Contents

Chapter 0

Introduction

Here I'll try to give you a flavour of what category theory is like, without actually mentioning categories.

A major theme of this course—and an important theme in mathematics as a whole—is that of 'universal property'. The further you go in mathematics, especially pure mathematics, the more often you'll meet this idea. So most of this introduction will be about universal properties.

Following that is a short description of the contents of the course.

0.1 Universal properties

Example 0.1.1 Let V be a vector space with a basis $(v_i)_{i \in I}$. Then for any vector space W, there is a 'natural' one-to-one correspondence between

linear maps $V \longrightarrow W$

and

functions $I \longrightarrow W$.

This is true because any function defined on the basis elements extends uniquely to a linear map on V.

Let's rephrase this fact. Define a function $b : I \longrightarrow V$ by $b(i) = v_i$ $(i \in I)$. Then we can easily prove the following 'universal property':

$$\begin{tikzcd}
I \arrow[r, "b"] \arrow[dr, "\forall f"'] & V \arrow[d, dotted, "\exists!\text{linear } \overline{f}"] \\
 & \forall W.
\end{tikzcd}$$

This means: for all vector spaces W and all functions $f : I \longrightarrow W$, there exists a unique linear map $\overline{f} : V \longrightarrow W$ such that $\overline{f} \circ b = f$.

(The symbols '∀', '∃' and '!' mean 'for all', 'there exists', and 'unique'. 'There exists a unique carrot' means 'there is one and only one carrot'.)

The universal property gives the one-to-one correspondence described above. Indeed, what it says is precisely that the function

$$\begin{array}{ccc}\{\text{linear maps } V \longrightarrow W\} & \longrightarrow & \{\text{functions } I \longrightarrow W\} \\ \overline{f} & \longmapsto & \overline{f} \circ b\end{array}$$

is a bijection. Note that the equation $\overline{f} \circ b = f$ in the universal property is equivalent to the statement that $\overline{f}(v_i) = f(i)$ for all $i \in I$.

Example 0.1.2 Any set S can be turned into a topological space $D(S)$ by giving it the discrete topology: all subsets are open. With this topology, *any* map from S to a space X is continuous.

Again, let's rephrase that. We have a function

$$\begin{array}{cccc} i: & S & \longrightarrow & D(S) \\ & s & \longmapsto & s \end{array}$$

($s \in S$) and the following universal property:

$$\begin{tikzcd} S \arrow[r, "i"] \arrow[rd, "\forall\text{functions } f"'] & D(S) \arrow[d, dotted, "\exists!\text{continuous } \overline{f}"] \\ & \forall X. \end{tikzcd}$$

The continuous map $\overline{f}$ 'is' the function f.

You may feel that this universal property is almost too trivial to mean anything. But if you change the definition of $D(S)$—say from the discrete to the indiscrete topology, where the only open sets are $\emptyset$ and S—you'll see that the property becomes false. So this property really does say something about the discrete topology.

Example 0.1.3 If U, V and W are vector spaces, a **bilinear map** $f : U \times V \longrightarrow W$ is a function f that's linear in each variable:

$$\begin{array}{rcl} f(u, v_1 + \lambda v_2) & = & f(u, v_1) + \lambda f(u, v_2), \\ f(u_1 + \lambda u_2, v) & = & f(u_1, v) + \lambda f(u_2, v) \end{array}$$

for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$, and scalars λ. A good example is the dot product (scalar product), which defines a bilinear map

$$\begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ (\mathbf{u}, \mathbf{v}) & \longmapsto & \mathbf{u}.\mathbf{v} \end{array}$$

of real vector spaces.

Given vector spaces U and V, there is a 'universal blinear map out of $U \times V$'. In other words, there are a certain vector space T and a certain bilinear map $b : U \times V \longrightarrow T$ with the following universal property:

$$\begin{tikzcd}
U \times V \arrow[r, "b"] \arrow[dr, "\forall\text{linear } f"'] & T \arrow[d, dotted, "\exists!\text{bilinear } \overline{f}"] \\
& \forall W.
\end{tikzcd} \tag{0:1}$$

(T is called the **tensor product** of U and V, written $T = U \otimes V$. This construction reduces the study of bilinear maps to the study of linear maps, since a bilinear map out of $U \times V$ is the same as a linear map out of $U \otimes V$.)

A universal property uniquely determines the object concerned. (More precisely, it determines it up to isomorphism—but who could want more?) This is an absolutely essential fact about universal properties. For example, in 0.1.3, we have:

Lemma 0.1.4 *Let U and V be vector spaces. Suppose that $b : U \times V \longrightarrow T$ and $b' : U' \times V' \longrightarrow T'$ are both universal bilinear maps out of $U \times V$. Then $T \cong T'$.*

Incidentally, I haven't shown that there *exists* a universal bilinear map out of $U \times V$. As it happens, one does exist, but the proof doesn't depend on knowing that.

Try proving the lemma yourself before reading the proof below. This will force you to understand what the universal property really says. It doesn't matter what 'bilinear', 'linear' or even 'vector space' mean: the hard part is getting those 'for all's and 'there exist's straight. That done, you should be able to see that there's really only one possible proof. For instance, to use the universality of b, you're going to have to choose some bilinear map f out of $U \times V$, and there are only two in sight: b and b'.

Proof In (0:1), take $(U \times V \xrightarrow{f} W)$ to be $(U \times V \xrightarrow{b'} T')$. Then we get a linear map $i : T \longrightarrow T'$ satisfying $i \circ b = b'$. Similarly, using the universality of b', we get a linear map $i' : T' \longrightarrow T$ satisfying $i' \circ b' = b$. Picture:

$$\begin{tikzcd}
& T \arrow[d, "i"] \\
U \times V \arrow[ur, "b"] \arrow[r, "b'"] \arrow[dr, "b"'] & T' \arrow[d, "i'"] \\
& T.
\end{tikzcd}$$

Now $i' \circ i : T \longrightarrow T$ is a linear map satisfying $(i' \circ i) \circ b = b$. Also $\mathrm{id}_T \circ b = b$. So by the uniqueness part of the universal property of b, we have $i' \circ i = \mathrm{id}_T$. (Here we took the '$f$' of (0:1) to be b.) Similarly, $i \circ i' = \mathrm{id}_{T'}$. So $T \cong T'$. $\square$

Exercise Prove analogous results for the universal properties in Examples 0.1.1 and 0.1.2.

Moral: once you've found a universal property of an object, you can forget how it was constructed.

For instance, if you look through a pile of algebra books you'll find several different ways of constructing the tensor product of vector spaces. But once you've proved that the tensor product satisfies the universal property, you can forget the construction. The universal property tells you all you need to know, because it determines the object uniquely.

Category theory typically takes this high-up viewpoint. We get a bird's-eye view of the world of mathematical objects and interactions between them. Details become invisible; only the main features can be seen.

Example 0.1.5 Let $\theta : G \longrightarrow H$ be a homomorphism of groups. Associated with θ is a diagram

$$\begin{tikzcd}
\ker(\theta) \arrow[r, hook, "\iota"] & G \arrow[r, shift left, "\theta"] \arrow[r, shift right, "\varepsilon"'] & H
\end{tikzcd}$$

where ι is the inclusion of $\ker(\theta)$ into G and ε is the trivial homomorphism, with constant value 1. (The symbol $\hookrightarrow$ is often used for inclusions: it's a combination of a subset symbol $\subset$ and an arrow.)

The map ι into G satisfies $\theta \circ \iota = \varepsilon \circ \iota$ and 'is universal as such'. Exercise: make this precise. In other words, write down a universal property similar to those in the previous examples.

Here is a final example of a universal property.

Example 0.1.6 Let X be a topological space covered by two open subsets: $X = U \cup V$. The diagram

$$\begin{tikzcd}
U \cap V \arrow[r, hook, "i"] \arrow[d, hook, "j"'] & U \arrow[d, hook, "j'"] \\
V \arrow[r, hook, "i'"'] & X
\end{tikzcd}$$

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

$$\begin{tikzcd}
U \cap V \arrow[r, hook, "i"] \arrow[d, hook, "j"'] & U \arrow[d, hook, "j'"] \arrow[ddr, "\forall f"] & \\
V \arrow[r, hook, "i'"'] \arrow[drr, "\forall g"'] & X \arrow[dr, dotted, "\exists! h"] & \\
 & & \forall Y.
\end{tikzcd}$$

This means: given Y, f and g such that $f \circ i = g \circ j$, there is exactly one $h : X \longrightarrow Y$ such that $h \circ j' = f$ and $h \circ i' = g$.

Under favourable conditions, the induced diagram

$$\begin{tikzcd}
\pi_1(U \cap V) \arrow[r, "i_*"] \arrow[d, "j_*"'] & \pi_1(U) \arrow[d, "j'_*"] \\
\pi_1(V) \arrow[r, "i'_*"'] & \pi_1(X)
\end{tikzcd}$$

of fundamental groups has the same property in the world of groups and group homomorphisms. This is **van Kampen's Theorem**. In fact, van Kampen stated his theorem in a much more complicated way, as he was working in the 1930s and categories had not yet been invented. Category theory lets you state the theorem in transparent language.

0.2 Outline of the course

Here are the chapter headings.

1. Categories, functors and natural transformations

A category is a system consisting of some objects, some maps between objects, and a way of composing strings of maps. A functor is a map between categories. For example, homology is a functor: historically, one of the first examples.

Naturality is a very important idea in mathematics. The attempt to formalize it, in the context of algebraic topology, was what led to the birth of category theory. By the early 1940s, people had started to use the phrase 'natural transformation' in an informal and imprecise way. Two algebraic topologists, Samuel

Eilenberg and Saunders Mac Lane, saw that a precise definition of natural transformation was needed. To define natural transformation, they had first to define functor; and to define functor, they had first to define category. Nowadays, the uses of category theory have spread far beyond algebraic topology.

Here are two examples of how the word 'natural' is used.

- For finite-dimensional vector spaces V, there is a 'natural' isomorphism $V \cong V^{**}$. Informally, what this means is that the isomorphism is God-given: to define it, you don't need to make any arbitrary choices. Contrast this with the single dual: to specify an isomorphism between V and V^*, you need to make an arbitrary choice of basis.

- For topological spaces W, X and Y, there is a 'natural' correspondence between

  continuous maps $W \longrightarrow X \times Y$

  and

  pairs $(W \longrightarrow X, W \longrightarrow Y)$ of continuous maps.

  The product topology is defined in order to make this true: a map into $X \times Y$ is continuous just when it is continuous in both its X-coordinate and its Y-coordinate.

Continuing the chapter headings:

2. Adjoints
3. Representables
4. Limits

These are all ways of formalizing the idea of 'universal property'. For instance, Examples 0.1.1 and 0.1.2 can most readily be described in terms of adjoints, Example 0.1.3 via representables, and Examples 0.1.5 and 0.1.6 via limits.

5. Adjoints, representables and limits

Here we compare the three approaches to universality. In principle, anything that can be described in one of the three formalisms can also be described in the others. It's a bit like cartesian and polar coordinates: anything that can be done in polar coordinates can in principle be done in cartesian coordinates, and vice versa, but it might be more graceful in one system than the other.

Depending on time and demand, we may then do one or both of the following chapters.

6. Monads

You may have noticed some similarities between the various branches of algebra that you've studied: linear algebra, group theory, ring theory, etc. For instance, in all of them there are notions of homomorphism, isomorphism, and subspace/subgroup/subring; in all of them there are Isomorphism Theorems. It turns out to be possible to study these common features systematically.

A monad is an ‘algebraic theory’. For example, there is one monad corresponding to the theory of groups, another for the theory of rings, and another for the theory of vector spaces over your favourite field.

7. Monoidal categories

There is an emerging subject known as ‘higher-dimensional algebra’. Ordinary algebra involves formulas—strings of symbols written along a (one-dimensional) line. In higher-dimensional algebra, these are replaced by diagrams living in two or more dimensions. Monoidal categories give the first taste of this subject, and are increasingly useful in modern mathematics.

Chapter 1

Categories, Functors and Natural Transformations

1.1 Categories

Definition 1.1.1 A **category** $\mathcal{A}$ consists of

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

  $$\begin{array}{ccc} \mathcal{A}(B,C) \times \mathcal{A}(A,B) & \longrightarrow & \mathcal{A}(A,C), \\ (g,f) & \longmapsto & g \circ f \end{array}$$

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

satisfying

- **associativity**: if $f \in \mathcal{A}(A,B)$, $g \in \mathcal{A}(B,C)$ and $h \in \mathcal{A}(C,D)$ then $(h \circ g) \circ f = h \circ (g \circ f)$
- **unit laws**: if $f \in \mathcal{A}(A,B)$ then $f \circ 1_A = f = 1_B \circ f$.

Notation 1.1.2 We often write:

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

People often write $\mathcal{A}(A,B)$ as $\mathrm{Hom}_{\mathcal{A}}(A,B)$ (longer) or $\mathrm{Hom}(A,B)$ (longer and less informative). 'Hom' stands for homomorphism, from one of the earliest examples of a category.

Remarks 1.1.3 a. The definition of category is set up so that any string

$$\begin{tikzcd} A_0 \arrow[r, "f_1"] & A_1 \arrow[r, "f_2"] & \cdots \arrow[r, "f_n"] & A_n \end{tikzcd}$$

of arrows in $\mathcal{A}$ (where $n \geq 0$) gives rise to exactly one arrow

$$\begin{tikzcd} A_0 \arrow[r] & A_n. \end{tikzcd}$$

It's safe to omit brackets and write this arrow as $f_n \cdots f_2 f_1$.

b. We speak of **commutative diagrams**: e.g. if

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

displays objects and arrows in some category, the diagram is said to **commute** if $gf = lkh$. More generally, a diagram is said to commute if whenever there are two paths from an object X to an object Y, the arrow obtained by composing along one path is equal to the arrow obtained by composing along the other.

c. The word 'class' means *roughly* the same as 'set'. A class is just a collection of things, with no structure on it. We'll come back to the set/class distinction later.

d. If $f \in \mathcal{A}(A,B)$, we call A the **domain** and B the **codomain** of f. We also write $A = \mathrm{dom}(f)$ and $B = \mathrm{cod}(f)$.

For the next few lectures we'll be discussing the interaction *between* categories, rather than what goes on *inside* them. We will, however, need the following definition.

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

Exercise 1.1.5 (Uniqueness of inverses) Show that if also $g' : B \longrightarrow A$ satisfies $g'f = 1_A$ and $fg' = 1_B$ then $g' = g$.

We call g the **inverse** of f and write $g = f^{-1}$. If there exists an isomorphism from A to B, we say that A and B are **isomorphic** and write $A \cong B$.

