.1.8 Let $\mathbf{Toph}_*$ be the category whose objects are topological spaces equipped with a basepoint and whose arrows are homotopy classes of basepoint-preserving continuous maps. Let $S^1 \in \mathbf{Toph}_*$ be the circle. Then for any object $X \in \mathbf{Toph}_*$, the maps $S^1 \to X$ in $\mathbf{Toph}_*$ are the elements of the fundamental group $\pi_1(X)$. Formally, this says that the composite functor

$$\mathbf{Toph}_* \xrightarrow{\pi_1} \mathbf{Grp} \xrightarrow{U} \mathbf{Set}$$

is isomorphic to $\mathbf{Toph}_*(S^1, -)$. In particular, it is representable.

Example 3.1.9 Fix a field k and vector spaces U and V over k. There is a functor

$$\mathbf{Bilin}(U, V; -) \colon \mathbf{Vect}_k \to \mathbf{Set}$$

whose value $\mathbf{Bilin}(U, V; W)$ at $W \in \mathbf{Vect}_k$ is the set of bilinear maps $U \times V \to W$. It can be shown that this functor is representable; in other words, there is a space T with the property that

$$\mathbf{Bilin}(U, V; W) \cong \mathbf{Vect}_k(T, W)$$

naturally in W. This T is the tensor product $U \otimes V$, which we met just after the proof of Lemma **??**.

Adjunctions give rise to representable functors in the following way.

Lemma 3.1.10 *Let* $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$ *with* $\perp$ *be locally small categories, and let* $A \in \mathscr{A}$. *Then the functor*

$$\mathscr{A}(A, G(-)) \colon \mathscr{B} \to \mathbf{Set}$$

(that is, the composite $\mathscr{B} \xrightarrow{G} \mathscr{A} \xrightarrow{H^A} \mathbf{Set}$*) is representable.*

Proof We have

$$\mathscr{A}(A, G(B)) \cong \mathscr{B}(F(A), B)$$

for each $B \in \mathscr{B}$. If we can show that this isomorphism is natural in B, then we will have proved that $\mathscr{A}(A, G(-))$ is isomorphic to $H^{F(A)}$ and is therefore representable. So, let $B \xrightarrow{q} B'$ be a map in $\mathscr{B}$. We must show that the square

$$\begin{tikzcd}
\mathscr{A}(A, G(B)) \arrow[r] \arrow[d, "G(q)\circ -"'] & \mathscr{B}(F(A), B) \arrow[d, "q\circ -"] \\
\mathscr{A}(A, G(B')) \arrow[r] & \mathscr{B}(F(A), B')
\end{tikzcd}$$

commutes, where the horizontal arrows are the bijections provided by the adjunction. For $f \colon A \to G(B)$, we have

$$\begin{tikzcd}
f \arrow[r, mapsto] \arrow[d, mapsto] & \bar{f} \arrow[d, mapsto] \\
G(q) \circ f \arrow[r, mapsto] & \frac{q \circ \bar{f}}{\overline{G(q) \circ f}},
\end{tikzcd}$$

so we must prove that $q \circ \bar{f} = \overline{G(q) \circ f}$. This follows immediately from the naturality condition (2.2) in the definition of adjunction (with $g = \bar{f}$). □

You would not expect a randomly-chosen functor into **Set** to be representable. In some sense, rather few functors are. However, forgetful functors do tend to be representable:

Proposition 3.1.11 *Any set-valued functor with a left adjoint is representable.*

Proof Let $G\colon \mathscr{A} \to \mathbf{Set}$ be a functor with a left adjoint F. Write 1 for the one-point set. Then

$$G(A) \cong \mathbf{Set}(1, G(A))$$

naturally in $A \in \mathscr{A}$ (by Example 3.1.4), that is, $G \cong \mathbf{Set}(1, G(-))$. So by Lemma 3.1.10, G is representable; indeed, $G \cong H^{F(1)}$. □

Example 3.1.12 Several of the examples of representables mentioned above arise as in Proposition 3.1.11. For instance, $U\colon \mathbf{Top} \to \mathbf{Set}$ has a left adjoint D (Example 2.1.5), and $D(1) \cong 1$, so we recover the result that $U \cong H^1$. Similarly, Exercise **??** asked you to construct a left adjoint D to the objects functor ob : $\mathbf{Cat} \to \mathbf{Set}$. This functor D satisfies $D(1) \cong \mathbf{1}$, proving again that ob $\cong H^{\mathbf{1}}$.

Example 3.1.13 The forgetful functor $U\colon \mathbf{Vect}_k \to \mathbf{Set}$ is representable, since it has a left adjoint. Indeed, if F denotes the left adjoint then $F(1)$ is the 1-dimensional vector space k, so $U \cong H^k$. This is also easy to see directly: a map from k to a vector space V is uniquely determined by the image of 1, which can be any element of V; hence $\mathbf{Vect}_k(k, V) \cong U(V)$ naturally in V.

Example 3.1.14 Examples 2.1.3 began with the declaration that forgetful functors between categories of algebraic structures usually have left adjoints. Take the category **CRing** of commutative rings and the forgetful functor $U\colon \mathbf{CRing} \to \mathbf{Set}$. This general principle suggests that U has a left adjoint, and Proposition 3.1.11 then tells us that U is representable.

Let us see how this works explicitly. Given a set S, let $\mathbb{Z}[S]$ be the ring of polynomials over $\mathbb{Z}$ in commuting variables x_s ($s \in S$). (This was called $F(S)$ in Example 1.2.4(b).) Then $S \mapsto \mathbb{Z}[S]$ defines a functor $\mathbf{Set} \to \mathbf{CRing}$, and this is left adjoint to U. Hence $U \cong H^{\mathbb{Z}[x]}$. Again, this can be verified directly: for any ring R, the maps $\mathbb{Z}[x] \to R$ correspond one-to-one with the elements of R (Exercises **??** and 3.1.29).

We have defined, for each object A of our category $\mathscr{A}$, a functor $H^A \in [\mathscr{A}, \mathbf{Set}]$. This describes how A sees the world. As A varies, the view varies. On the other hand, it is always the same world being seen, so the different views from different objects are somehow related. (Compare aerial photos taken from a moving aeroplane, which agree well enough on their overlaps that they can be patched together to make one big picture.) So the family $(H^A)_{A \in \mathscr{A}}$ of 'views' has some consistency to it. What this means is that whenever there is a map between objects A and A', there is also a map between H^A and $H^{A'}$.

Precisely, a map $A' \xrightarrow{f} A$ induces a natural transformation

$$\begin{tikzcd} \mathscr{A} \arrow[r, bend left, "H^A"{name=U}] \arrow[r, bend right, "H^{A'}"'{name=D}] & \mathbf{Set}, \arrow[from=U, to=D, Rightarrow, "H^f"] \end{tikzcd}$$

whose B-component (for $B \in \mathscr{A}$) is the function

$$\begin{array}{ccc} H^A(B) = \mathscr{A}(A, B) & \to & H^{A'}(B) = \mathscr{A}(A', B) \\ p & \mapsto & p \circ f. \end{array}$$

Again, H^f goes by a variety of other names: $\mathscr{A}(f, -)$, f^*, and $- \circ f$.