Examples 1.1.6 Categories of mathematical structures:

a. There is a category **Gp** whose objects are groups and whose maps are group homomorphisms. (Officially, I also have to tell you what the composition and identities are; but you can guess from the names.)

b. Similarly, there is a category **Ring** of rings and ring homomorphisms.

c. **Set**: sets and functions.

d. **Top**: topological spaces and continuous maps.

Exercise 1.1.7 What are the isomorphisms in all these categories? If you were teaching an introductory course on group theory, exactly how would you define *isomorphism* of groups?

Examples 1.1.8 Categories as mathematical structures:

a. A **monoid** is a 'group without inverses': precisely, 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.

   Consider a category $\mathcal{A}$ with just one object. It doesn't matter what letter or symbol we use to denote the object; let's call it A. Then $\mathcal{A}$ consists of a set (or class) $\mathcal{A}(A, A)$, an associative composition function

$$\circ : \mathcal{A}(A, A) \times \mathcal{A}(A, A) \longrightarrow \mathcal{A}(A, A),$$

   and a two-sided unit $1_A \in \mathcal{A}(A, A)$. In other words, a one-object category is the same thing as a monoid.

   (Beginners sometimes find this argument difficult, worrying about the question of what the object is called. It matters exactly as much as whether you choose x or y or t to denote some variable in an algebra problem: in other words, not at all. Later we'll define 'equivalence' of categories, and then we can prove a precise statement: the category of monoids is equivalent to the category of one-object categories.)

b. In particular, a group is the same thing as a one-object category is which every map is an isomorphism.

   The first time you meet the idea that a group is a kind of category, it's tempting to dismiss it as a coincidence or a trick. It's not: there's real content. To see this, suppose your education had been shuffled and you took a course on category theory before ever learning what a group was. Someone comes to you and says:

   'There are these structures called "groups", and the idea is this: a group is what you get when you collect together all the symmetries of a given thing.'

‘What do you mean by a “symmetry”?’ you ask.

‘Well, a symmetry of an object X is a way of transforming or mapping X into itself, in an invertible way.’

‘Oh,’ you reply, ‘that’s a special case of an idea I’ve met before. A category is the structure formed by *lots* of objects and mappings between them—not necessarily invertible. A group’s just the very special case where you’ve only got one object, and all the maps happen to be invertible.’

c. A **preorder** is a reflexive transitive relation. A **preordered set** $(S, \leq)$ is a set S equipped 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 $\subseteq$; $S = \mathbb{Z}$ and $a \leq b$ means that a divides b.

A preordered set can be regarded as a category $\mathcal{A}$ in which, for each $A, B \in \mathcal{A}$, there is at most one map from A to B. For given a preordered set $(S, \leq)$, we can define a category $\mathcal{A}$ by taking $\text{ob}(\mathcal{A}) = S$ and taking $\mathcal{A}(A, B)$ to have one element if $A \leq B$, and none otherwise; composition and identities are then uniquely determined. (Again, it doesn’t matter what letter we use to denote the map $A \longrightarrow B$, when $A \leq B$. You can think of this map as ‘the 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 $A = B$. This says that if $A \cong B$ in the corresponding category then $A = B$. Ordered sets are also called **partially ordered sets** or **posets**.)

The last example teaches us an important lesson:

Maps in a category need not be remotely like functions.

Constructions 1.1.9 New categories from old:

a. Every category $\mathcal{A}$ has an **opposite** (or **dual**) category $\mathcal{A}^{\text{op}}$, defined by ‘reversing arrows’. Formally, $\text{ob}(\mathcal{A}^{\text{op}}) = \text{ob}(\mathcal{A})$ and $\mathcal{A}^{\text{op}}(B, A) = \mathcal{A}(A, B)$; composition in $\mathcal{A}^{\text{op}}$ is the same as in $\mathcal{A}$ but back to front; the identities are the same. Clearly $(\mathcal{A}^{\text{op}})^{\text{op}} = \mathcal{A}$.

So, arrows $A \longrightarrow B$ in $\mathcal{A}$ correspond to arrows $B \longrightarrow A$ in $\mathcal{A}^{\text{op}}$. According to the definition above, if $f : A \longrightarrow B$ is an arrow in $\mathcal{A}$ then the corresponding arrow $B \longrightarrow A$ in $\mathcal{A}^{\text{op}}$ is also called f. Beginners sometimes find it less confusing to call it something else, such as f^{op} or f^*.

b. Every pair $(\mathcal{A}, \mathcal{B})$ of categories has a **product category** $\mathcal{A} \times \mathcal{B}$, in which

$$\begin{aligned} \text{ob}(\mathcal{A} \times \mathcal{B}) &= \text{ob}(\mathcal{A}) \times \text{ob}(\mathcal{B}), \\ (\mathcal{A} \times \mathcal{B})((A, B), (A', B')) &= \mathcal{A}(A, A') \times \mathcal{B}(B, B'). \end{aligned}$$

(I haven’t said how composition and identities are defined, but there’s only one sensible possibility. What is it?)

1.2 Functors

Category theory always asks: 'What are the maps?' We can ask this about categories themselves. A map between categories is called a 'functor'.

Definition 1.2.1 Let $\mathcal{A}$ and $\mathcal{B}$ be categories. A **functor** $F : \mathcal{A} \longrightarrow \mathcal{B}$ consists of

- a function
$$\mathrm{ob}(\mathcal{A}) \longrightarrow \mathrm{ob}(\mathcal{B}),$$
written $A \longmapsto F(A)$
- for each $A, A' \in \mathcal{A}$, a function
$$\mathcal{A}(A, A') \longrightarrow \mathcal{B}(F(A), F(A')),$$
written $f \longmapsto F(f)$,

such that

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

Remarks 1.2.2 a. The definition of functor is set up so that any diagram
$$A_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} A_n$$
in $\mathcal{A}$ (where $n \geq 0$) gives rise to exactly one map
$$F(A_0) \longrightarrow F(A_n)$$
in $\mathcal{B}$.

b. We're familiar with the idea that structures and the structure-preserving maps between them form a category (e.g. **Gp**, **Ring**, ...). In particular, this applies to categories and functors: there is a category **CAT** whose objects are categories and whose maps are functors. An important part of this statement is that functors can be composed.

Examples 1.2.3 a. **Forgetful functors** For instance,

- There is a functor $U : \mathbf{Gp} \longrightarrow \mathbf{Set}$ that forgets the group structure. If G is a group then $U(G)$ is the underlying set of G, and if $f : G \longrightarrow H$ is a group homomorphism then $U(f)$ is the function f itself.
- Let **Ab** be the category of abelian groups and their homomorphisms. Then there is an inclusion functor $U : \mathbf{Ab} \longrightarrow \mathbf{Gp}$ 'forgetting that the group is abelian'.

The first functor forgets the *structure* of the objects, but the second forgets a *property*. Nevertheless, it turns out to be convenient to use the same word, 'forgetful', for both types of functor. (There's no precise definition of forgetful functor; it's an informal term.)

b. **Free functors** For instance,

- Let k be a field, and write $\mathbf{Vect}_k$ for the category of vector spaces over k and linear maps. There is a functor $F : \mathbf{Set} \longrightarrow \mathbf{Vect}_k$, defined on objects by taking $F(S)$ to be a vector space with basis S: so
$$F(S) = \{\text{formal } k\text{-linear combinations of elements of } S\}.$$
(For instance, if S is a finite set with n elements then $F(S) \cong k^n$.)
- Similarly, there is a functor $F : \mathbf{Set} \longrightarrow \mathbf{Gp}$ for which $F(S)$ is the free group on S (which we will meet again later). The elements of $F(S)$ are formal words such as $x^{-1}yx^2zy^{-3}$ where $x, y, z \in S$.

c. **Homology and homotopy** For each $n \in \mathbb{N}$, there is a functor $H_n : \mathbf{Top} \longrightarrow \mathbf{Ab}$ assigning to each space its nth homology (in whatever sense of the word you prefer).

Let $\mathbf{Top}_*$ be the category of topological spaces equipped with a basepoint, and continuous basepoint-preserving maps. Then there is a functor $\pi_1 : \mathbf{Top}_* \longrightarrow \mathbf{Gp}$ assigning to each space with basepoint its fundamental group (= first homotopy group).

d. **Functors between monoids** Let G and H be groups (or more generally, monoids), regarded as one-object categories $\mathcal{G}$ and $\mathcal{H}$. A functor $F : \mathcal{G} \longrightarrow \mathcal{H}$ must send the unique object of $\mathcal{G}$ to the unique object of $\mathcal{H}$, so it is determined by its effect on maps. Hence F amounts to a function $F : G \longrightarrow H$ such that $F(g' \circ g) = F(g') \circ F(g)$ for all $g', g \in G$, and $F(1) = 1$. In other words, a functor $\mathcal{G} \longrightarrow \mathcal{H}$ is just a homomorphism $G \longrightarrow H$.

e. **Actions and representations** Let G be a group (or monoid), regarded as a one-object category $\mathcal{G}$. A functor $F : \mathcal{G} \longrightarrow \mathbf{Set}$ consists of a set S (the value of F at the unique object of $\mathcal{G}$) together with, for each $g \in G$, a functor $F(g) : S \longrightarrow S$, satisfying the functoriality axioms. Writing $(F(g))(s) = g \cdot s$ whenever $g \in G$ and $s \in S$, we see that F amounts to a set S together with a function

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

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

Similarly, a functor $\mathcal{G} \longrightarrow \mathbf{Vect}_k$ is exactly a k-linear representation of G. If you don't know what a representation of a group is, take that as the definition.

f. **Functors between ordered sets** If A and B are (pre)ordered sets then a functor between the corresponding categories is just an **order-preserving map**, that is, a function f such that $a \leq a' \Rightarrow f(a) \leq f(a')$.

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

Definition 1.2.4 Let $\mathcal{A}$ and $\mathcal{B}$ be categories. A **contravariant functor** from $\mathcal{A}$ to $\mathcal{B}$ is a functor $\mathcal{A}^{\mathrm{op}} \longrightarrow \mathcal{B}$.

(We don't write 'a contravariant functor $\mathcal{A} \longrightarrow \mathcal{B}$': that would be confusing.)

Functors $\mathcal{C} \longrightarrow \mathcal{D}$ correspond one-to-one with functors $\mathcal{C}^{\mathrm{op}} \longrightarrow \mathcal{D}^{\mathrm{op}}$, so a contravariant functor from $\mathcal{A}$ to $\mathcal{B}$ can also be described as a functor $\mathcal{A} \longrightarrow \mathcal{B}^{\mathrm{op}}$. You might find this description more natural than the one in the definition; it doesn't matter an enormous amount, but in the long run the $\mathcal{A}^{\mathrm{op}} \longrightarrow \mathcal{B}$ definition makes life easier.

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

Examples 1.2.5 a. **Functions on a space** You can tell a lot about a space by examining the functions on it. (The importance of this idea in 20th- and 21st-Century mathematics can hardly be overstated.) For example, given a topological space X, let $C(X)$ be the ring of continuous real-valued functions on X, where the ring operations are defined pointwise (e.g. if $p_1, p_2 : X \longrightarrow \mathbb{R}$ then $(p_1 + p_2)(x) = p_1(x) + p_2(x)$, for $x \in X$). A continuous map $f : X \longrightarrow Y$ induces a ring homomorphism $C(f) : C(Y) \longrightarrow C(X)$, defined on $q \in C(Y)$ 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. (You'll sometimes hear people say that 'algebra is dual to geometry'.) After checking some axioms, we conclude that C is a contravariant functor from **Top** to **Ring**.

b. **Functions on a vector space** Taking duals of vector spaces gives a contravariant functor $(\ )^*$ from $\mathbf{Vect}_k$ to itself:

$$(\ )^* : \mathbf{Vect}_k^{\mathrm{op}} \longrightarrow \mathbf{Vect}_k$$

defined by

$$\begin{tikzcd}
V \arrow[r, phantom, "\longmapsto"] \arrow[d, "f"'] & V^* \\
\arrow[r, phantom, "\longmapsto"] & \\
W \arrow[r, phantom, "\longmapsto"] & W^*. \arrow[u, "f^*"']
\end{tikzcd}$$

c. **Cohomology** For each $n \in \mathbb{N}$, there is a functor $H^n : \mathbf{Top}^{\mathrm{op}} \longrightarrow \mathbf{Ab}$ assigning to each space its nth cohomology.

d. **Right actions** Let G be a group. A functor $G^{\mathrm{op}} \longrightarrow \mathbf{Set}$ is a *right* G-set.

Definition 1.2.6 Let $\mathcal{A}$ be a category. A **presheaf** on $\mathcal{A}$ is a functor $\mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$.

The definition is motivated by the following special case. Let X be a topological space and let $\mathbf{Open}(X)$ be the poset of open subsets of X, ordered by inclusion. A **presheaf** on the space X is a presheaf on the category $\mathbf{Open}(X)$. Presheaves, and a certain class of presheaves called sheaves, are particularly important in algebraic geometry.

Definition 1.2.7 A functor $F : \mathcal{A} \longrightarrow \mathcal{B}$ is **faithful** (respectively, **full**) if for each $A, A' \in \mathcal{A}$, the function

$$\begin{array}{ccc}
\mathcal{A}(A, A') & \longrightarrow & \mathcal{B}(F(A), F(A')) \\
f & \longmapsto & F(f)
\end{array}$$

is injective (respectively, surjective).

Example 1.2.8 A **subcategory** $\mathcal{S}$ of a category $\mathcal{C}$ consists of a subclass $\mathrm{ob}(\mathcal{S})$ of $\mathrm{ob}(\mathcal{C})$ together with, for each $S, S' \in \mathrm{ob}(\mathcal{S})$, a subclass $\mathcal{S}(S, S')$ of $\mathcal{C}(S, S')$, such that $\mathcal{S}$ is closed under composition and identities. The inclusion of $\mathcal{S}$ into $\mathcal{C}$ defines a faithful functor $\mathcal{S} \longrightarrow \mathcal{C}$. A subcategory is called **full** when the inclusion functor is full. For instance, **Ab** is a full subcategory of **Gp**. To specify a full subcategory of a known category $\mathcal{C}$, you only need to say what its objects are.

1.3 Natural transformations

Further resource for this section: The Catsters, Natural Transformations, http://youtube.com/TheCatsters

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

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

To see how this might work, let's consider a special case. Let $\mathcal{A}$ be the category whose objects are the natural numbers $0, 1, 2, \ldots$, and with no maps except for the identity on each object (which, as a category, it's obliged to have). A functor F from $\mathcal{A}$ to another category $\mathcal{B}$ is simply a sequence $(F_0, F_1, F_2, \ldots)$ of objects of $\mathcal{B}$. (If you like, think of F as 'one object of $\mathcal{B}$ every day from now to eternity'.) Let G be another functor from $\mathcal{A}$ to $\mathcal{B}$, consisting of another sequence $(G_0, G_1, \ldots)$ of objects of $\mathcal{B}$. What might a 'map from F to G' sensibly be?