Note the reversal of direction! Each functor H^A is covariant, but they come together to form a *contravariant* functor, as in the following definition.

Definition 3.1.15 Let $\mathscr{A}$ be a locally small category. The functor

$$H^\bullet \colon \mathscr{A}^{\mathrm{op}} \to [\mathscr{A}, \mathbf{Set}]$$

is defined on objects A by $H^\bullet(A) = H^A$ and on maps f by $H^\bullet(f) = H^f$.

The symbol $\bullet$ is another type of blank, like $-$.

All of the definitions presented so far in this chapter can be dualized. At the formal level, this is trivial: reverse all the arrows, so that every $\mathscr{A}$ becomes an $\mathscr{A}^{\mathrm{op}}$ and vice versa. But in our usual examples, the flavour is different. We are no longer asking what objects *see*, but how they are *seen*.

Let us first dualize Definition 3.1.1.

Definition 3.1.16 Let $\mathscr{A}$ be a locally small category and $A \in \mathscr{A}$. We define a functor

$$H_A = \mathscr{A}(-, A) \colon \mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$$

as follows:

- for objects $B \in \mathscr{A}$, put $H_A(B) = \mathscr{A}(B, A)$;
- for maps $B' \xrightarrow{g} B$ in $\mathscr{A}$, define

  $$H_A(g) = \mathscr{A}(g, A) = g^* = - \circ g \colon \mathscr{A}(B, A) \to \mathscr{A}(B', A)$$

  by

  $$p \mapsto p \circ g$$

  for all $p \colon B \to A$.

If you know about dual vector spaces, this construction will seem familiar. In particular, you will not be surprised that a map $B' \to B$ induces a map in the opposite direction, $H_A(B) \to H_A(B')$.

We now define representability for *contravariant* set-valued functors. Strictly speaking, this is unnecessary, as a contravariant functor on $\mathscr{A}$ is a covariant functor on $\mathscr{A}^{\mathrm{op}}$, and we already know what it means for a covariant set-valued functor to be representable. But it is useful to have a direct definition.

Definition 3.1.17 Let $\mathscr{A}$ be a locally small category. A functor $X\colon \mathscr{A}^{\mathrm{op}} \to$ **Set** is **representable** if $X \cong H_A$ for some $A \in \mathscr{A}$. A **representation** of X is a choice of an object $A \in \mathscr{A}$ and an isomorphism between H_A and X.

Example 3.1.18 There is a functor

$$\mathscr{P} : \mathbf{Set}^{\mathrm{op}} \to \mathbf{Set}$$

sending each set B to its power set $\mathscr{P}(B)$, and defined on maps $g\colon B' \to B$ by $(\mathscr{P}(g))(U) = g^{-1}U$ for all $U \in \mathscr{P}(B)$. (Here $g^{-1}U$ denotes the inverse image or preimage of U under g, defined by $g^{-1}U = \{x' \in B' \mid g(x') \in U\}$.) As we saw in Section **??**, a subset amounts to a map into the two-point set 2. Precisely put, $\mathscr{P} \cong H_2$.

Example 3.1.19 Similarly, there is a functor

$$\mathscr{O} : \mathbf{Top}^{\mathrm{op}} \to \mathbf{Set}$$

defined on objects B by taking $\mathscr{O}(B)$ to be the set of open subsets of B. If S denotes the two-point topological space in which exactly one of the two singleton subsets is open, then continuous maps from a space B into S correspond naturally to open subsets of B (Exercise 3.1.30). Hence $\mathscr{O} \cong H_S$, and $\mathscr{O}$ is representable.

Example 3.1.20 In Example 1.2.11, we defined a functor $C\colon \mathbf{Top}^{\mathrm{op}} \to \mathbf{Ring}$, assigning to each space the ring of continuous real-valued functions on it. The composite functor

$$\mathbf{Top}^{\mathrm{op}} \xrightarrow{C} \mathbf{Ring} \xrightarrow{U} \mathbf{Set}$$

is representable, since by definition, $U(C(X)) = \mathbf{Top}(X, \mathbb{R})$ for topological spaces X.

Previously, we assembled the covariant representables $(H^A)_{A\in\mathscr{A}}$ into one big functor $H^\bullet$. We now do the same for the contravariant representables $(H_A)_{A\in\mathscr{A}}$.

Any map $A \xrightarrow{f} A'$ in $\mathscr{A}$ induces a natural transformation

$$\begin{tikzcd}\mathscr{A}^{\mathrm{op}} \arrow[r, bend left, "H_A"{name=U}] \arrow[r, bend right, "H_{A'}"'{name=D}] & \mathbf{Set} \arrow[from=U, to=D, Rightarrow, "H_f"]\end{tikzcd}$$

(also called $\mathscr{A}(-, f)$, f_* or $f \circ -$), whose component at an object $B \in \mathscr{A}$ is

$$\begin{array}{ccc} H_A(B) = \mathscr{A}(B, A) & \to & H_{A'}(B) = \mathscr{A}(B, A') \\ p & \mapsto & f \circ p. \end{array}$$

Definition 3.1.21 Let $\mathscr{A}$ be a locally small category. The **Yoneda embedding** of $\mathscr{A}$ is the functor

$$H_\bullet \colon \mathscr{A} \to [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$$

defined on objects A by $H_\bullet(A) = H_A$ and on maps f by $H_\bullet(f) = H_f$.

Here is a summary of the definitions so far.

For each $A \in \mathscr{A}$, we have a functor	$\mathscr{A} \xrightarrow{H^A} \mathbf{Set}$.
Putting them all together gives a functor	$\mathscr{A}^{\mathrm{op}} \xrightarrow{H^\bullet} [\mathscr{A}, \mathbf{Set}]$.
For each $A \in \mathscr{A}$, we have a functor	$\mathscr{A}^{\mathrm{op}} \xrightarrow{H_A} \mathbf{Set}$.
Putting them all together gives a functor	$\mathscr{A} \xrightarrow{H_\bullet} [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$.

The second pair of functors is the dual of the first. Both involve contravariance; it cannot be avoided.

In the theory of representable functors, it does not make much difference whether we work with the first or the second pair. Any theorem that we prove about one dualizes to give a theorem about the other. We choose to work with the second pair, the H_As and $H_\bullet$. In a sense to be explained, $H_\bullet$ 'embeds' $\mathscr{A}$ into $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$. This can be useful, because the category $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ has some good properties that $\mathscr{A}$ might not have.

Exercise 3.1.27 asks you to prove that $H_\bullet$ is injective on isomorphism classes of objects. It is strongly recommended that you do it before reading on, as it encapsulates the key ideas of the rest of this chapter.

There is one more functor to define. It unifies the first and second pairs of functors shown above.

Definition 3.1.22 Let $\mathscr{A}$ be a locally small category. The functor

$$\mathrm{Hom}_{\mathscr{A}} : \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \to \mathbf{Set}$$

is defined by

$$\begin{tikzcd}
(A,B) \arrow[dd, shift left, "g"] & \mapsto & \mathscr{A}(A,B) \arrow[dd, "g\circ-\circ f"] \\
 & \mapsto & \\
(A',B') \arrow[uu, shift left, "f"] & \mapsto & \mathscr{A}(A',B').
\end{tikzcd}$$

In other words, $\mathrm{Hom}_{\mathscr{A}}(A,B) = \mathscr{A}(A,B)$ and $(\mathrm{Hom}_{\mathscr{A}}(f,g))(p) = g \circ p \circ f$, whenever $A' \xrightarrow{f} A \xrightarrow{p} B \xrightarrow{g} B'$.

Remarks 3.1.23 (a) The existence of the functor $\mathrm{Hom}_{\mathscr{A}}$ is something like the fact that for a metric space (X,d), the metric is itself a continuous map $d\colon X \times X \to \mathbb{R}$. (If we take two points and move each one slightly, the distance between them changes only slightly.)

(b) In terms of Exercise 1.2.25, $\mathrm{Hom}_{\mathscr{A}}$ is the functor $\mathscr{A}^{\mathrm{op}} \times \mathscr{A} \to \mathbf{Set}$ corresponding to the families of functors $(H^A)_{A\in\mathscr{A}}$ and $(H_B)_{B\in\mathscr{A}}$.

(c) In Example 2.1.6, we saw that for any set B, there is an adjunction $(-\times B) \dashv (-)^B$ of functors $\mathbf{Set} \to \mathbf{Set}$. Similarly, for any category $\mathscr{B}$, there is an adjunction $(- \times \mathscr{B}) \dashv [\mathscr{B}, -]$ of functors $\mathbf{CAT} \to \mathbf{CAT}$; in other words, there is a canonical bijection

$$\mathbf{CAT}(\mathscr{A} \times \mathscr{B}, \mathscr{C}) \cong \mathbf{CAT}(\mathscr{A}, [\mathscr{B}, \mathscr{C}])$$

for $\mathscr{A}, \mathscr{B}, \mathscr{C} \in \mathbf{CAT}$. Under this bijection, the functors

$$\mathrm{Hom}_{\mathscr{A}} : \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \to \mathbf{Set}, \qquad H^{\bullet}\colon \mathscr{A}^{\mathrm{op}} \to [\mathscr{A}, \mathbf{Set}]$$

correspond to one another. Thus, $\mathrm{Hom}_{\mathscr{A}}$ carries the same information as $H^{\bullet}$ (or $H_{\bullet}$), presented slightly differently.

Remark 3.1.24 We can now explain the naturality in the definition of adjunction (Definition 2.1.1). Take categories and functors $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$. They give rise to functors

$$\begin{tikzcd}
\mathscr{A}^{\mathrm{op}} \times \mathscr{B} \arrow[r, "1\times G"] \arrow[d, "F^{\mathrm{op}}\times 1"'] & \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \arrow[d, "\mathrm{Hom}_{\mathscr{A}}"] \\
\mathscr{B}^{\mathrm{op}} \times \mathscr{B} \arrow[r, "\mathrm{Hom}_{\mathscr{B}}"'] & \mathbf{Set}.
\end{tikzcd}$$

The composite functor $\downarrow_{\rightarrow}$ sends (A,B) to $\mathscr{B}(F(A),B)$; it can be written as $\mathscr{B}(F(-),-)$. The composite $^{\rightarrow}\!\downarrow$ sends (A,B) to $\mathscr{A}(A,G(B))$. Exercise 3.1.32 asks you to show that these two functors

$$\mathscr{B}(F(-),-),\ \mathscr{A}(-,G(-))\colon \mathscr{A}^{\mathrm{op}} \times \mathscr{B} \to \mathbf{Set}$$

are naturally isomorphic if and only if F and G are adjoint. This justifies the claim in Remark 2.1.2(a): the naturality requirements (2.2) and (2.3) in the definition of adjunction simply assert that two particular functors are naturally isomorphic.

Objects of an arbitrary category do not have elements in any obvious sense. However, *sets* certainly have elements, and we have observed that an element of a set A is the same thing as a map $1 \to A$. This inspires the following definition.

Definition 3.1.25 Let A be an object of a category. A **generalized element** of A is a map with codomain A. A map $S \to A$ is a generalized element of A of **shape** S.

'Generalized element' is nothing more than a synonym of 'map', but sometimes it is useful to think of maps as generalized elements.

For example, when A is a set, a generalized element of A of shape 1 is an ordinary element of A, and a generalized element of A of shape $\mathbb{N}$ is a sequence in A. In the category of topological spaces, the generalized elements of shape 1 (the one-point space) are the points, and the generalized elements of shape S^1 (the circle) are, by definition, loops. As this suggests, in categories of geometric objects, we might equally well say 'figures of shape S'.

In algebra, we are often interested in solutions to equations such as $x^2+y^2 = 1$. Perhaps we begin by being particularly interested in solutions in $\mathbb{Q}$, but then realize that in order to study rational solutions, it will be helpful to study solutions in other rings first. (This is often a fruitful strategy.) Given a ring A, a pair $(a, b) \in A \times A$ satisfying $a^2 + b^2 = 1$ amounts to a homomorphism of rings

$$\mathbb{Z}[x, y]/(x^2 + y^2 - 1) \to A.$$

Thus, the solutions to our equation (in any ring) can be seen as the generalized elements of shape $\mathbb{Z}[x, y]/(x^2 + y^2 - 1)$.

For an object S of a category $\mathscr{A}$, the functor

$$H^S : \mathscr{A} \to \mathbf{Set}$$

sends an object to its set of generalized elements of shape S. The functoriality tells us that any map $A \to B$ in $\mathscr{A}$ transforms S-elements of A into S-elements of B. For example, taking $\mathscr{A} = \mathbf{Top}$ and $S = S^1$, any continuous map $A \to B$ transforms loops in A into loops in B.

Exercises

3.1.26 Find three examples of representable functors not mentioned above.

3.1.27 Let $\mathscr{A}$ be a locally small category, and let $A, A' \in \mathscr{A}$ with $H_A \cong H_{A'}$. Prove directly that $A \cong A'$.

3.1.28 Let p be a prime number. Show that the functor $U_p \colon \mathbf{Grp} \to \mathbf{Set}$ defined in Example 3.1.5 is isomorphic to $\mathbf{Grp}(\mathbb{Z}/p\mathbb{Z}, -)$. (To check that there is an isomorphism of functors – that is, a *natural* isomorphism – you will first need to define U_p on maps. There is only one sensible way to do this.)