One answer is that a map α from F to G should consist of a sequence

$$(F_0 \xrightarrow{\alpha_0} G_0,\ F_1 \xrightarrow{\alpha_1} G_1,\ \ldots)$$

of maps in $\mathcal{B}$. And indeed, that's exactly what a natural transformation $\alpha : F \longrightarrow G$ is—for this particular category $\mathcal{A}$.

This suggests that in the general case, a natural transformation between functors $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, shift right, "G"'] & \mathcal{B} \end{tikzcd}$ might consist of maps $\alpha_A : F(A) \longrightarrow G(A)$, one for each $A \in \mathcal{A}$. The category $\mathcal{A}$ will usually be less trivial than in our example, and we will then demand some kind of compatibility between the maps α_A and the structure of the categories concerned.

Definition 1.3.1 Let $\mathcal{A}$ and $\mathcal{B}$ be categories and let $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, shift right, "G"'] & \mathcal{B} \end{tikzcd}$ be functors. A **natural transformation** $\alpha : F \longrightarrow G$ is a family $(F(A) \xrightarrow{\alpha_A} G(A))_{A \in \mathcal{A}}$ of maps in $\mathcal{B}$ such that for every map $A \xrightarrow{f} A'$ in $\mathcal{A}$, the diagram

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, "\alpha_A"'] & F(A') \arrow[d, "\alpha_{A'}"] \\ G(A) \arrow[r, "G(f)"'] & G(A') \end{tikzcd} \qquad (1{:}1)$$

commutes.

Remarks 1.3.2 a. The definition of natural transformation is set up so that any map $A \xrightarrow{f} A'$ in $\mathcal{A}$ gives rise to exactly one map $F(A) \longrightarrow G(A')$ in $\mathcal{B}$. When $f = 1_A$, this map is α_A (called the **component of** α **at** A). For general f, it is the diagonal of the square (1:1) (called the **naturality square** of α at f); 'exactly one' implies that the square commutes.

b. A functor is a map of categories; a natural transformation is a map of functors. (It goes no further!) We often write

$$\begin{tikzcd}
\mathcal{A} \arrow[r, bend left=50, "F"{name=U}] \arrow[r, bend right=50, "G"'{name=D}] & \mathcal{B} \arrow[Rightarrow, from=U, to=D, "\alpha"]
\end{tikzcd}$$

to indicate a natural transformation as above.

Examples 1.3.3 a. A category is **discrete** if it contains no maps except for identities. Such a category amounts to just a class of objects. For instance, the category $\mathcal{A}$ at the beginning of this section is the discrete category corresponding to the set $\mathbb{N}$.

Let $\mathcal{A}$ be a discrete category and let $F, G : \mathcal{A} \longrightarrow \mathcal{B}$ be functors to another category $\mathcal{B}$. Then F and G are just families $(F(A))_{A\in\mathcal{A}}$ and $(G(A))_{A\in\mathcal{A}}$ of objects of $\mathcal{B}$. A natural transformation $\alpha : F \longrightarrow G$ is just a family $(F(A) \xrightarrow{\alpha_A} G(A))_{A\in\mathcal{A}}$ of maps in $\mathcal{B}$. *A priori*, this family must satisfy the naturality axiom (1:1) for every map f in $\mathcal{A}$; but the only maps in $\mathcal{A}$ are identities, and when f is an identity the naturality axiom holds automatically.

b. Recall from 1.1.8(a) that a group (or more generally, a monoid) G can be regarded as a one-object category. Previously we used $\mathcal{G}$ to denote the category corresponding to the group G; from now on we will use G to denote them both. Also recall from 1.2.3(e) that a functor from G to **Set** is nothing but a left G-set. Take two G-sets, S and T. What is a natural transformation between (the functors corresponding to) S and T? Since G has just one object, a natural transformation from S to T consists of just one map in **Set**, satisfying some axioms. Precisely, a natural transformation from S to T is a function $\alpha : S \longrightarrow T$ 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**).

c. Fix a natural number n. For any commutative ring R, the $n \times n$ matrices with entries in R form a monoid $M_n(R)$ under multiplication. Any ring homomorphism $R \longrightarrow S$ induces a monoid homomorphism $M_n(R) \longrightarrow$

$M_n(S)$, and we have a functor $M_n : \mathbf{CRing} \longrightarrow \mathbf{Mon}$ 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 $U : \mathbf{CRing} \longrightarrow \mathbf{Mon}$.

Now, any $n \times n$ 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 $\det_R : M_n(R) \longrightarrow U(R)$ is a monoid homomorphism. This suggests that determinant might define a natural transformation

$$\begin{tikzcd} \mathbf{CRing} \arrow[r, bend left=50, "M_n", ""{name=T, below}] \arrow[r, bend right=50, "U"', ""{name=B, above}] & \mathbf{Mon}, \arrow[Rightarrow, from=T, to=B, "\det"] \end{tikzcd}$$

and indeed it does. The fact that the naturality squares commute (check!) reflects the fact that determinant is defined in the same way for all rings. In general, the naturality axiom (1:1) is meant to capture the idea that the family $(\alpha_A)_{A \in \mathcal{A}}$ is defined in a uniform way across all $A \in \mathcal{A}$.

Natural transformations are a kind of map, so we would expect to be able to compose them. This is simple: given natural transformations

$$\begin{tikzcd} \mathcal{A} \arrow[r, bend left=70, "F", ""{name=F, below}] \arrow[r, "G" description, ""{name=G1, above}, ""{name=G2, below}] \arrow[r, bend right=70, "H"', ""{name=H, above}] & \mathcal{B}, \arrow[Rightarrow, from=F, to=G1, "\alpha"] \arrow[Rightarrow, from=G2, to=H, "\beta"] \end{tikzcd}$$

there is a composite natural transformation

$$\begin{tikzcd} \mathcal{A} \arrow[r, bend left=50, "F", ""{name=T, below}] \arrow[r, bend right=50, "H"', ""{name=B, above}] & \mathcal{B} \arrow[Rightarrow, from=T, to=B, "\beta\circ\alpha"] \end{tikzcd}$$

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

$$\begin{tikzcd} \mathcal{A} \arrow[r, bend left=50, "F", ""{name=T, below}] \arrow[r, bend right=50, "F"', ""{name=B, above}] & \mathcal{A} \arrow[Rightarrow, from=T, to=B, "1_F"] \end{tikzcd}$$

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

Examples 1.3.4 a. Let 2 be the discrete category with two objects. A functor from 2 to a category $\mathcal{B}$ is a pair of objects of $\mathcal{B}$, and a natural transformation is a pair of maps. The functor category $[2, \mathcal{B}]$ is therefore isomorphic to the product category $\mathcal{B} \times \mathcal{B}$ (defined in 1.1.9(b)). This fits well with the alternative notation for the functor category, $\mathcal{B}^2$.

b. If G is a group then $[G, \mathbf{Set}]$ is the category of left G-sets and $[G^{\mathrm{op}}, \mathbf{Set}]$ is the category of right G-sets (see 1.2.5(d)).

c. Let A and B be ordered sets, regarded as categories (1.1.8(c), 1.2.3(f)). Then the category $[A, B]$ also corresponds to an ordered set: namely, the set of order-preserving maps from A to B, with $f \leq g$ whenever $f(a) \leq g(a)$ for all $a \in A$. (So there are three orders involved here: one on A, one on B, and one on $[A, B]$.)

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 applies, in particular, when the category concerned is a functor category. In other words, given two functors $F, G : \mathcal{A} \longrightarrow \mathcal{B}$, we usually do not care whether they are equal (which would imply that the objects $F(A)$ and $G(A)$ of $\mathcal{B}$ are equal for all $A \in \mathcal{A}$). What really matters is whether they are naturally isomorphic.

Definition 1.3.5 Let $\mathcal{A}$ and $\mathcal{B}$ be categories. A **natural isomorphism** of functors from $\mathcal{A}$ to $\mathcal{B}$ is an isomorphism in $[\mathcal{A}, \mathcal{B}]$.

An equivalent form of the definition is often useful:

Lemma 1.3.6 *Let* $\begin{tikzcd} \mathcal{A} \arrow[r, bend left, "F", ""{name=U, below}] \arrow[r, bend right, "G"', ""{name=D}] & \mathcal{B} \arrow[Rightarrow, from=U, to=D, "\alpha"] \end{tikzcd}$ *be a natural transformation. Then α is a natural isomorphism if and only if the map $\alpha_A : F(A) \longrightarrow G(A)$ is an isomorphism for all $A \in \mathcal{A}$.*

This is like the fact that the inverse of a bijective group homomorphism is also a homomorphism.

Proof Exercise. □

Take functors $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, shift right, "G"'] & \mathcal{B} \end{tikzcd}$. Recall that we use $\cong$ to denote isomorphism of two objects of a category; in particular, when the category is $[\mathcal{A}, \mathcal{B}]$, we use $F \cong G$ to mean that there is a natural isomorphism between F and G. We sometimes use the phrase '$F(A) \cong G(A)$ **naturally in** A' to mean the same thing.

Examples 1.3.7 a. Let $F, G : \mathcal{A} \longrightarrow \mathcal{B}$ be functors from a discrete category $\mathcal{A}$ to a category $\mathcal{B}$. Then $F \cong G$ if and only if $F(A) \cong G(A)$ for all $A \in \mathcal{A}$.

b. Let **FDVect** be the category of finite-dimensional vector spaces over some chosen field. The dual vector space construction defines a contravariant functor from **FDVect** to itself; the double dual construction defines a covariant functor from **FDVect** to itself. Moreover, we have for each $V \in \mathbf{FDVect}$ a canonical isomorphism $\varepsilon_V : V \longrightarrow V^{**}$ (evaluation), giving a natural transformation

$$\begin{tikzcd}
\mathbf{FDVect} \arrow[r, bend left=50, "1_{\mathbf{FDVect}}", ""{name=U, below}] \arrow[r, bend right=50, "(\ )^{**}"', ""{name=D, above}] & \mathbf{FDVect} \arrow[Rightarrow, from=U, to=D, "\varepsilon"]
\end{tikzcd}$$

from the identity functor to the double dual functor. By Lemma 1.3.6, ε is a natural isomorphism. So $1_{\mathbf{FDVect}} \cong (\ )^{**}$; put differently, $V \cong V^{**}$ naturally in $V \in \mathbf{FDVect}$.

c. There exist categories and functors $\mathcal{A} \overset{F}{\underset{G}{\rightrightarrows}} \mathcal{B}$ such that $F(A) \cong G(A)$ for all A, but not *naturally* in A. On Sheet 2 there is an example from combinatorics.

Two elements of a set are either equal or not. Two objects of a category might be equal, or isomorphic, or not. But between categories themselves, even isomorphism is an unreasonably strict relation: for if $\mathcal{A} \cong \mathcal{B}$ then there are functors

$$\begin{tikzcd}
\mathcal{A} \arrow[r, shift left, "F"] & \mathcal{B} \arrow[l, shift left, "G"]
\end{tikzcd} \tag{1:2}$$

such that (among other things) $G(F(A))$ is *equal* to A for all $A \in \mathcal{A}$. The most useful notion of sameness of categories, called 'equivalence', is looser than isomorphism.

Definition 1.3.8 An **equivalence** between categories $\mathcal{A}$ and $\mathcal{B}$ consists of a pair (1:2) of functors between them, together with natural isomorphisms

$$\eta : 1_{\mathcal{A}} \longrightarrow G \circ F, \qquad \varepsilon : F \circ G \longrightarrow 1_{\mathcal{B}}.$$

In this situation, we say that $\mathcal{A}$ and $\mathcal{B}$ are **equivalent** and write $\mathcal{A} \simeq \mathcal{B}$. We also say that the functors F and G are **equivalences**.

You might wonder about the choice of directions for η and ε. Of course, since they're isomorphisms, it doesn't make much difference. When we come to adjunctions, you'll see the reason for this choice.

Warning 1.3.9 The symbol $\cong$ is used for isomorphism of objects of categories, and in particular for isomorphism of categories (which are objects of **CAT**). The symbol $\simeq$ is used for equivalence of categories.

Exercise 1.3.10 Show that equivalence of categories is an equivalence relation. (Not quite as obvious as it looks. If you get stuck, try again when you've read to the end of the section.)

There is a very useful alternative description of when a functor is an equivalence. First we need a definition.

Definition 1.3.11 A functor $F : \mathcal{A} \longrightarrow \mathcal{B}$ is **essentially surjective on objects** if for all $B \in \mathcal{B}$, there exists $A \in \mathcal{A}$ such that $F(A) \cong B$.

Proposition 1.3.12 *A functor is an equivalence if and only if it is full, faithful and essentially surjective on objects.*

Proof Exercise. For 'if', begin by choosing for each $B \in \mathcal{B}$ an object $A \in \mathcal{A}$ such that $F(A) \cong B$. □

Examples 1.3.13

a. Let $\mathcal{A}$ be any category, and let $\mathcal{B}$ be any full subcategory containing at least one object from every isomorphism class of $\mathcal{A}$. Then the inclusion is full, faithful (like any inclusion of subcategories), and essentially surjective on objects; hence $\mathcal{B} \simeq \mathcal{A}$.

   So if you 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. Conversely, if you take a category and remove all but one of the objects in each isomorphism class, the slimmed-down version is equivalent to the original.

   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 $\mathcal{B}$ be the full subcategory of **FinSet** with objects $\mathbf{0}, \mathbf{1}, \ldots$. (In this course, 0 is a natural number.) Then $\mathcal{B} \simeq$ **FinSet**, even though $\mathcal{B}$ is in some sense much smaller than **FinSet**.

b. An equivalence of the form $\mathcal{A}^{\mathrm{op}} \simeq \mathcal{B}$ is sometimes called a **duality** between $\mathcal{A}$ and $\mathcal{B}$; one says that $\mathcal{A}$ is **dual** to $\mathcal{B}$. There are many famous dualities in which $\mathcal{A}$ is a category of algebras and $\mathcal{B}$ is family of spaces: recall the motto 'algebra is dual to geometry' from 1.2.5(a).

   Here are some examples, clearly not for examination.

   - In algebraic geometry, there is a duality, for any algebraically closed field k, between the category of finitely generated reduced k-algebras and the category of affine varieties over k. (A variety is an algebraic geometer's notion of a 'space'.)
   - Gelfand duality: the category of commutative C^*-algebras is dual to the category of compact Hausdorff spaces. (C^*-algebras are certain algebraic structures important in functional analysis.)