3.1.29 Using the result of Exercise **????**, prove that the forgetful functor **CRing** → **Set** is isomorphic to $\mathbf{CRing}(\mathbb{Z}[x], -)$, as in Example 3.1.14.

3.1.30 The **Sierpiński space** is the two-point topological space S in which one of the singleton subsets is open but the other is not. Prove that for any topological space X, there is a canonical bijection between the open subsets of X and the continuous maps $X \to S$. Use this to show that the functor $\mathscr{O}$: $\mathbf{Top}^{\mathrm{op}} \to \mathbf{Set}$ of Example 3.1.19 is represented by S.

3.1.31 Let $M \colon \mathbf{Cat} \to \mathbf{Set}$ be the functor that sends a small category $\mathscr{A}$ to the set of all maps in $\mathscr{A}$. Prove that M is representable.

3.1.32 Take locally small categories $\mathscr{A}$ and $\mathscr{B}$, and functors $\begin{tikzcd}\mathscr{A} \arrow[r, shift left, "F"] & \mathscr{B} \arrow[l, shift left, "G"]\end{tikzcd}$. Show that F is left adjoint to G if and only if the two functors

$$\mathscr{B}(F(-), -),\ \mathscr{A}(-, G(-)) \colon \mathscr{A}^{\mathrm{op}} \times \mathscr{B} \to \mathbf{Set}$$

of Remark 3.1.24 are naturally isomorphic. (Hint: this is made easier by using either Exercise 1.3.29 or Exercise 2.1.14.)

3.2 The Yoneda lemma

What do representables see?

Recall from Definition 1.2.15 that functors $\mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$ are sometimes called ‘presheaves’ on $\mathscr{A}$. So for each $A \in \mathscr{A}$ we have a representable presheaf H_A, and we are asking how the rest of the presheaf category $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ looks from the viewpoint of H_A. In other words, if X is another presheaf, what are the maps $H_A \to X$?

Newcomers to category theory commonly find that the material presented in this section is where they first get stuck. Typically, the core of the difficulty is in understanding the question just asked. Let us ask it again.

We start by fixing a locally small category $\mathscr{A}$. We then take an object $A \in \mathscr{A}$ and a functor $X \colon \mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$. The object A gives rise to another functor

$H_A = \mathscr{A}(-, A) \colon \mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$. The question is: what are the maps $H_A \to X$? Since H_A and X are both objects of the presheaf category $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$, the 'maps' concerned are maps in $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$. So, we are asking what natural transformations

$$\begin{tikzcd} \mathscr{A}^{\mathrm{op}} \arrow[r, bend left, "H_A"] \arrow[r, bend right, "X"'] \arrow[r, phantom, "\Downarrow"] & \mathbf{Set} \end{tikzcd} \tag{3.1}$$

there are. The set of such natural transformations is called

$$[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X).$$

(This is a special case of the notation $\mathscr{B}(B, B')$ for the set of maps $B \to B'$ in a category $\mathscr{B}$. Here, $\mathscr{B} = [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$, $B = H_A$, and $B' = X$.) We want to know what this set is.

There is an informal principle of general category theory that allows us to guess the answer. Look back at Remarks 1.1.2(b), 1.2.2(a) and 1.3.2(a) on the definitions of category, functor and natural transformation. Each remark is of the form 'from input of one type, it is possible to construct exactly one output of another type'. For example, in Remark 1.1.2(b), the input is a sequence of maps $A_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} A_n$, the output is a map $A_0 \to A_n$, and the statement is that no matter what we do with the input data $f_1, \ldots, f_n$, there is only one map $A_0 \to A_n$ that we can construct.

Let us apply this principle to our question. We have just seen how, given as input an object $A \in \mathscr{A}$ and a presheaf X on $\mathscr{A}$, we can construct a set, namely, $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X)$. Are there any other ways to construct a set from the same input data (A, X)? Yes: simply take the set $X(A)$! The informal principle suggests that these two sets are the same:

$$[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \cong X(A) \tag{3.2}$$

for all $A \in \mathscr{A}$ and $X \in [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$. This turns out to be true; and that is the Yoneda lemma.

Informally, then, the Yoneda lemma says that for any $A \in \mathscr{A}$ and presheaf X on $\mathscr{A}$:

> *A natural transformation $H_A \to X$ is an element of $X(A)$.*

Here is the formal statement. The proof follows shortly.

Theorem 3.2.1 (Yoneda) *Let $\mathscr{A}$ be a locally small category. Then*

$$[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \cong X(A) \tag{3.3}$$

naturally in $A \in \mathscr{A}$ and $X \in [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$.

This is exactly what was stated in (3.2), except that the word 'naturally' has appeared. Recall from Definition 1.3.12 that for functors $F, G\colon \mathscr{C} \to \mathscr{D}$, the phrase '$F(C) \cong G(C)$ naturally in C' means that there is a natural isomorphism $F \cong G$. So the use of this phrase in the Yoneda lemma suggests that each side of (3.3) is functorial in both A and X. This means, for instance, that a map $X \to X'$ must induce a map

$$[\mathscr{A}^{\text{op}}, \mathbf{Set}](H_A, X) \to [\mathscr{A}^{\text{op}}, \mathbf{Set}](H_A, X'),$$

and that not only does the isomorphism (3.3) hold for *every* A and X, but also, the isomorphisms can be chosen in a way that is compatible with these induced maps. Precisely, the Yoneda lemma states that the composite functor

$$\begin{tikzcd}
\mathscr{A}^{\text{op}} \times [\mathscr{A}^{\text{op}}, \mathbf{Set}] \arrow[r, "H_\bullet^{\text{op}} \times 1"] & {[\mathscr{A}^{\text{op}}, \mathbf{Set}]^{\text{op}} \times [\mathscr{A}^{\text{op}}, \mathbf{Set}]} \arrow[r, "\text{Hom}_{[\mathscr{A}^{\text{op}}, \mathbf{Set}]}"] & \mathbf{Set} \\
(A, X) \arrow[r, mapsto] & (H_A, X) \arrow[r, mapsto] & {[\mathscr{A}^{\text{op}}, \mathbf{Set}](H_A, X)}
\end{tikzcd}$$

is naturally isomorphic to the evaluation functor

$$\begin{array}{ccc}
\mathscr{A}^{\text{op}} \times [\mathscr{A}^{\text{op}}, \mathbf{Set}] & \to & \mathbf{Set} \\
(A, X) & \mapsto & X(A).
\end{array}$$

If the Yoneda lemma were false then the world would look much more complex. For take a presheaf $X\colon \mathscr{A}^{\text{op}} \to \mathbf{Set}$, and define a new presheaf X' by

$$X' = [\mathscr{A}^{\text{op}}, \mathbf{Set}](H_\bullet, X)\colon \mathscr{A}^{\text{op}} \to \mathbf{Set},$$

that is, $X'(A) = [\mathscr{A}^{\text{op}}, \mathbf{Set}](H_A, X)$ for all $A \in \mathscr{A}$. Yoneda tells us that $X'(A) \cong X(A)$ naturally in A; in other words, $X' \cong X$. If Yoneda were false then starting from a single presheaf X, we could build an infinite sequence $X, X', X'', \ldots$ of new presheaves, potentially all different. But in reality, the situation is very simple: they are all the same.