- Stone duality: the category of Boolean algebras is dual to the category of totally disconnected compact Hausdorff spaces.
- Pontryagin duality: the category of locally compact abelian topological groups is dual to itself. (The words 'topological group' tell you that both sides of the duality are algebraic *and* geometric.) Surprising as it may seem, Pontryagin duality is an abstraction of the properties of the Fourier transform, from complex analysis.

c. There's no point in having a category of structured objects if the maps don't respect that structure. For instance, let $\mathcal{A}$ 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 $U : \mathcal{A} \longrightarrow \mathbf{Set}_{\neq\emptyset}$ is full and faithful. It's a fact (not very profound) that every nonempty set admits at least one group structure, so U is essentially surjective on objects. Hence U is an equivalence and $\mathcal{A} \simeq \mathbf{Set}_{\neq\emptyset}$.

Remarks 1.3.14 Here's a kind of review of the course so far. We've defined

- categories
- functors between categories
- natural transformations between functors
- composition of functors

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

  (and accompanying identities)
- composition of natural transformations

$$\begin{tikzcd}[column sep=large] \cdot \arrow[r, bend left=60] \arrow[r] \arrow[r, bend right=60] \arrow[r, phantom, shift left=2.5ex, "\Downarrow"] \arrow[r, phantom, shift right=2.5ex, "\Downarrow"] & \cdot \end{tikzcd}$$

  (and accompanying identities).

This composition of natural transformations is sometimes called **vertical composition**. There's also a **horizontal composition**, applicable to natural transformations of the following type:

$$\begin{tikzcd} \mathcal{A} \arrow[r, bend left=50, "F"] \arrow[r, bend right=50, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathcal{A}' \arrow[r, bend left=50, "F'"] \arrow[r, bend right=50, "G'"'] \arrow[r, phantom, "\Downarrow\alpha'"] & \mathcal{A}'' \end{tikzcd}$$

and giving a natural transformation

$$\begin{tikzcd}
\mathcal{A} \arrow[r, bend left=50, "F' \circ F"{name=U}] \arrow[r, bend right=50, "G' \circ G"'{name=D}] & \mathcal{A}'' \arrow[Rightarrow, from=U, to=D]
\end{tikzcd}$$

traditionally written $\alpha' * \alpha$. The component of $\alpha' * \alpha$ at $A \in \mathcal{A}$ is 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 either as $\alpha'_{G(A)} \circ F'(\alpha_A)$ or as $G'(\alpha_A) \circ \alpha'_{F(A)}$: it doesn't matter which, as they're equal.

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

$$\begin{tikzcd}
\mathcal{A} \arrow[r, "F"] & \mathcal{A}' \arrow[r, bend left=50, "F'"{name=U}] \arrow[r, bend right=50, "G'"'{name=D}] & \mathcal{A}'' \arrow[Rightarrow, from=U, to=D, "\alpha'"] & \text{gives rise to} & \mathcal{A} \arrow[r, bend left=50, "F' \circ F"{name=U2}] \arrow[r, bend right=50, "G' \circ F"'{name=D2}] & \mathcal{A}'' \arrow[Rightarrow, from=U2, to=D2, "\alpha' F"]
\end{tikzcd}$$

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

$$\begin{tikzcd}
\mathcal{A} \arrow[r, bend left=50, "F"{name=U}] \arrow[r, bend right=50, "G"'{name=D}] & \mathcal{A}' \arrow[r, "F'"] & \mathcal{A}'' \arrow[Rightarrow, from=U, to=D, "\alpha"] & \text{gives rise to} & \mathcal{A} \arrow[r, bend left=50, "F' \circ F"{name=U2}] \arrow[r, bend right=50, "F' \circ G"'{name=D2}] & \mathcal{A}'' \arrow[Rightarrow, from=U2, to=D2, "F' \alpha"]
\end{tikzcd}$$

where $(F'\alpha)_A = F'(\alpha_A)$. These constructions are needed to solve Exercise 1.3.10.

Vertical and horizontal composition interact well: natural transformations

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

obey the **interchange law**,

$$(\beta' \circ \alpha') * (\beta \circ \alpha) = (\beta' * \beta) \circ (\alpha' * \alpha) : F' \circ F \longrightarrow H' \circ H.$$

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

All of this enables us to construct a functor

$$[\mathcal{A}', \mathcal{A}''] \times [\mathcal{A}, \mathcal{A}'] \longrightarrow [\mathcal{A}, \mathcal{A}'']$$

given on objects by $(F', F) \longmapsto F' \circ F$ and on maps by $(\alpha', \alpha) \longmapsto \alpha' * \alpha$. In particular, if $F' \cong G'$ and $F \cong G$ then $F' \circ F \cong G' \circ G$.

The last couple of pages are filled with diagrams containing not only objects (0-dimensional) and arrows $\rightarrow$ (1-dimensional), but also double-arrows $\Rightarrow$ sweeping out a 2-dimensional region between arrows. What we are doing, implicitly, is called 2-category theory. If you're really serious about categories, you have to get serious about 2-categories. And if you're really serious about 2-categories, you have to get serious about 3-categories... and before you know it, you're studying ∞-categories. But in this course, we will be content to remain on the first rung or two of this dizzying ladder.

Chapter 2

Adjoints

> *Further resource for this chapter: The Catsters, Adjunctions, http://youtube.com/TheCatsters*

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

$$\boxed{\textbf{Adjoint functors arise everywhere.}}$$

He demonstrates this by giving examples of adjoint functors from diverse parts of mathematics. We'll look at some of them here. We'll also try to understand adjointness by coming at it from three different angles, each of which carries its own intuition, and then proving that the three approaches are equivalent.

2.1 Basics

Definition 2.1.1 Let $\mathcal{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{B}$ be categories and functors. If

$$\mathcal{B}(F(A), B) \cong \mathcal{A}(A, G(B)) \tag{2:1}$$

naturally in $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we say that F is **left adjoint** to G, and G is **right adjoint** to F, and write $F \dashv G$. An **adjunction** between F and G is a choice of natural isomorphism (2:1).

Remarks 2.1.2 a. This says that for any $A \in \mathcal{A}$ and $B \in \mathcal{B}$, a map $F(A) \longrightarrow B$ is essentially the same thing as a map $A \longrightarrow G(B)$. The correspondence is denoted by a horizontal bar:

$$\begin{array}{ccc} (FA \xrightarrow{g} B) & \longmapsto & (A \xrightarrow{\overline{g}} GB), \\ (FA \xrightarrow{\overline{f}} B) & \longleftarrow\!\!\shortmid & (A \xrightarrow{f} GB), \end{array}$$

so that $\overline{\overline{f}} = f$ and $\overline{\overline{g}} = g$. We call $\overline{q}$ the **transpose** of q. I'll explain the meaning of 'naturally' later; for now, interpret it in the usual informal sense of 'defined without making any random choices'.

b. Not only do adjoint functors arise everywhere; better, whenever you see a pair of functors $\mathcal{A} \rightleftarrows \mathcal{B}$, there's an excellent chance that they're 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 may not know what either of those things is, but she goes 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. What you should do at this point is bet her money that one process is adjoint to the other. This trick almost always works.

c. A given functor G 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 will prove this later.

   You might ask 'what do we gain from knowing that two functors are adjoint?' The uniqueness is a crucial part of the answer. Let's return to the example of (b). It would take you only a few minutes to learn what Lie algebras and 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'll find in most algebra books takes much longer to understand, and to get there you'd first have to learn some preliminary concepts. However, you can bypass all that by just knowing that F is the left adjoint of G. Since G can only have *one* left adjoint (up to isomorphism), this characterizes F completely. In a sense, it tells you all you need to know.

Examples 2.1.3 (Algebra: free $\dashv$ forgetful) Any forgetful functor between categories of algebraic structures has a left adjoint. 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 (as in 1.2.3(a)) and F is the free functor of 1.2.3(b). Adjointness says that given a set S and a vector space V, a linear map $F(S) \longrightarrow V$ is essentially the same thing as a function $S \longrightarrow U(V)$. We saw this in Example 0.1.1.

   To analyze this example more precisely, first we need to get straight exactly what the functors F and U are.

   Firstly, F sends a set S to some vector space with basis S. Any two such vector spaces are isomorphic, but it's perhaps not obvious that there is any such vector space at all, so we have to construct one. A construction of sorts is in 1.2.3(b), where $F(S)$ is described as the set of 'formal k-linear combinations of elements of S' (with an unmentioned vector space structure). I will now make this more precise. A formal k-linear combination of elements of S is an expression $\sum_{s\in S}\lambda_s s$ where each λ_s is a scalar and λ_s is only nonzero for finitely many values of $s \in S$. (You can only do *finite* sums in a vector space.) Any two elements of $F(S)$ can be added—

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

—and there is a scalar multiplication on $F(S)$—

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

In this way, $F(S)$ becomes a vector space.

(If you want to be 100% precise and avoid talking about 'expressions', you can define $F(S)$ to be the set of all functions $\lambda : S \longrightarrow k$ such that $\{s \in S \mid \lambda(s) \neq 0\}$ is finite. Think of λ as corresponding to the 'expression' $\sum_{s\in S}\lambda(s)s$; this idea should suggest how to define the vector space structure on $F(S)$.)

Secondly, U sends a vector space V to its underlying set, or set of elements, $U(V)$. The symbols for forgetful functors are often omitted, by a slight abuse of notation: here, this would mean writing 'V' to mean $U(V)$. In this example I'll be fussy and keep the U in.

Now let's check adjointness in detail. Given a linear map $g : F(S) \longrightarrow V$, we may define a function $\overline{g} : S \longrightarrow U(V)$ by $\overline{g}(s) = g(s)$ for all $s \in S$. This gives a function

$$\begin{array}{ccc}\mathbf{Vect}_k(F(S),V) & \longrightarrow & \mathbf{Set}(S,U(V)) \\ g & \longmapsto & \overline{g}.\end{array}$$

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

$$\begin{array}{ccc}\mathbf{Set}(S,U(V)) & \longrightarrow & \mathbf{Vect}_k(F(S),V) \\ f & \longmapsto & \overline{f}.\end{array}$$

These two functions 'bar' are mutually inverse: if $g : F(S) \longrightarrow V$ is a linear map then

$$\overline{\overline{g}}(\sum\lambda_s s) = \sum\lambda_s\overline{g}(s) = \sum\lambda_s g(s) = g(\sum\lambda_s s)$$

for all $\sum\lambda_s s \in F(S)$, so $\overline{\overline{g}} = g$, and if $f : S \longrightarrow U(V)$ is a function then

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

for all $s \in S$, so $\overline{\overline{f}} = f$. We therefore have a 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.

b. Similarly, there is an adjunction

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

where F and U are the free and forgetful functors of 1.2.3(a,b).

The free group functor is quite tricky to construct explicitly. Later we will meet 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; compare Remark 2.1.2(c).

c. There is an adjunction

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

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

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

where η is the natural map from G to its quotient G_{ab} and A is any abelian group.

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

d. There are adjunctions

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

between the categories of groups and monoids. The middle functor U is inclusion. If M is a monoid then $R(M)$ is the submonoid of M consisting of all the invertible elements. Again, the left adjoint is 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 will guarantee the existence of this adjoint.

Gp is both a **reflective** and a **coreflective** subcategory of **Mon**. This means that the inclusion functor $\mathbf{Gp} \hookrightarrow \mathbf{Mon}$ has both a left and a right adjoint. Another example: **Ab** is a reflective subcategory of **Gp**.

e. Let **Field** be the category of fields (with ring homomorphisms as maps). The forgetful functor $\mathbf{Field} \longrightarrow \mathbf{Set}$ does *not* have a left adjoint. The theory of fields is unlike other algebraic theories, because the operation $x \longmapsto x^{-1}$ is not defined for all x (only for $x \neq 0$).

Example 2.1.4 (Spaces and sets) There are adjunctions

$$\begin{tikzcd}
\mathbf{Top} \arrow[d, "U" description] \\
\mathbf{Set} \arrow[u, shift left=6ex, "D"] \arrow[u, phantom, shift left=3ex, "\dashv"] \arrow[u, phantom, shift right=3ex, "\dashv"] \arrow[u, shift right=6ex, "I"']
\end{tikzcd}$$

where U sends a space to its set of points, D equips a set with the discrete topology (all subsets are open), and I equips a set with the indiscrete topology (only the empty set and the whole set are open).

Example 2.1.5 (Products and exponentials) Let Y be a set. For any set X, we may form the product set $X \times Y$, and this defines a functor

$$- \times Y : \mathbf{Set} \longrightarrow \mathbf{Set}.$$

(The $-$ is a blank or empty slot for a variable to go into. So the value of this functor at a set X is $X \times Y$.) Also, for any set Z, we may form the set Z^Y of functions from Y to Z, and this defines a functor

$$(-)^Y : \mathbf{Set} \longrightarrow \mathbf{Set}.$$

Moreover, there is a natural bijection

$$\mathbf{Set}(X \times Y, Z) \cong \mathbf{Set}(X, Z^Y)$$

for any sets X, Y, Z. (Why?) So there is an adjunction

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

Example 2.1.6 (Initial and terminal objects) Let $\mathcal{A}$ be a category. An object $I \in \mathcal{A}$ is **initial** if for every $A \in \mathcal{A}$, there is exactly one map $I \longrightarrow A$. For example, the trivial group is an initial object of **Gp** and $\mathbb{Z}$ is an initial object of **Ring** (the category of rings *with identity*). If I and I' are initial objects of the same category then $I \cong I'$, and in fact there is a *unique* isomorphism $I \longrightarrow I'$ (exercise).

Initial objects can be described as adjoints. Let $\mathbf{1}$ be the category with just one object and one arrow (necessarily the identity). Let $\mathcal{A}$ be any category. There is precisely one functor $\mathcal{A} \longrightarrow \mathbf{1}$, and a functor $\mathbf{1} \longrightarrow \mathcal{A}$ is just an object of $\mathcal{A}$. A left adjoint to $\mathcal{A} \longrightarrow \mathbf{1}$ is exactly an initial object of $\mathcal{A}$.

Similarly, an object T of a category $\mathcal{A}$ is **terminal** if for every $A \in \mathcal{A}$, there is exactly one map $A \longrightarrow T$. For example, the trivial group is also a terminal object of **Gp**, and **1** is itself terminal in **Cat**. A right adjoint to the unique functor $\mathcal{A} \longrightarrow \mathbf{1}$ is exactly a terminal object of $\mathcal{A}$.

Remark 2.1.7 Every concept, result and proof in category theory has a **dual**, obtained by reversing all the arrows. For example, the dual concept to 'initial object' is 'terminal object'. Another way of saying this is that a terminal object in a category $\mathcal{A}$ is the same as an initial object in its opposite category $\mathcal{A}^{\mathrm{op}}$.