The proof of the Yoneda lemma is the longest proof so far. Nevertheless, there is essentially only one way to proceed at each stage. If you suspect that you are one of those newcomers to category theory for whom the Yoneda lemma presents the first serious challenge, an excellent exercise is to work out the proof before reading it. No ingenuity is required, only an understanding of all the terms in the statement.

Proof of the Yoneda lemma We have to define, for each A and X, a bijection between the sets $[\mathscr{A}^{\text{op}}, \mathbf{Set}](H_A, X)$ and $X(A)$. We then have to show that our bijection is natural in A and X.

First, fix $A \in \mathscr{A}$ and $X \in [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$. We define functions

$$[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \underset{\tilde{(\ )}}{\overset{\hat{(\ )}}{\rightleftarrows}} X(A) \tag{3.4}$$

and show that they are mutually inverse. So we have to do four things: define the function $\hat{(\ )}$, define the function $\tilde{(\ )}$, show that $\hat{\tilde{(\ )}}$ is the identity, and show that $\tilde{\hat{(\ )}}$ is the identity.

- Given $\alpha\colon H_A \to X$, define $\hat{\alpha} \in X(A)$ by $\hat{\alpha} = \alpha_A(1_A)$. (How else could we possibly define it?)
- Let $x \in X(A)$. We have to define a natural transformation $\tilde{x}\colon H_A \to X$. That is, we have to define for each $B \in \mathscr{A}$ a function

  $$\tilde{x}_B\colon H_A(B) = \mathscr{A}(B, A) \to X(B)$$

  and show that the family $\tilde{x} = (\tilde{x}_B)_{B \in \mathscr{A}}$ satisfies naturality.

  Given $B \in \mathscr{A}$ and $f \in \mathscr{A}(B, A)$, define

  $$\tilde{x}_B(f) = (X(f))(x) \in X(B).$$

  (How else could we possibly define it?) This makes sense, since $X(f)$ is a map $X(A) \to X(B)$. To prove naturality, we must show that for any map $B' \xrightarrow{g} B$ in $\mathscr{A}$, the square

  $$\begin{tikzcd}
  \mathscr{A}(B,A) \arrow[r, "H_A(g) = -\circ g"] \arrow[d, "\tilde{x}_B"'] & \mathscr{A}(B',A) \arrow[d, "\tilde{x}_{B'}"] \\
  X(B) \arrow[r, "X(g)"'] & X(B')
  \end{tikzcd}$$

  commutes. To reduce clutter, let us write $X(g)$ as Xg, and so on. Now for all $f \in \mathscr{A}(B, A)$, we have

  $$\begin{tikzcd}
  f \arrow[r, mapsto] \arrow[d, mapsto] & f \circ g \arrow[d, mapsto] \\
  (Xf)(x) \arrow[r, mapsto] & \begin{matrix}(X(f \circ g))(x) \\ (Xg)((Xf)(x)),\end{matrix}
  \end{tikzcd}$$

  and $X(f \circ g) = (Xg) \circ (Xf)$ by functoriality, so the square does commute.
- Given $x \in X(A)$, we have to show that $\hat{\tilde{x}} = x$, and indeed,

  $$\hat{\tilde{x}} = \tilde{x}_A(1_A) = (X1_A)(x) = 1_{X(A)}(x) = x.$$

- Given $\alpha\colon H_A \to X$, we have to show that $\tilde{\hat{\alpha}} = \alpha$. Two natural transformations are equal if and only if all their components are equal; so, we have to show that $\left(\tilde{\hat{\alpha}}\right)_B = \alpha_B$ for all $B \in \mathscr{A}$. Each side of this equation is a function from $H_A(B) = \mathscr{A}(B, A)$ to $X(B)$, and two functions are equal if and only if they take equal values at every element of the domain; so, we have to show that

  $$\left(\tilde{\hat{\alpha}}\right)_B(f) = \alpha_B(f)$$

  for all $B \in \mathscr{A}$ and $f\colon B \to A$ in $\mathscr{A}$. The left-hand side is by definition

  $$\left(\tilde{\hat{\alpha}}\right)_B(f) = (Xf)(\hat{\alpha}) = (Xf)(\alpha_A(1_A)),$$

  so it remains to prove that

  $$(Xf)(\alpha_A(1_A)) = \alpha_B(f). \qquad (3.5)$$

  By naturality of α (the only tool at our disposal), the square

  $$\begin{tikzcd} \mathscr{A}(A,A) \arrow[r, "H_A(f) = - \circ f"] \arrow[d, "\alpha_A"'] & \mathscr{A}(B,A) \arrow[d, "\alpha_B"] \\ X(A) \arrow[r, "Xf"'] & X(B) \end{tikzcd}$$

  commutes, which when taken at $1_A \in \mathscr{A}(A, A)$ gives equation (3.5).

(The proof is not over yet, but it is worth pausing to consider the significance of the fact that $\tilde{\hat{\alpha}} = \alpha$. Since $\hat{\alpha}$ is the value of α at 1_A, this implies:

> *A natural transformation $H_A \to X$ is determined by its value at 1_A.*

Just *how* a natural transformation $H_A \to X$ is determined by its value at 1_A is described in equation (3.5).)

This establishes the bijection (3.4) for each $A \in \mathscr{A}$ and $X \in [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$. We now show that the bijection is natural in A and X.

We employ two mildly labour-saving devices. First, in principle we have to prove naturality of both $(\hat{\ })$ and $(\tilde{\ })$, but by Lemma 1.3.11, it is enough to prove naturality of just one of them. We prove naturality of $(\hat{\ })$. Second, by Exercise 1.3.29, $(\hat{\ })$ is natural in the pair (A, X) if and only if it is natural in A for each fixed X and natural in X for each fixed A. So, it remains to check these two types of naturality.

Naturality in A states that for each $X \in [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ and $B \xrightarrow{f} A$ in $\mathscr{A}$, the square

$$\begin{tikzcd}
{[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X)} \arrow[r, "- \circ H_f"] \arrow[d, "\hat{(\ )}"'] & {[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_B, X)} \arrow[d, "\hat{(\ )}"] \\
X(A) \arrow[r, "Xf"'] & X(B)
\end{tikzcd}$$

commutes. For $\alpha \colon H_A \to X$, we have

$$\begin{tikzcd}
\alpha \arrow[r, mapsto] \arrow[dd, mapsto] & \alpha \circ H_f \arrow[d, mapsto] \\
 & (\alpha \circ H_f)_B(1_B) \\
\alpha_A(1_A) \arrow[r, mapsto] & (Xf)(\alpha_A(1_A)),
\end{tikzcd}$$

so we have to show that $(\alpha \circ H_f)_B(1_B) = (Xf)(\alpha_A(1_A))$. Indeed, where the first step is by definition of composition in $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$, the second is by definition of H_f, and the last is by equation (3.5).

Naturality in X states that for each $A \in \mathscr{A}$ and map

$$\begin{tikzcd}
\mathscr{A}^{\mathrm{op}} \arrow[r, bend left, "X"{name=U}] \arrow[r, bend right, "X'"'{name=D}] & \mathbf{Set}
\arrow[from=U, to=D, Rightarrow, "\theta"]
\end{tikzcd}$$

in $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$, the square

$$\begin{tikzcd}
{[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X)} \arrow[r, "\theta \circ -"] \arrow[d, "\hat{(\ )}"'] & {[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X')} \arrow[d, "\hat{(\ )}"] \\
X(A) \arrow[r, "\theta_A"'] & X'(A)
\end{tikzcd}$$

commutes. For $\alpha \colon H_A \to X$, we have

$$\begin{tikzcd}
\alpha \arrow[r, mapsto] \arrow[dd, mapsto] & \theta \circ \alpha \arrow[d, mapsto] \\
 & (\theta \circ \alpha)_A(1_A) \\
\alpha_A(1_A) \arrow[r, mapsto] & \theta_A(\alpha_A(1_A)),
\end{tikzcd}$$

and $(\theta \circ \alpha)_A = \theta_A \circ \alpha_A$ by definition of composition in $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$, so the square does commute. This completes the proof. □

Exercises

3.2.2 State the dual of the Yoneda lemma.

3.2.3 One way to understand the Yoneda lemma is to examine some special cases. Here we consider one-object categories.

Let M be a monoid. The underlying set of M can be given a right M-action by multiplication: $x \cdot m = xm$ for all $x, m \in M$. This M-set is called the **right regular representation** of M. Let us write it as $\underline{M}$.

(a) When M is regarded as a one-object category, functors $M^{\mathrm{op}} \to \mathbf{Set}$ correspond to right M-sets (Example 1.2.14). Show that the M-set corresponding to the unique representable functor $M^{\mathrm{op}} \to \mathbf{Set}$ is the right regular representation.

(b) Now let X be any right M-set. Show that for each $x \in X$, there is a unique map $\alpha\colon \underline{M} \to X$ of right M-sets such that $\alpha(1) = x$. Deduce that there is a bijection between {maps $\underline{M} \to X$ of right M-sets} and X.

(c) Deduce the Yoneda lemma for one-object categories.

3.3 Consequences of the Yoneda lemma

The Yoneda lemma is fundamental in category theory. Here we look at three important consequences.

Notation 3.3.1 An arrow decorated with a ~, as in $A \xrightarrow{\sim} B$, denotes an isomorphism.

A representation is a universal element

Corollary 3.3.2 *Let $\mathscr{A}$ be a locally small category and $X\colon \mathscr{A}^{\mathrm{op}} \to \mathbf{Set}$. Then a representation of X consists of an object $A \in \mathscr{A}$ together with an element $u \in X(A)$ such that:*

$$\text{for each } B \in \mathscr{A} \text{ and } x \in X(B), \text{ there is a unique map } \bar{x}\colon B \to A \text{ such that } (X\bar{x})(u) = x. \tag{3.6}$$

To clarify the statement, first recall that by definition, a representation of X is an object $A \in \mathscr{A}$ together with a natural isomorphism $\alpha\colon H_A \xrightarrow{\sim} X$. Corollary 3.3.2 states that such pairs (A, α) are in natural bijection with pairs (A, u) satisfying condition (3.6).

Pairs (B, x) with $B \in \mathscr{A}$ and $x \in X(B)$ are sometimes called **elements** of the presheaf X. (Indeed, the Yoneda lemma tells us that x amounts to a generalized element of X of shape H_B.) An element u satisfying condition (3.6) is sometimes called a **universal** element of X. So, Corollary 3.3.2 says that a representation of a presheaf X amounts to a universal element of X.

Proof By the Yoneda lemma, we have only to show that for $A \in \mathscr{A}$ and $u \in X(A)$, the natural transformation $\tilde{u}\colon H_A \to X$ is an isomorphism if and only if (3.6) holds. (Here we are using the notation introduced in the proof of the Yoneda lemma.) Now, $\tilde{u}$ is an isomorphism if and only if for all $B \in \mathscr{A}$, the function

$$\tilde{u}_B\colon H_A(B) = \mathscr{A}(B, A) \to X(B)$$

is a bijection, if and only if for all $B \in \mathscr{A}$ and $x \in X(B)$, there is a unique $\bar{x} \in \mathscr{A}(B, A)$ such that $\tilde{u}_B(\bar{x}) = x$. But $\tilde{u}_B(\bar{x}) = (X\bar{x})(u)$, so this is exactly condition (3.6). □

Our examples will use the dual form, for covariant set-valued functors:

Corollary 3.3.3 *Let $\mathscr{A}$ be a locally small category and $X\colon \mathscr{A} \to$ **Set**. Then a representation of X consists of an object $A \in \mathscr{A}$ together with an element $u \in X(A)$ such that:*

for each $B \in \mathscr{A}$ and $x \in X(B)$, there is a unique map $\bar{x}\colon A \to B$ such that $(X\bar{x})(u) = x$. (3.7)

Proof Follows immediately by duality. □

Example 3.3.4 Fix a set S and consider the functor

$$\begin{array}{rccc} X = \mathbf{Set}(S, U(-))\colon & \mathbf{Vect}_k & \to & \mathbf{Set} \\ & V & \mapsto & \mathbf{Set}(S, U(V)). \end{array}$$

Here are two familiar (and true!) statements about X:

(a) there exist a vector space $F(S)$ and an isomorphism

$$\mathbf{Vect}_k(F(S), V) \cong \mathbf{Set}(S, U(V)) \tag{3.8}$$

natural in $V \in \mathbf{Vect}_k$ (Example 2.1.3(a));

(b) there exist a vector space $F(S)$ and a function $u\colon S \to U(F(S))$ such that:

for each vector space V and function $f\colon S \to U(V)$, there is a unique linear map $\bar{f}\colon F(S) \to V$ such that

$$\begin{tikzcd} S \arrow[r, "u"] \arrow[dr, "f"'] & U(F(S)) \arrow[d, "U(\bar{f})"] \\ & U(V) \end{tikzcd}$$

commutes

(as in the introduction to Section 2.3, where u was called by its usual name, η_S).