Invoking the principle of duality can save you work: whenever you have a theorem you can reverse the arrows throughout its statement and proof to obtain a dual theorem. For example, once you've shown that any two initial objects of a category are isomorphic, it follows by duality that any two terminal objects of a category are isomorphic.

Remark 2.1.8 Adjunctions can be composed. Given adjunctions

$$\begin{tikzcd}
\mathcal{A} \arrow[r, shift left=2, "F"] \arrow[r, phantom, "\perp"] & \mathcal{A}' \arrow[l, shift left=2, "G"] \arrow[r, shift left=2, "F'"] \arrow[r, phantom, "\perp"] & \mathcal{A}'' \arrow[l, shift left=2, "G'"]
\end{tikzcd}$$

(where the $\perp$ symbols are rotated versions of $\dashv$), we obtain an adjunction

$$\begin{tikzcd}
\mathcal{A} \arrow[r, shift left=2, "F' \circ F"] \arrow[r, phantom, "\perp"] & \mathcal{A}'' \arrow[l, shift left=2, "G \circ G'"]
\end{tikzcd}$$

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

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

naturally in A and A''.

2.2 Units and counits

In this section and the next we will meet two other ways of saying what an adjunction is. The one in this section is most useful for theoretical purposes; the one in the next is more oriented towards examples.

Before we do anything, we have to make precise the naturality requirement in the definition of adjunction (2.1.1). What it ought to say is that as A and B vary, the isomorphism between $\mathcal{B}(F(A), B)$ and $\mathcal{A}(A, G(B))$ varies in a way that is compatible with all the structure already in place (namely, the composition in the categories $\mathcal{A}$ and $\mathcal{B}$ and the action of the functors F and G). For example, suppose we have maps

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

in $\mathcal{B}$. There are two things we can do: either compose and take the transpose, giving a map $\overline{q \circ g} : A \longrightarrow G(B')$, or take the transpose of g and compose it with $G(q)$, giving a potentially different map $G(q) \circ \overline{g} : A \longrightarrow G(B')$. Naturality will say that these maps are, in fact, the same. Indeed, the naturality requirement is that

$$\overline{(F(A) \xrightarrow{g} B \xrightarrow{q} B')} = (A \xrightarrow{\overline{g}} G(B) \xrightarrow{G(q)} G(B')) \qquad (2{:}2)$$

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

$$\overline{(A' \xrightarrow{p} A \xrightarrow{f} G(B))} = (F(A') \xrightarrow{F(p)} F(A) \xrightarrow{\overline{f}} B) \qquad (2{:}3)$$

for all A, A', B, B', f, p, g, q. It doesn't matter whether we put the big bar over the left or the right of these equations, since bar is self-inverse.

In a later chapter we will see that the naturality requirement can also be cast in terms of natural isomorphism between two functors.

Exercise 2.2.1 Show that the naturality equations (2:2) and (2:3) could equivalently be replaced by the single equation

$$\overline{(A' \xrightarrow{p} A \xrightarrow{f} G(B) \xrightarrow{G(q)} G(B'))} = (F(A') \xrightarrow{F(p)} F(A) \xrightarrow{\overline{f}} B \xrightarrow{q} B').$$

Now take an adjunction
$$\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathcal{B} \arrow[l, shift left, "G"] \end{tikzcd}$$
. For each $A \in \mathcal{A}$ we have a map

$$(A \xrightarrow{\eta_A} G(F(A))) = \overline{(F(A) \xrightarrow{1} F(A))}.$$

Similarly, for each $B \in \mathcal{B}$ we have a map

$$(F(G(B)) \xrightarrow{\varepsilon_B} B) = \overline{(G(B) \xrightarrow{1} G(B))}.$$

These define natural transformations

$$\eta : 1_{\mathcal{A}} \longrightarrow G \circ F, \qquad \varepsilon : F \circ G \longrightarrow 1_{\mathcal{B}},$$

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

Example 2.2.2 Take the usual adjunction
$$\begin{tikzcd} \mathbf{Vect}_k \arrow[r, shift left, "U"] \arrow[r, phantom, "\top"] & \mathbf{Set} \arrow[l, shift left, "F"] \end{tikzcd}$$
. The unit $\eta : 1_{\mathbf{Set}} \longrightarrow U \circ F$ has components

$$\begin{array}{rccc} \eta_S : & S & \longrightarrow & UF(S) = \{\text{formal } k\text{-linear sums } \sum_{s \in S} \lambda_s s\} \\ & s & \longmapsto & s, \end{array}$$

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

$$\varepsilon_V : F(U(V)) \longrightarrow V$$

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

Note that $FU(V)$ is *huge*. 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 $F(U(V))$ is a vector space with one basis element for every element of $\mathbb{R}^2$. So ε_V is a map from this infinite-dimensional space to the 2-dimensional space V.

Lemma 2.2.3 *Given an adjunction $F \dashv G$ with unit η and counit ε, the triangles*

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

commute.

Remark 2.2.4 These are called the **triangle identities**. They are commutative diagrams in the functor categories $[\mathcal{A}, \mathcal{B}]$ and $[\mathcal{B}, \mathcal{A}]$. For an explanation of the notation, see 1.3.14 (and particularly the 'special cases' on p. 16). Another way of expressing them is that

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

commute for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. (We are beginning to omit unnecessary brackets. For instance, $FGF(A)$ means $F(G(F(A)))$.)

Proof of Lemma 2.2.3 Let $A \in \mathcal{A}$. Since $\overline{1_{GF(A)}} = \varepsilon_{F(A)}$, equation (2:3) gives

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

But the left-hand side is $\overline{\eta_A} = \overline{\overline{1_{F(A)}}} = 1_{F(A)}$, so the right-hand side is $1_{F(A)}$, as required. The other identity is proved similarly (exercise). □

The unit and counit are important because they determine the whole adjunction. This might seem surprising, because they appear to know only what the transposes of identities are. Nevertheless, it's true:

Proposition 2.2.5 *Take categories and functors* $\mathcal{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{B}$. *Then there is a one-to-one correspondence between*

a. adjunctions $F \dashv G$, *and*

b. pairs $(1_{\mathcal{A}} \xrightarrow{\eta} GF,\ FG \xrightarrow{\varepsilon} 1_{\mathcal{B}})$ *of natural transformations satisfying the triangle identities.*

(Recall that by definition, an adjunction $F \dashv G$ is a choice of isomorphism (2:1) for each A and B, that when denoted by a bar satisfies the naturality requirements (2:2) and (2:3).)

Proof We show that given a pair (η, ε) of natural transformations satisfying the triangle identities, there is a unique adjunction $F \dashv G$ with unit η and counit ε.

Uniqueness In any adjunction $F \dashv G$ with unit η and counit ε, we have

$$\overline{(F(A) \xrightarrow{g} B)} = \overline{(F(A) \xrightarrow{1} F(A) \xrightarrow{g} B)} = (A \xrightarrow{\eta_A} GF(A) \xrightarrow{G(g)} G(B))$$

for any map $g : F(A) \longrightarrow B$. Similarly, $\overline{f} = \varepsilon_B \circ F(f)$ for any map $f : A \longrightarrow G(B)$. So η and ε determine the adjunction.

Existence Take natural transformations η and ε as in (b). For each A and B, define functions

$$\mathcal{B}(F(A), B) \rightleftarrows \mathcal{A}(A, G(B)), \tag{2:4}$$

both denoted by a bar, by $\overline{f} = \varepsilon_B \circ F(f)$ and $\overline{g} = G(g) \circ \eta_A$.

I claim that these two functions are mutually inverse. Indeed, given a map $g : F(A) \longrightarrow B$ in $\mathcal{B}$, we have a commutative diagram

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

The composite map from $F(A)$ to B by one route around the outside of the diagram is $\varepsilon_B \circ FG(g) \circ F(\eta_A) = \varepsilon_B \circ F(\overline{g}) = \overline{\overline{g}}$, and by the other is $g \circ 1 = g$, so $\overline{\overline{g}} = g$. Similarly, $\overline{\overline{f}} = f$ for any map $f : A \longrightarrow G(B)$ in $\mathcal{A}$. This proves the claim.

Moreover, it follows from the naturality of η and ε that the naturality equations (2:2) and (2:3) hold. So the correspondence (2:4) defines 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 similarly for the counit. □

Corollary 2.2.6 *Take categories and functors* $\mathcal{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{B}$. *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

$$\forall a \in A, b \in B, \qquad f(a) \leq b \iff a \leq g(b). \tag{2:5}$$

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 nonempty. (Also, 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 1.3.4(c) 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 $c \in C$. In the adjunction above, the unit and counit are the statements that $a \leq gf(a)$ for all a and $fg(b) \leq b$ for all b. The triangle identities say nothing, since they assert the equality of two natural transformations between ordered sets.

In the case of ordered sets, Corollary 2.2.6 says that (2:5) is equivalent to the conditions

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

This can also be shown directly.

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

$$\mathcal{P}(X) \underset{i}{\overset{\mathrm{Cl}}{\rightleftarrows}} \mathcal{C}(X)$$

where i is the inclusion and Cl is closure. Indeed, 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 $B \subseteq X$, or equivalently that $A \subseteq \mathrm{Cl}(A)$ for all $A \subseteq X$ and $\mathrm{Cl}(B) \subseteq B$ for all closed $B \subseteq X$.

Remark 2.2.8 Proposition 2.2.5 says 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 (1.3.8) is not necessarily an adjunction. It *is* true that F is left adjoint to G, but η and ε are not necessarily the unit and the counit (because there's no reason why they should satisfy the triangle identities).

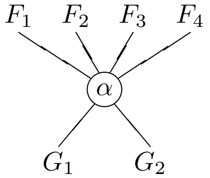
Digression 2.2.9 There is a way of drawing natural transformations that makes the triangle identities intuitive. Suppose, for instance, that we have categories and functors

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

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

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

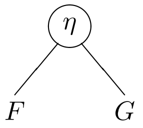
However, we can also draw α as a **string diagram**:



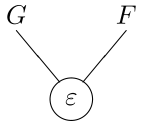
There's clearly nothing special about 4 and 2: we could replace them by any natural numbers m and n. If $m = 0$ then $\mathcal{A} = \mathscr{B}$ and the domain of α is $1_{\mathcal{A}}$. (The composite of no maps should be interpreted as the identity, just as the sum of no numbers is interpreted as zero.) In that case, the disk labelled α will have no strings coming into the top. Similar statements hold for $n = 0$.

Vertical and horizontal composition of natural transformations correspond to putting string diagrams on top of each other and side by side. The identity on a functor F is drawn as a simple string,

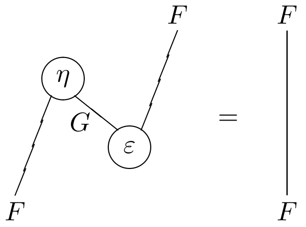
Now let's apply this notation to adjunctions. The unit and counit are drawn as



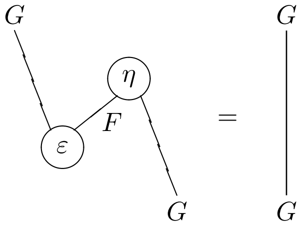
and



The triangle identities now become the topologically plausible



and



—'pull the string straight'.

2.3 Adjunctions via initial objects

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

Consider once more the adjunction

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

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

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

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

$$\begin{array}{rccc} \eta_S : & S & \longrightarrow & UF(S) \\ & s & \longmapsto & s. \end{array}$$

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 $\overline{f} \in \mathbf{Vect}_k(F(S), V)$ such that the diagram
>
> $$\begin{tikzcd}
S \arrow[r, "\eta_S"] \arrow[rd, "f"'] & U(F(S)) \arrow[d, "U(\overline{f})"] \\
& U(V)
\end{tikzcd} \qquad (2{:}6)$$
>
> commutes.

(Compare Example 0.1.1.) The purpose of this section is to show that this statement is equivalent to the statement that F is left adjoint to U with unit η.

To do this, we first need a definition.

Definition 2.3.1 Given categories and functors

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

the **comma category** $(P \Rightarrow Q)$ (often written $(P \downarrow Q)$) is the category with

- objects: triples (A, h, B) with $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $h : P(A) \longrightarrow Q(B)$ in $\mathcal{C}$
- maps $(A, h, B) \longrightarrow (A', h', B')$: pairs $(f : A \longrightarrow A', g : B \longrightarrow B')$ of maps such that

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

  commutes.

Remark 2.3.2 There are canonical functors and a natural transformation as shown:

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

In a sense that we won't go into, $(P \Rightarrow Q)$ is universal with this property.

Examples 2.3.3 a. Let $\mathcal{A}$ be a category and $A \in \mathcal{A}$. Functors $\mathbf{1} \longrightarrow \mathcal{A}$ correspond to objects of $\mathcal{A}$ (Example 2.1.6), and we write $A : \mathbf{1} \longrightarrow \mathcal{A}$ for the functor corresponding to $A \in \mathcal{A}$. Now consider the comma category $(1_{\mathcal{A}} \Rightarrow A)$, as in the diagram

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

An object of $(1_{\mathcal{A}} \Rightarrow A)$ is a pair (X, h) where $X \in \mathcal{A}$ and $h : X \longrightarrow A$ in $\mathcal{A}$. (*A priori* it's a triple (X, h, B) where B is an object of $\mathbf{1}$, but $\mathbf{1}$ only has one object, so we might as well leave the B out.) A map $(X, h) \longrightarrow (X', h')$ is a map $f : X \longrightarrow X'$ in $\mathcal{A}$ making the triangle

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

commute. We write $(1_{\mathcal{A}} \Rightarrow A)$ as $\mathcal{A}/A$, the **slice category** of $\mathcal{A}$ over A. Dually (turn all the arrows around), there is a **coslice category** $A/\mathcal{A} = (A \Rightarrow 1_{\mathcal{A}})$.

b. Take a functor $G : \mathcal{B} \longrightarrow \mathcal{A}$ and an object $A \in \mathcal{A}$, and consider the comma category $(A \Rightarrow G)$, as in the diagram

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

Its objects are pairs $(B \in \mathcal{B}, f : A \longrightarrow G(B))$. A map $(B, f) \longrightarrow (B', f')$ is a map $q : B \longrightarrow B'$ in $\mathcal{B}$ 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 the last diagram resembles (2:6). We're going to use comma categories $(A \Rightarrow G)$ to capture the kind of universal property discussed in the vector space example.