Each of these two statements says that X is representable. Statement (a) says that there is an isomorphism $X(V) \cong \mathbf{Vect}(F(S), V)$ natural in V, that is, an isomorphism $X \cong H^{F(S)}$. So X is representable, by definition of representability. Statement (b) says that $u \in X(F(S))$ satisfies condition (3.7). So X is representable, by Corollary 3.3.3.

You will have noticed that the first way of saying that X is representable is substantially shorter than the second. Indeed, it is clear that if the situation of (b) holds then there is an isomorphism

$$\mathbf{Vect}_k(F(S), V) \xrightarrow{\sim} \mathbf{Set}(S, U(V))$$

natural in V, defined by $g \mapsto U(g) \circ u$. But it looks at first as if (b) says rather more than (a), since it states that the two functors are not only naturally isomorphic, but naturally isomorphic in a rather special way. Corollary 3.3.3 tells us that this is an illusion: all natural isomorphisms (3.8) arise in this way. It is the word 'natural' in (a) that hides the explicit detail.

Example 3.3.5 The same can be said for any other adjunction $\begin{tikzcd} \mathscr{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, shift left, "G"] \end{tikzcd}$. Fix $A \in \mathscr{A}$ and put

$$X = \mathscr{A}(A, G(-)) \colon \mathscr{B} \to \mathbf{Set}.$$

Then X is representable, and this can be expressed in either of the following ways:

(a) $\mathscr{A}(A, G(B)) \cong \mathscr{B}(F(A), B)$ naturally in B; in other words, $X \cong H^{F(A)}$ (as in Lemma 3.1.10);

(b) the unit map $\eta_A \colon A \to G(F(A))$ is an initial object of the comma category $(A \Rightarrow G)$; that is, $\eta_A \in X(F(A))$ satisfies condition (3.7).

This observation can be developed into an alternative proof of Theorem 2.3.6, the reformulation of adjointness in terms of initial objects.

Example 3.3.6 For any group G and element $x \in G$, there is a unique homomorphism $\phi \colon \mathbb{Z} \to G$ such that $\phi(1) = x$. This means that $1 \in U(\mathbb{Z})$ is a universal element of the forgetful functor $U \colon \mathbf{Grp} \to \mathbf{Set}$; in other words, condition (3.7) holds when $\mathscr{A} = \mathbf{Grp}$, $X = U$, $A = \mathbb{Z}$ and $u = 1$. So $1 \in U(\mathbb{Z})$ gives a representation $H^{\mathbb{Z}} \xrightarrow{\sim} U$ of U.

On the other hand, the same is true with -1 in place of 1. The isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} U$ coming from 1 and -1 are not equal, because Corollary 3.3.3 provides a *one-to-one* correspondence between universal elements and representations.

The Yoneda embedding

Here is a second corollary of the Yoneda lemma.

Corollary 3.3.7 *For any locally small category $\mathscr{A}$, the Yoneda embedding*

$$H_\bullet \colon \mathscr{A} \to [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$$

is full and faithful.

Informally, this says that for $A, A' \in \mathscr{A}$, a map $H_A \to H_{A'}$ of presheaves is the same thing as a map $A \to A'$ in $\mathscr{A}$.

Proof We have to show that for each $A, A' \in \mathscr{A}$, the function

$$\begin{array}{ccc} \mathscr{A}(A, A') & \to & [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, H_{A'}) \\ f & \mapsto & H_f \end{array} \tag{3.9}$$

is bijective. By the Yoneda lemma (taking 'X' to be $H_{A'}$), the function

$$(\tilde{\ })\colon H_{A'}(A) \to [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}](H_A, H_{A'}) \tag{3.10}$$

is bijective, so it is enough to prove that the functions (3.9) and (3.10) are equal. Thus, given $f\colon A \to A'$, we have to prove that $\tilde{f} = H_f$, or equivalently, $\widehat{H_f} = f$. And indeed,

$$\widehat{H_f} = (H_f)_A(1_A) = f \circ 1_A = f,$$

as required. □

In mathematics at large, the word 'embedding' is used (sometimes informally) to mean a map $A \to B$ that makes A isomorphic to its image in B. For example, an injection of sets $i\colon A \to B$ might be called an embedding, because it provides a bijection between A and the subset iA of B. Similarly, a map $i\colon A \to B$ of topological spaces might be called an embedding if it is a homeomorphism to its image, so that $A \cong iA$. Corollary 1.3.19 tells us that in category theory, a full and faithful functor $\mathscr{A} \to \mathscr{B}$ can reasonably be called an embedding, as it makes $\mathscr{A}$ equivalent to a full subcategory of $\mathscr{B}$.

In the case at hand, the Yoneda embedding $H_\bullet\colon \mathscr{A} \to [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ embeds $\mathscr{A}$ into its own presheaf category (Figure 3.1). So, $\mathscr{A}$ is equivalent to the full subcategory of $[\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ whose objects are the representables.

In general, full subcategories are the easiest subcategories to handle. For instance, given objects A and A' of a full subcategory, we can speak unambiguously of the 'maps' from A to A'; it makes no difference whether this is understood to mean maps in the subcategory or maps in the whole category. Similarly, we can speak unambiguously of isomorphism of objects of the subcategory, as in the following lemma.

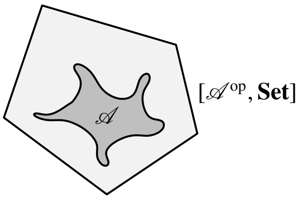


Figure 3.1 A category $\mathscr{A}$ embedded into its presheaf category.

Lemma 3.3.8 *Let $J\colon \mathscr{A} \to \mathscr{B}$ be a full and faithful functor and $A, A' \in \mathscr{A}$. Then:*

(a) *a map f in $\mathscr{A}$ is an isomorphism if and only if the map $J(f)$ in $\mathscr{B}$ is an isomorphism;*
(b) *for any isomorphism $g\colon J(A) \to J(A')$ in $\mathscr{B}$, there is a unique isomorphism $f\colon A \to A'$ in $\mathscr{A}$ such that $J(f) = g$;*
(c) *the objects A and A' of $\mathscr{A}$ are isomorphic if and only if the objects $J(A)$ and $J(A')$ of $\mathscr{B}$ are isomorphic.*

Proof Exercise 3.3.15. □

Example 3.3.9 In Example 3.3.6, we considered the representations of the forgetful functor $U\colon \mathbf{Grp} \to \mathbf{Set}$, and found two different isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} U$. Did we find all of them?

Since $H^{\mathbb{Z}} \cong U$, there are as many isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} U$ as there are isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} H^{\mathbb{Z}}$. By Corollary 3.3.7 and Lemma 3.3.8(b), there are as many of *these* as there are group isomorphisms $\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}$. There are precisely two such (corresponding to the two generators ±1 of $\mathbb{Z}$), so we did indeed find all the isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} U$. Differently put, there are exactly two universal elements of $U(\mathbb{Z})$.