If we're being casual, we say that $f : A \longrightarrow G(B)$ is an object of $(A \Rightarrow G)$, when what we should really say is that the pair $(B, f : A \longrightarrow G(B))$ is an object of $(A \Rightarrow G)$. There's potential for confusion here, since there may be different objects B, B' of $\mathcal{B}$ with $G(B) = G(B')$. Nevertheless, we will often use this casual convention when it seems clear which 'B' is meant.

Now let's see what relevance this has to adjunctions.

Lemma 2.3.4 *Take an adjunction* $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathcal{B} \arrow[l, shift left, "G"] \end{tikzcd}$ *and an object* $A \in \mathcal{A}$*. Then the unit map* $\eta_A : A \longrightarrow GF(A)$ *is an initial object of* $(A \Rightarrow G)$.

Recall from 2.1.6 that an object I of a category is initial if for every object X, there is exactly one map $I \longrightarrow X$. Also recall that initial objects are unique up to isomorphism (if one exists at all).

Proof Let $(B, f : A \longrightarrow G(B))$ 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 $(F(A), \eta_A) \longrightarrow (B, f)$ in $(A \Rightarrow G)$ is a map $q : F(A) \longrightarrow B$ in $\mathcal{B}$ such that

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[rd, "f"'] & GF(A) \arrow[d, "G(q)"] \\ & G(B) \end{tikzcd} \tag{2:7}$$

commutes. But in the 'uniqueness' part of the proof of Proposition 2.2.5, we saw that $G(q) \circ \eta_A = \overline{q}$. So (2:7) commutes if and only if $f = \overline{q}$, if and only if $q = \overline{f}$. So $\overline{f}$ is the unique map $(F(A), \eta_A) \longrightarrow (B, f)$ in $(A \Rightarrow G)$. □

We now meet our third and final formulation of adjointness:

Proposition 2.3.5 *Take categories and functors* $\mathcal{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{B}$. *Then there is a one-to-one correspondence between*

a. adjunctions $F \dashv G$, *and*

b. natural transformations $\eta : 1 \longrightarrow GF$ *such that* $\eta_A : A \longrightarrow GF(A)$ *is initial in* $(A \Rightarrow G)$ *for every* $A \in \mathcal{A}$.

Proof We have just seen how every adjunction $F \dashv G$ gives rise to a natural transformation η as in (b), its unit. To prove the theorem, we have to show that every η as in (b) is the unit of exactly one adjunction $F \dashv G$. But by Proposition 2.2.5, an adjunction $F \dashv G$ amounts to a pair (η, ε) of natural transformations satisfying the triangle identities; so an equivalent task is to show that for every η as in (b), there exists a unique natural transformation $\varepsilon : FG \longrightarrow 1_{\mathcal{B}}$ making the triangle identities hold.

So, take a natural transformation η as in (b).

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

$$\begin{tikzcd} G(B) \arrow[r, "\eta_{G(B)}"] \arrow[rd, "1_{G(B)}"'] & G(FG(B)) \arrow[d, "G(\varepsilon_B)"] \\ & G(B) \end{tikzcd} \tag{2:8}$$

commutes. This says that ε_B is a map

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

in $(G(B) \Rightarrow G)$. The same is true for 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 $B \in \mathcal{B}$, define $\varepsilon_B : FG(B) \longrightarrow B$ to be the unique map $(FG(B), \eta_{G(B)}) \longrightarrow (B, 1_{G(B)})$ in $(G \Rightarrow B)$. (So by definition of ε_B, triangle (2:8) commutes.) We show that $(\varepsilon_B)_{B\in\mathcal{B}}$ is a natural transformation $FG \longrightarrow 1$ such that η and ε satisfy the triangle identities.

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

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

So $q \circ \varepsilon_B$ and $\varepsilon_{B'} \circ FG(q)$ are both maps $\eta_{G(B)} \longrightarrow G(q)$ 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 (2:8) holds. The other says that for $A \in \mathcal{A}$,

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

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

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

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

The Proposition proves the claim made in the introduction to this section: that having an an adjunction $F \dashv G$ amounts to having a natural transformation $\eta : 1 \longrightarrow GF$ with the universal property described there.

Later we will meet the Adjoint Functor Theorems, which provide conditions under which a functor is guaranteed to have a left adjoint. The following corollary is the starting point for the proofs of the Adjoint Functor Theorems.

Corollary 2.3.6 *Let $G : \mathcal{B} \longrightarrow \mathcal{A}$ be a functor. Then G has a left adjoint if and only if for each $A \in \mathcal{A}$, the category $(A \Rightarrow G)$ has an initial object.*

Proof Lemma 2.3.4 proves 'only if'. To prove 'if', choose for each $A \in \mathcal{A}$ an initial object $(F(A), \eta_A : A \longrightarrow GF(A))$ of $(A \Rightarrow G)$. For each map $f : A \longrightarrow A'$ in $\mathcal{A}$, let $F(f) : F(A) \longrightarrow F(A')$ 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 $\eta_A \longrightarrow \eta_{A'} \circ f$ in $(A \Rightarrow G)$). It is easily checked that F is a functor $\mathcal{A} \longrightarrow \mathcal{B}$, and η is then clearly a natural transformation $1 \longrightarrow GF$. So $F \dashv G$ by Proposition 2.3.5. $\square$

Chapter $2\frac{1}{2}$

Interlude on sets

> *Recommended further reading for this chapter: Part I of Lawvere and Schanuel,* Conceptual Mathematics. *See also Lawvere and Rosebrugh,* Sets for Mathematics*: not in the library, but I have a copy you can borrow.*
>
> *Not recommended: almost every other book on set theory.*

The notions of set and function are ubiquitous in mathematics. You might have the impression that they are most strongly connected to the pure end of the subject, but this is an illusion: think of probability density functions in statistics, data sets in computer science, planetary motion in astronomy, or flow in fluid dynamics.

Category theory is often used to shed light on common constructions and patterns in mathematics. If we hope to do this kind of thing, we're first going to have to get straight these basic notions of set and function. That's the aim of the first section of this short chapter. Most of it will be revision, but some parts will probably be new.

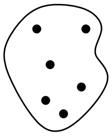
The definition of category mentions a *class* of objects and *classes* of maps. We will see in the second section that some classes of things are 'too big to be sets', which leads to a distinction between 'small' and 'large' classes. This distinction will be needed later in the course.

The final section takes a historical look at set theory. It also explains why the approach to sets taken here is better than the approach you'll find in most books. None of this is necessary for the course, and it hardly needs saying that it's non-examinable, but it may provide useful perspective.

$2\frac{1}{2}$.1 What can you do with sets?

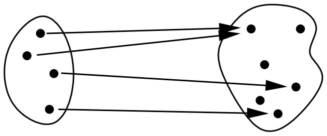
We don't have a definition of 'set' or of 'function'. Nevertheless, guided by our intuition, we'll list some properties that we expect the world of sets and functions to have. For instance, we'll describe some of the sets that we think ought to exist and some ways of building new sets from old. We'll discuss the logical status of all this at the end of the section.

Intuitively, a set is a bag of points:

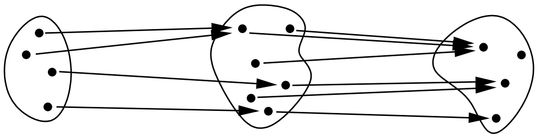


These points, or elements, aren't related to one another in any way. They're not in any order, they don't come with any algebraic structure (e.g. there's no specified way of multiplying elements together), and there's no notion of what it means for one point to be close to another. In particular examples, you might have some extra structure in mind: for instance, you know that the set $\mathbb{Z}$ of integers can be equipped with a ring structure and an order. But when you view it as a *set*, you're ignoring all that structure—regarding it as no more than a bunch of featureless points.

Intuitively, a function $A \longrightarrow B$ is an assignment of a point in bag B to each point in bag A:



We can do one function after another:



and the resulting composition of functions is associative: $h \circ (g \circ f) = (h \circ g) \circ f$. There is also an identity function on every set. Hence:

> *Sets and functions form a category, denoted* **Set***.*

This doesn't pin things down much: there are many categories, mostly quite unlike the category of sets. So let's list some of the special features of sets.

The empty set There is a set with no elements, $\emptyset$.

Suppose that someone hands you a pair of sets, A and B, and tells you to specify a function from A to B. Then your task is to specify for each element of A an element of B. The larger A is, the longer the task; the smaller A is, the shorter the task. In particular, if A is empty then the task takes no time at all: there's nothing to do. So there is a function from $\emptyset$ to B specified by doing nothing. On the other hand, there aren't two different ways to do nothing, so there's only one function from $\emptyset$ to B. Hence:

$\emptyset$ *is an initial object of* **Set**.

Perhaps you're not convinced by that argument. If not, consider the following. Suppose we have a set A with subsets A_1 and A_2 such that $A_1 \cup A_2 = A$ and $A_1 \cap A_2 = \emptyset$. Then a function from A to B amounts to a function from A_1 to B together with a function from A_2 to B. So if all the sets are finite, we should have the rule

$$\begin{aligned}(\text{number of functions from } A \text{ to } B) &= (\text{number of functions from } A_1 \text{ to } B) \\ &\quad \times(\text{number of functions from } A_2 \text{ to } B).\end{aligned}$$

In particular, we could take $A_1 = A$ and $A_2 = \emptyset$. This would force the number of functions from $\emptyset$ to B to be 1. So if we want this rule to hold (and surely we do!), we'd better say that there's exactly one function from $\emptyset$ to B.

As for functions *into* $\emptyset$: there's exactly one function $\emptyset \longrightarrow \emptyset$ (the identity), but if A is not empty then there are no functions $A \longrightarrow \emptyset$ (because there's nowhere for elements of A to go).

The one-point set There is a set with exactly one element. I'll call this set 1.

For any set A, there is exactly one function from A to 1, since every element of A must be mapped to the unique element of 1. That is:

1 *is a terminal object of* **Set**.

A function *from* 1 *to* a set B is just a choice of an element of B; in short,

The functions $1 \longrightarrow B$ *are the elements of* B.

So the concept of 'element' can be derived from the basic concepts of set and function.

Products Any two sets A and B have a product, $A \times B$. Its elements are ordered pairs (a, b) with $a \in A$ and $b \in B$. You're used to ordered pairs from coordinate geometry. All that matters about them is the following property: for $a, a' \in A$ and $b, b' \in B$,

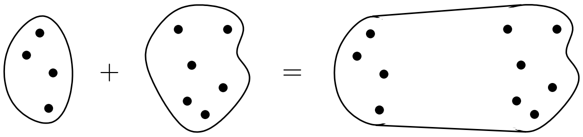
$$(a, b) = (a', b') \iff a = a' \text{ and } b = b'.$$

In fact, take any set I and family $(A_i)_{i \in I}$ of sets. Then this family has a product $\prod_{i \in I} A_i$, whose elements are families $(a_i)_{i \in I}$ with $a_i \in A_i$ for all i.

An important special case is when all the A_i's are equal. Let's write A for their common value. Then we have a set $\prod_{i \in I} A$ whose elements are the families $(a_i)_{i \in I}$ with $a_i \in A$ for all i. In other words, $\prod_{i \in I} A$ is the set of functions from I to A. We usually write $\prod_{i \in I} A$ as A^I.

Sums Any two sets A and B have a **sum** $A+B$.

This construction may be less familiar. Thinking of sets as bags of points, the sum of two sets is what you get when you put all the points into one big bag:



If A and B are finite sets with m and n elements respectively, then $A+B$ always has $m+n$ elements.

There are inclusion functions

$$\begin{tikzcd} A \arrow[r, "i"] & A+B & B \arrow[l, "j"'] \end{tikzcd}$$

such that the union of the images of i and j is all of $A+B$ and the intersection of the images is empty.

Sum is sometimes called **disjoint union** and written $A \amalg B$. Don't confuse sum with (ordinary) union $\cup$. For a start, you can take the sum of *any* two sets A and B, whereas $A \cup B$ only makes sense when A and B come as subsets of some larger set. Even then, $A+B$ and $A \cup B$ are often different. For example, take the subsets $A = \{1,2,3\}$ and $B = \{3,4\}$ of $\mathbb{N}$. Then $A \cup B = \{1,2,3,4\}$, but $A+B$ is a set with $3+2=5$ elements.

More generally, any family $(A_i)_{i \in I}$ of sets has a sum $\sum_{i \in I} A_i$.

Digression on arithmetic We are using notation reminiscent of arithmetic: $A \times B$, A^B and $A+B$. There is good reason for this: if A is a finite set with m elements and B a finite set with n elements then $A \times B$ has $m \times n$ elements, and similarly for A^B and $A+B$. This also fits with our notation 1 for a one-element set, and the alternative notation 0 for the empty set $\emptyset$. All the usual laws of arithmetic have their counterparts for sets:

$$\begin{aligned} A \times (B+C) &\cong (A \times B) + (A \times C), \\ A^{B+C} &\cong A^B \times A^C, \end{aligned}$$

and so on, where $\cong$ is isomorphism in the category of sets. These are true for *any* sets. The usual laws of arithmetic can be deduced by restricting to *finite* sets and counting the number of elements on each side.

The two-point set Let 2 be the set $1+1$ (a set with two elements!). For reasons that will emerge, I'll write the elements of 2 as `true` and `false`.

Given a subset S of A, we obtain a function $\chi_S : A \longrightarrow 2$, where for $a \in A$,

$$\chi_S(a) = \begin{cases} \texttt{true} & \text{if } a \in S \\ \texttt{false} & \text{if } a \notin S. \end{cases}$$

Conversely, given a function $f : A \longrightarrow 2$, we obtain a subset

$$f^{-1}\{\texttt{true}\} = \{a \in A \mid f(a) = \texttt{true}\}$$

of A. These two processes are mutually inverse. Hence:

> *Subsets of A correspond one-to-one with functions $A \longrightarrow 2$.*

We already know that the functions from A to 2 form a set, 2^A. When we are thinking of 2^A as being the set of all subsets of A, we call it the **powerset** of A and write it $\mathcal{P}(A)$.

Equalizers It would be nice to say that, given a set A, you can define a subset S of A by specifying a property that the elements of S are to satisfy:

$$S = \{a \in A \mid \text{some property of } a \text{ holds}\}.$$

It's hard to give a general definition of 'property'. There is, however, a special type of property that's easy to handle: equality of two functions. Precisely, given sets and functions $A \underset{g}{\overset{f}{\rightrightarrows}} B$, there is a set

$$\{a \in A \mid f(a) = g(a)\}.$$

This set is called the **equalizer** of f and g, as it's the part of A on which the two functions are equal.

Quotients by equivalence relations You're familiar with quotient groups and quotient rings (also known as factor groups and factor rings) in algebra. Quotients also come up everywhere in topology, as for instance when you glue two opposite sides of a square together to make a cylinder. But the most basic context for quotients is that of sets.

Let A be a set and $\sim$ an equivalence relation on A. Then there is a set $A/\sim$, the **quotient of A by $\sim$**, whose elements are the equivalence classes. There is also a map

$$p : A \longrightarrow A/\sim$$

sending an element of A to its equivalence class; it is surjective and has the property that $p(a) = p(a') \iff a \sim a'$.

We have now listed the most important properties of sets and functions. Here are two further properties.

Infinity As far as the properties listed so far are concerned, it could be the case that all sets are finite. The following property guarantees the existence of an infinite set, $\mathbb{N}$.

There is a function $s : \mathbb{N} \longrightarrow \mathbb{N}$ defined by $s(n) = n + 1$. There is also an element 0 of $\mathbb{N}$. It turns out that the crucial property of $\mathbb{N}$ is this: for every set X, function $t : X \longrightarrow X$, and element x of X, there is a unique function $f : \mathbb{N} \longrightarrow X$ such that $f(0) = x$ and $f \circ s = t \circ f$.

A function with domain $\mathbb{N}$ is usually called a **sequence**. The property says that for every set X, function $t : X \longrightarrow X$ and element x of X, there is a unique sequence (x_n) such that $x_0 = x$ and $x_{n+1} = t(x_n)$ for all n.

I only mention this property for the sake of completeness; we won't really use it.

Choice Let $f : A \longrightarrow B$ be a map in a category $\mathcal{A}$. A **section** of f is a map $i : B \longrightarrow A$ in $\mathcal{A}$ such that $f \circ i = 1_B$.

In the category of sets, any map with a section is surjective (exercise). The converse statement is called the Axiom of Choice:

> *Every surjective function has a section.*

It's called 'choice' because specifying a section of $f : A \longrightarrow B$ amounts to choosing, for each $b \in B$, an element of the nonempty set $\{a \in A \mid f(a) = b\}$.

The properties that we've set out aren't theorems, since we don't have rigorous definitions of set and function. They read more like a wishlist. Let's try to understand what's going on.

Definitions in mathematics usually depend on previous definitions. A vector space is defined as an abelian group with a scalar multiplication. An abelian group is defined as a group with a certain property. A group is defined as a set with certain extra structure. A set is defined as... well, what?

We can't keep going back indefinitely, otherwise we quite literally wouldn't know what we were talking about. We have to start somewhere. In other words, there have to be some 'basic' concepts not defined in terms of anything else. The concept of 'set' is usually taken to be one of the basic ones, which is why you've probably never read a sentence beginning 'Definition: a **set** is...'. We'll also take 'function' as a basic concept.

But now there seems to be a problem. If these basic concepts aren't defined in terms of anything else, how are we to know what they really are? How are we going to reason in the watertight, logical way that mathematics depends on? It's no use trusting our intuitions: your intuitive idea of what a set is might be slightly different from mine, and if it came to a dispute we'd have no way of deciding who was right.

The problem is solved as follows. Instead of *defining* a set to be a such-and-such and a function to be a such-and-such else, we list some *properties* that we declare sets and functions to have. In other words, we never attempt to say what sets and functions *are*; we just say what you can *do* with them.

What we are doing is often referred to as 'foundations'. In this metaphor, the foundation consists of the basic concepts (set and function), which are not built on anything else. Instead, one states a list of properties that they are declared to satisfy. On top of the foundations are built various definitions; on top of those, further definitions; and so on, towering upwards.

The properties above are stated informally, but they can be formalized using some categorical language. (See *Sets for Mathematics.*) In the formal version, we'd begin by saying that sets and functions form a category, **Set**. We'd then list some properties of this category, such as having an initial and a terminal object. The properties described informally under the headings Products and Equalizers are made formal by the statement that **Set** 'has limits'. Later we'll learn what 'has limits' means; for now, all that matters is that it's a property that an arbitrary category may or may not have, and we're declaring that **Set** does have it.

While we were making the list, we were guided by our intuition about sets. But now that it's made, our intuition plays no further official role: any disputes about the nature of sets are settled by consulting the list of properties.

With this in mind, let's look again at the section on the empty set. You might have felt that I was on shaky ground when trying to convince you that $\emptyset$ is initial. But the point is that I don't need to convince you that this is a *true statement*; I only need to convince you that it's a *convenient assumption.* Indeed, it doesn't even make sense to ask whether it's 'true', since we don't have definitions of set and function, and at that point we had made no assumptions about how sets and functions behave.

We can make whatever assumptions we like, but some lead to more interesting mathematics than others. If you want to assume that there are *no* functions from $\emptyset$ to any other set, you can: no problem. It's just that the tower of mathematics built on that foundation will look different from what you're used to—for instance, the 'number of functions' rule fails.

$2\frac{1}{2}$.2 Small vs. large

We have now made some assumptions about how sets behave. One of the consequences of these assumptions is this: you can't collect together any old bunch of things and call it a set.

To show this, we need a lemma.

Lemma $2\frac{1}{2}$.2.1 *There is no injection $\mathcal{P}(A) \longrightarrow A$ for any set A.*

Recall that $\mathcal{P}(A)$ is the powerset of A. The lemma is easy for finite sets, since if A has n elements then $\mathcal{P}(A)$ has 2^n elements, and $2^n > n$. The proof is often called **Cantor's diagonal argument**.

Proof (not for exam) Suppose there is an injection $i : \mathcal{P}(A) \longrightarrow A$. Call a subset S of A **good** if $i(S) \in S$, and **bad** otherwise. Let

$$R = \{a \in A \mid a = i(S) \text{ for some bad } S \in \mathcal{P}(A)\}.$$

I claim that for all $S \in \mathcal{P}(A)$,

$$S \text{ is bad} \iff i(S) \in R.$$

'$\Rightarrow$' is immediate. For '$\Leftarrow$', if $i(S) \in R$ then $i(S) = i(S')$ for some bad S'; but i is injective, so $S = S'$, so S is bad. This proves the claim. Now taking $S = R$, we have

$$R \text{ is bad} \iff R \text{ is good},$$

a contradiction. □

Theorem $2\frac{1}{2}$.2.2 *There is no set Q such that $A \in Q$ for all sets A.*

In other words, there is no 'set of all sets'.

Proof (not for exam) Suppose there is such a set Q. There is an injection $i : \mathcal{P}(Q) \longrightarrow Q$, defined by $i(S) = S$ for subsets S of Q. This contradicts the Lemma. □

We use the word **class** to mean any collection of mathematical objects. All sets are classes, but the moral of the Theorem is that some classes are too big to be sets. A class will be called **small** if it is a set, and **large** (or a **proper class**) otherwise.

A category $\mathcal{A}$ is **small** if the class of all maps in $\mathcal{A}$ is small, and **large** otherwise. If $\mathcal{A}$ is small then the class of objects of $\mathcal{A}$ is small too, since objects correspond one-to-one with identity maps. I will follow the custom of using one typeface ($\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, ...) for small categories and another ($\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, ...) for arbitrary categories.

A category $\mathcal{A}$ is **locally small** if for each $A, B \in \mathcal{A}$, the class $\mathcal{A}(A, B)$ is small. Many authors take local smallness to be part of the definition of category. The class $\mathcal{A}(A, B)$ is often called the **hom-set** from A to B; strictly speaking, we should only say this when $\mathcal{A}$ is locally small.

A category is small if and only if it is locally small and the class of objects is small. If this seems mysterious, try to prove the following similar fact: a category $\mathcal{A}$ is finite (i.e. the class of all maps in $\mathcal{A}$ is finite) if and only if it is locally finite (i.e. each class $\mathcal{A}(A, B)$ is finite) and the class of objects is finite.

Recall that the category of *all* categories and functors is written **CAT**. The category of *small* categories and functors is written **Cat**.

Examples $2\frac{1}{2}$.2.3

a. Monoids (e.g. groups) are by definition *sets* equipped with certain structure, so the one-object categories that they correspond to are small. The categories corresponding to ordered sets are also small.

b. The category **Set** is large, since the Theorem tells us that its class of objects is not a set. The point to get clear is that

   While any *individual* set is small, the class of *all* sets is large.

   But **Set** is *locally* small, since if $A, B \in \mathbf{Set}$ then $\mathbf{Set}(A, B)$ is the set B^A. For similar reasons, other familiar categories of mathematical structures such as **Top**, **Gp**, **Ring** and $\mathbf{Vect}_k$ are large but locally small.

This may all seem rather murky. If so, don't worry: we won't pay much attention to these delicate issues of 'size'. Occasionally we'll need to make the distinction between small and large collections, but that's all.

$2\frac{1}{2}$.3 Historical remarks

The set theory that we began to develop in the first section is rather different from what most mathematicians think of as set theory. Here I'll explain what this dominant version of set theory is, why it is the object of widespread scorn, and why our kind of set theory is better.

Cantor's set theory The first person to think systematically about sets was the German–Russian mathematician Georg Cantor, in the late 19th century. Previously, sets hadn't been regarded as entities worthy of study in their own right; but Cantor, originally motivated by a problem in Fourier analysis, developed an extensive theory of sets. Among many other things, he showed that there are different 'sizes' of infinity, proving, for instance, that there is no bijection between $\mathbb{N}$ and $\mathbb{R}$.

Cantor was met with all the resistance that typically accompanies a really new idea. His work was criticized as nonsensical, as meaningless, as far too abstract; later, as all very well but of no use to the 'mainstream' of mathematics; Kronecker, an important mathematician of the day, called him a charlatan and a 'corrupter of youth'. Nowadays, the basics of his work are on almost every undergraduate mathematics syllabus. Times change: in the modern style of mathematics, almost every definition, when unravelled sufficiently, depends on the notion of set. But pre-Cantor this wasn't so. It is interesting to try to understand the mindset of mathematicians of the time, who had successfully developed sophisticated subjects such as complex analysis and Galois theory without depending on this notion that we now regard as so fundamental.

Before we continue with the history, we need to discuss another fundamental concept.

Types Suppose someone asks you 'is $\sqrt{2} = \pi$?' Your answer, of course, is 'no'. Now suppose someone asks you 'is $\sqrt{2} = \log$?' You might frown and wonder if you'd heard right, and perhaps your answer would again be 'no'—but it would be a different kind of 'no'. After all, $\sqrt{2}$ is a number, whereas log is a function, so it's inconceivable that they might be equal. A better answer might be 'your question makes no sense'.

This illustrates the idea of **types**. $\sqrt{2}$ is a real number, $\mathbb{Q}$ is a field, S_3 is a group, log is a function from $(0, \infty)$ to $\mathbb{R}$, and $\frac{d}{dx}$ is an operation that turns one function into another. Formally, we say that the type of $\sqrt{2}$ is 'real number', and so on. We have an inbuilt sense of type, and it wouldn't usually occur to us to ask whether two things of different type were equal.

You may have met this idea before if you have programmed computers. Many programming languagues require you to declare the type of a variable before you first use it. For example, you might declare that x is to be a variable of type 'real number', n a variable of type 'integer', M a variable of type '3×3 array', and so on.

The distinction between objects of different types has always been instinctively understood. However, things took a strange turn at the beginning of the 20th century.

Membership-based set theory Those who came after Cantor sought to compile a definitive list of assumptions to be made about sets (an **axiomatization** of set theory). The list they arrived at, in the early years of the 20th century, is known as ZFC (Zermelo–Fraenkel with Choice). It soon became the standard. If you choose a random mathematician and say the words 'set theory' to them, they will almost certainly think of ZFC.

From what I've just said, you might imagine that the axiomatization of Zermelo *et al.* was similar to the one that we were working towards in the first section of this chapter. In spirit it was, but there was a crucial difference: whereas we used sets and *functions* as our basic concepts, they took sets and *membership* ($\in$).

At first sight this difference may look innocuous. But when the membership-based approach is used as a foundation on which to build the rest of mathematics, several bizarre features become apparent:

- In this approach, *everything* is a set. For instance, a function is defined as a set with certain properties. Many other things that you wouldn't think of as being sets are, nevertheless, regarded as sets: the number $\sqrt{2}$ is a set, the field $\mathbb{Q}$ (*including* its field structure) is a set, the group S_3 (*including* its group structure) is a set, the function log is a set, the operator d/dx is a set, and so on.

  You might wonder how on earth this is possible. Perhaps it's useful to compare data storage in a computer, where files of all different types—sound files, text files, image files—are ultimately encoded as sequences of 0s and 1s. To give an example, in the membership-based set theory that you'll find in most books, the number 4 is encoded as the set

  $$\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}.$$

- The virtue of this approach is its simplicity: *everything* is a set! But the price to be paid is very high: we lose the fundamental notion of type. This is because everything is regarded as being of type 'set'. Returning to the example of the Types section, if we use membership-based foundations then the two questions 'is $\sqrt{2} = \pi$?' and 'is $\sqrt{2} = \log$?' make equal sense.

- This approach begins with the declaration that there are some things (called sets) and there is a binary relation on those things (called membership, $\in$). This is a 'global' relation, that is, for *any* two sets A and B it makes sense to ask whether $A \in B$. (Compare the approach we took to elements in the first section: there, an element of a set B is a *function* from 1 to B, so is not itself a set.)

> Since *everything* is being viewed as a set, it makes sense to ask, for *any* two things A and B, whether $A \in B$. For instance, 'is $\mathbb{Q} \in \sqrt{2}$?' makes just as much sense as 'is $\sqrt{2} \in \mathbb{Q}$?'.
>
> Further still, the axioms of ZFC imply that we can form the intersection $A \cap B$ of *any* sets A and B. (Its elements are those sets C for which $C \in A$ and $C \in B$.) So it makes sense to ask such apparently nonsensical questions as 'is $S_3 \cap \log = \emptyset$?'

The answers to all these nonsensical questions depend entirely on the fine detail of how mathematical objects (numbers, functions, groups, etc) are encoded as sets. Even devotees of the membership-based approach would agree that this encoding is entirely a matter of convention, just like a word processor's encoding of a document as a string of 0s and 1s. So the answers to these questions are quite meaningless. To borrow the words of Hermann Weyl, membership-based set theory is too full of sand.

Set theory today From the previous section it should be clear why most modern-day mathematicians take a dim view of set theory. However often they are told that it is 'the foundation of mathematics', they feel that much of it is artificial and irrelevant to their concerns.

To a large extent this is justified. But it is also a symptom of the historical dominance of membership-based set theory: most mathematicians don't realize there's any other kind. Taking sets and functions (rather than sets and membership) as the basic concepts leads to a theory containing all of the meaningful set theory of Cantor and others, with none of the meaningless aspects that arise from assuming a global membership relation. In particular, the function-based approach respects the fundamental notion of type.