In Section **??**, we will see that every presheaf can be built from representables, in very roughly the same way that every positive integer can be built from primes.

Isomorphism of representables

In Exercise 3.1.27, you were asked to prove directly that if $H_A \cong H_{A'}$ then $A \cong A'$. The proof contains all the main ideas in the proof of the Yoneda lemma. The result itself can also be deduced from the Yoneda lemma, as follows.

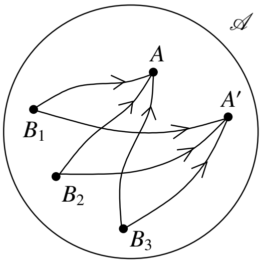


Figure 3.2 If $\mathscr{A}(B, A) \cong \mathscr{A}(B, A')$ naturally in B, then $A \cong A'$.

Corollary 3.3.10 *Let $\mathscr{A}$ be a locally small category and $A, A' \in \mathscr{A}$. Then*

$$H_A \cong H_{A'} \iff A \cong A' \iff H^A \cong H^{A'}.$$

Proof By duality, it is enough to prove the first '$\iff$'. This follows from Corollary 3.3.7 and Lemma 3.3.8(c). □

Since functors always preserve isomorphism (Exercise 1.2.21), the force of this statement is that

$$H_A \cong H_{A'} \implies A \cong A'.$$

In other words, if $\mathscr{A}(B, A) \cong \mathscr{A}(B, A')$ naturally in B, then $A \cong A'$. Thinking of $\mathscr{A}(B, A)$ as 'A viewed from B', the corollary tells us that two objects are the same if and only if they look the same from all viewpoints (Figure 3.2). (If it looks like a duck, walks like a duck, and quacks like a duck, then it probably is a duck.)

Example 3.3.11 Consider Corollary 3.3.10 in the case $\mathscr{A} = \mathbf{Grp}$. Take two groups A and A', and suppose someone tells us that A and A' 'look the same from B' (meaning that $H_A(B) \cong H_{A'}(B)$) for all groups B. Then, for instance:

- $H_A(1) \cong H_{A'}(1)$, where 1 is the trivial group. But $H_A(1) = \mathbf{Grp}(1, A)$ is a one-element set, as is $H_{A'}(1)$, no matter what A and A' are. So this tells us nothing at all.
- $H_A(\mathbb{Z}) \cong H_{A'}(\mathbb{Z})$. We know that $H_A(\mathbb{Z})$ is the underlying set of A, and similarly for A'. So A and A' have isomorphic underlying sets. But for all we know so far, they might have entirely different group structures.
- $H_A(\mathbb{Z}/p\mathbb{Z}) \cong H_{A'}(\mathbb{Z}/p\mathbb{Z})$ for every prime p, so by Example 3.1.5, A and A' have the same number of elements of each prime order.

Each of these isomorphisms gives only partial information about the similarity of A and A'. But if we know that $H_A(B) \cong H_{A'}(B)$ for all groups B, and *naturally* in B, then $A \cong A'$.

Example 3.3.12 The category of sets is very unusual in this respect. For any set A, we have

$$A \cong \mathbf{Set}(1, A) = H_A(1),$$

so $H_A(1) \cong H_{A'}(1)$ implies $A \cong A'$. In other words, two objects of **Set** are the same if they look the same from the point of view of the one-element set. This is a familiar feature of sets: the only thing that matters about a set is its elements!

For a general category, Corollary 3.3.10 tells us that two objects are the same if they have the same generalized elements of all shapes. But the category of sets has a special property: if I choose an object and tell you only what its generalized elements of shape 1 are, then you can deduce exactly what my object must be.

Example 3.3.13 Let $G\colon \mathscr{B} \to \mathscr{A}$ be a functor, and suppose that both F and F' are left adjoint to G. Then for each $A \in \mathscr{A}$, we have

$$\mathscr{B}(F(A), B) \cong \mathscr{A}(A, G(B)) \cong \mathscr{B}(F'(A), B)$$

naturally in $B \in \mathscr{B}$, so $H^{F(A)} \cong H^{F'(A)}$, so $F(A) \cong F'(A)$ by Corollary 3.3.10. In fact, this isomorphism is natural in A, so that $F \cong F'$. This shows that left adjoints are unique, as claimed in Remark 2.1.2(d). Dually, right adjoints are unique. See also Exercise 3.3.18.

Example 3.3.14 Corollary 3.3.10 implies that if a set-valued functor is isomorphic to both H^A and $H^{A'}$ then $A \cong A'$. So the functor *determines* the representing object, if one exists. For instance, take the functor

$$\mathbf{Bilin}(U, V; -)\colon \mathbf{Vect}_k \to \mathbf{Set}$$

of Example 3.1.9. Corollary 3.3.10 implies that up to isomorphism, there is *at most one* vector space T such that

$$\mathbf{Bilin}(U, V; W) \cong \mathbf{Vect}_k(T, W)$$

naturally in W. It can be shown that there does, in fact, exist such a vector space T. Since all such spaces T are isomorphic, it is legitimate to refer to any of them as *the* tensor product of U and V.

Exercises

3.3.15 Prove Lemma 3.3.8.

3.3.16 Let $\mathscr{A}$ be a locally small category. Prove each of the following statements directly (without using the Yoneda lemma).

(a) $H_\bullet \colon \mathscr{A} \to [\mathscr{A}^{\mathrm{op}}, \mathbf{Set}]$ is faithful.
(b) $H_\bullet$ is full.
(c) Given $A \in \mathscr{A}$ and a presheaf X on $\mathscr{A}$, if $X(A)$ has an element u that is universal in the sense of Corollary 3.3.2, then $X \cong H_A$.

3.3.17 Interpret the theory of Chapter 3 in the case where the category $\mathscr{A}$ is discrete. For example, what do presheaves look like, and which ones are representable? What does the Yoneda lemma tell us? Does its proof become any shorter? What about the corollaries of the Yoneda lemma?

3.3.18 Let $\mathscr{B}$ be a category and $J \colon \mathscr{C} \to \mathscr{D}$ a functor. There is an induced functor

$$J \circ - \colon [\mathscr{B}, \mathscr{C}] \to [\mathscr{B}, \mathscr{D}]$$

defined by composition with J.

(a) Show that if J is full and faithful then so is $J \circ -$.
(b) Deduce that if J is full and faithful and $G, G' \colon \mathscr{B} \to \mathscr{C}$ with $J \circ G \cong J \circ G'$ then $G \cong G'$.
(c) Now deduce that right adjoints are unique: if $F \colon \mathscr{A} \to \mathscr{B}$ and $G, G' \colon \mathscr{B} \to \mathscr{A}$ with $F \dashv G$ and $F \dashv G'$ then $G \cong G'$. (Hint: the Yoneda embedding is full and faithful.)

Index