The function-based approach is, of course, categorical, and its advantages are related to more general points about how the world looks through categorical eyes. The categorical view could be characterized as 'external': it sees a class of objects and the maps that relate them, but the objects aren't assumed to have any kinds of innards or distinguishing features, beyond their place in the category. This applies to the categorical approach to any subject, and in particular set theory. Membership-based set theory is 'internal': it sees a class of sets and the inside of every set (because the membership relation tells you what the elements are).

Actually, a great strength of category theory is that it is both external *and* internal. We get inside an object by probing it with maps to or from other objects: for example, an element of a set A is a map $1 \longrightarrow A$, and a subset of A is a map $A \longrightarrow 2$. This is a major theme of the next chapter.

Chapter 3

Representables

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

Take, for instance, the category of topological spaces, and consider how it looks from the point of view of the one-point space 1. A map from 1 to any space X is the same thing as a point of X; we might say that 1 'sees points'. Similarly, a map from $\mathbb{R}$ to a space X is what could be called a curve in X; in this sense, $\mathbb{R}$ sees curves.

Or take, if you prefer, the category of groups. A map from the infinite cyclic group $\mathbb{Z}$ to any group G amounts to an element of G. (To prove this, think about the image of $1 \in \mathbb{Z}$.) 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.

In the category of fields (and ring homomorphisms), a map $K \longrightarrow L$ is a way of describing L as an extension of K, so what each field K sees is the ways in which other fields extend it. If K and L are fields of different characteristics then there are no maps between K and L; hence the category of fields is the union of disjoint subcategories $\mathbf{Field}_0$, $\mathbf{Field}_2$, $\mathbf{Field}_3$, $\mathbf{Field}_5$, $\ldots$ consisting of the fields of characteristics $0, 2, 3, 5, \ldots$. Each field sees only the fields of the same characteristic.

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

So far we've chosen an object and asked how it *sees* the world. But we can ask the dual question: given an object, how is it *seen by* the world? In other words, what are the maps into it?

For example, let S be the two-point topological space $\{a, b\}$ in which $\{a\}$ is open but $\{b\}$ is not. For any space X, the maps from X to S correspond one-to-one with open subsets of X (exercise; if stuck, compare the observations on the two-point set in section $2\frac{1}{2}$.1). So S is 'seen by open subsets'.

In the category of fields, the field $\mathbb{C}$ of complex numbers is seen by the subfields of $\mathbb{C}$; that is, the maps into $\mathbb{C}$ correspond to the subfields of $\mathbb{C}$. (Actually, this isn't quite true: it's really the *isomorphism classes* of maps into $\mathbb{C}$ that correspond to subfields. This will be explained when we come to do subobjects.)

In the ordered set $(\mathbb{R}, \leq)$, the object 0 is seen by the non-positive numbers.

We spend this chapter exploring this theme: how each object sees and is seen by the category that it lives in. We're naturally led to the notion of representable functor, which (after adjunctions) provides our second way of getting at the idea of universal property.

3.1 Basics

We've seen that for any topological space X, its underlying set $U(X)$ is in one-to-one correspondence with the set $\mathbf{Top}(1, X)$ of continuous maps from the one-point space 1 to X. We also know that U defines a functor $\mathbf{Top} \longrightarrow \mathbf{Set}$: that is, U is defined on maps in **Top**, as well as objects. This suggests that there should be a sensible way of defining $\mathbf{Top}(1, f)$ for any map f in **Top**, so that $\mathbf{Top}(1, -)$ defines a functor $\mathbf{Top} \longrightarrow \mathbf{Set}$.

This is indeed the case, as we're about to see. In fact, we'll see that in this respect, there's nothing special about the category **Top** or the object 1.

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

$$H^A = \mathcal{A}(A, -) : \mathcal{A} \longrightarrow \mathbf{Set}$$

as follows:

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

  $$H^A(g) = \mathcal{A}(A, g) : \mathcal{A}(A, B) \longrightarrow \mathcal{A}(A, B')$$

  by

  $$p \longmapsto g \circ p$$

  for all $p : A \longrightarrow B$.

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

b. $H^A(g)$ is sometimes called g_* or $g \circ -$.

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

Representable functors are sometimes just called 'representables'. Only set-valued functors can be representable!

Examples 3.1.4 a. All of the 'seeing' functors from the introduction to this chapter are representable. For example, the forgetful functor $U : \mathbf{Top} \longrightarrow \mathbf{Set}$ is isomorphic to $H^1 = \mathbf{Top}(1, -)$, and the forgetful functor $\mathbf{Gp} \longrightarrow \mathbf{Set}$ is isomorphic to $\mathbf{Gp}(\mathbb{Z}, -)$. For each prime p, there is a functor $V_p : \mathbf{Gp} \longrightarrow \mathbf{Set}$ defined on objects by

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

and as claimed in the introduction, $V_p \cong \mathbf{Gp}(\mathbb{Z}/p\mathbb{Z}, -)$ (exercise). Hence V_p is representable.

b. There is a functor $\mathrm{ob} : \mathbf{Cat} \longrightarrow \mathbf{Set}$ sending a small category to its set of objects. If $\mathbf{1}$ denotes the terminal category (with one object and only the identity map) then $\mathrm{ob} \cong \mathbf{Cat}(\mathbf{1}, -)$. Hence ob is representable.

c. Let M be a monoid, regarded as a one-object category. Recall from 1.2.3(e) 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.

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

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

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

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

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

whose value $\mathbf{Bilin}(U, V; W)$ at $W \in \mathbf{Vect}_k$ is the set of bilinear maps $U \times V \longrightarrow W$. 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 so-called tensor product $U \otimes V$, which we met in 0.1.3.)

f. Take an adjunction $\begin{tikzcd} \mathcal{A} \arrow[r, bend left, "F"] \arrow[r, phantom, "\perp"] & \mathcal{B} \arrow[l, bend left, "G"] \end{tikzcd}$. For each $A \in \mathcal{A}$, the functor

$$\mathcal{A}(A, G(-)) : \mathcal{B} \longrightarrow \mathbf{Set}$$

is isomorphic to $\mathcal{B}(F(A), -) = H^{F(A)}$, and is therefore representable.

You wouldn't expect a randomly-chosen set-valued functor to be representable; in some sense, rather few functors are. However, forgetful functors tend to be:

Proposition 3.1.5 *Any set-valued functor with a left adjoint is representable.*

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

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

naturally in $A \in \mathcal{A}$; hence $G \cong H^{F(1)}$. □

Examples 3.1.6 a. Several of the examples above (3.1.4) are covered by this result. In 2.1.4 we saw that $U : \mathbf{Top} \longrightarrow \mathbf{Set}$ has a left adjoint D, and $D(1) \cong 1$, so we recover the result that $U \cong H^1$. Similarly, in Sheet 4, q.4, you (hopefully) constructed a left adjoint D to the ob : $\mathbf{Cat} \longrightarrow \mathbf{Set}$, and $D(1) \cong \mathbf{1}$; this proves again that ob $\cong H^{\mathbf{1}}$.

b. The forgetful functor $U : \mathbf{Vect}_k \longrightarrow \mathbf{Set}$ 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.

c. Examples 2.1.3 began with the words 'Any forgetful functor between categories of algebraic structures has a left adjoint'. Take the category **CRing** of commutative rings and the forgetful functor $U : \mathbf{CRing} \longrightarrow \mathbf{Set}$. This general principle tells us that U has a left adjoint, and the Proposition then tells us that U is representable.

Let's see how this works explicitly. Given a set S, let $\mathbb{Z}[S]$ be the ring of polynomials in variables x_s ($s \in S$); for instance, if $S = 1$ then $\mathbb{Z}[S]$ is the ring $\mathbb{Z}[x]$ of polynomials in one variable. Then $S \longmapsto \mathbb{Z}[S]$ defines a functor $\mathbf{Set} \longrightarrow \mathbf{CRing}$, and this is the left adjoint to U. Hence $U \cong H^{\mathbb{Z}[x]}$. Again, this can be verified directly: for any ring R, the maps $\mathbb{Z}[x] \longrightarrow R$ correspond one-to-one with the elements of R (exercise).

We have defined, for each object A of our category $\mathcal{A}$, a functor $H^A \in [\mathcal{A}, \mathbf{Set}]$. This describes how A sees the world. As A varies, the view varies. On the other hand, it's always the same world being seen, so the different views obtained from different objects are somehow related. (It's a bit like 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 $(H^A)_{A \in \mathcal{A}}$ of 'views' has some consistency to it. What this means is that whenever there's a map between objects A and A', there's also a map between H^A and $H^{A'}$.

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

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

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

$$\begin{array}{ccc} H^A(B) = \mathcal{A}(A, B) & \longrightarrow & H^{A'}(B) = \mathcal{A}(A', B) \\ p & \longmapsto & p \circ f. \end{array}$$

Construction 3.1.7 Let $\mathcal{A}$ be a locally small category. We define a functor

$$H^\bullet : \mathcal{A}^{\mathrm{op}} \longrightarrow [\mathcal{A}, \mathbf{Set}]$$

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

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

Everything we've done so far can be dualized. At the formal level this is trivial: reverse all the arrows, so that every $\mathcal{A}$ becomes an $\mathcal{A}^{\mathrm{op}}$ and *vice versa*. But in our usual examples it will feel different: we're no longer asking how objects *see*, but how they are *seen*.

First we dualize Construction 3.1.1.

Construction 3.1.8 Let $\mathcal{A}$ be a locally small category and $A \in \mathcal{A}$. We define a functor

$$H_A = \mathcal{A}(-, A) : \mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$$

as follows:

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

$$H_A(g) = \mathcal{A}(g, A) = g^* = - \circ g : \mathcal{A}(B, A) \longrightarrow \mathcal{A}(B', A)$$

  by

$$p \longmapsto p \circ g$$

  for all $p : B \longrightarrow A$.

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

We now define representability for *contravariant* set-valued functors. Strictly speaking this is unnecessary, as a contravariant functor on $\mathcal{A}$ is a covariant functor on $\mathcal{A}^{\mathrm{op}}$, 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 3.1.9 Let $\mathcal{A}$ be a locally small category. A functor $X : \mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$ is **representable** if $X \cong H_A$ for some $A \in \mathcal{A}$. A **representation** of X is a choice of an $A \in \mathcal{A}$ and an isomorphism between H_A and X.

Examples 3.1.10 a. The functors in the latter part of the introduction to this chapter, describing how a particular object is 'seen by' the rest of the category, are representable. For example, there is a functor

$$\mathcal{O} : \mathbf{Top}^{\mathrm{op}} \longrightarrow \mathbf{Set}$$

defined on objects B by taking $\mathcal{O}(B)$ to be the set of open subsets of B, and on maps $g : B' \longrightarrow B$ by $(\mathcal{O}(g))(U) = g^{-1}(U)$ for all $U \in \mathcal{O}(B)$. (Of course, we're not assuming that g is invertible: $g^{-1}(U)$ denotes the inverse image or preimage of U under g, that is, $\{x' \in B' \mid g(x') \in U\}$.)

We described a certain two-point space S with the property that maps from any space B into S correspond naturally to open subsets of B. Made precise, this says that $\mathcal{O} \cong H_S$; hence $\mathcal{O}$ is representable.

b. This is a simpler version of the previous example. There is a functor

$$\mathcal{P} : \mathbf{Set}^{\mathrm{op}} \longrightarrow \mathbf{Set}$$

sending each set B to its powerset $\mathcal{P}(B)$, and defined on maps $g : B' \longrightarrow B$ by $(\mathcal{P}(g))(U) = g^{-1}(U)$ for all $U \in \mathcal{P}(B)$. As we saw in the last chapter, a subset amounts to a map into the two-point set 2, so $\mathcal{P} \cong H_2$.

c. In 1.2.5(a) we defined a functor $C : \mathbf{Top}^{\mathrm{op}} \longrightarrow \mathbf{Ring}$, 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 $U(C(B)) = \mathbf{Top}(B, \mathbb{R})$ for $B \in \mathbf{Top}$.

Previously we assembled the covariant representables $(H^A)_{A \in \mathcal{A}}$ into one big functor $H^{\bullet}$; now we do the same thing for the contravariant representables $(H_A)_{A \in \mathcal{A}}$. Any map $A \xrightarrow{f} A'$ in $\mathcal{A}$ induces a natural transformation

$$\begin{tikzcd}
\mathcal{A} \arrow[r, bend left=50, "H_A"{name=U}] \arrow[r, bend right=50, "H_{A'}"'{name=D}] & \mathbf{Set} \arrow[from=U, to=D, Rightarrow, "H_f"]
\end{tikzcd}$$

(also called $\mathcal{A}(-, f)$, f_* and $f \circ -$), whose B-component (for $B \in \mathcal{A}$) is

$$\begin{array}{ccc}
H_A(B) = \mathcal{A}(B, A) & \longrightarrow & H_{A'}(B) = \mathcal{A}(B, A') \\
p & \longmapsto & f \circ p.
\end{array}$$

Construction 3.1.11 Let $\mathcal{A}$ be a locally small category. We define a functor

$$H_{\bullet} : \mathcal{A} \longrightarrow [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$$

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. As usual, $\mathcal{A}$ is a locally small category.

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

The part below the line is the dual of the part above the line. Whichever you choose, you can't avoid contravariance! For example, you might have started out by preferring the H^A's because they're covariant: but when you put them together, you get a contravariant functor $H^\bullet$.

The rest of the chapter is about the theory of representable functors. It doesn't make much difference whether we work with the H^A's and $H^\bullet$ or with the H_A's and $H_\bullet$: any theorem that we prove about one dualizes to give a theorem about the other. We choose to work with the H_A's and $H_\bullet$.

In a sense to be explained, $H_\bullet$ 'embeds' $\mathcal{A}$ into $[\mathcal{A}^{\text{op}}, \mathbf{Set}]$. This can be useful, because the category $[\mathcal{A}^{\text{op}}, \mathbf{Set}]$ has all sorts of nice properties that $\mathcal{A}$ might not have. It's also one reason why we choose to work with the functors below the line, since we want to have an embedding of $\mathcal{A}$, not $\mathcal{A}^{\text{op}}$, into a nice category.

Exercise 3.1.12 (Highly recommended) Given $A, A' \in \mathcal{A}$ with $H_A \cong H_{A'}$, prove that $A \cong A'$. (In other words, $H_\bullet$ is 'injective on isomorphism classes of objects'.)

There is one more functor to define. It unifies the 'above the line' and 'below the line' functors.

Construction 3.1.13 Let $\mathcal{A}$ be a locally small category. The functor

$$\text{Hom}_{\mathcal{A}} : \mathcal{A}^{\text{op}} \times \mathcal{A} \longrightarrow \mathbf{Set}$$

is defined by

$$\begin{tikzcd}
(A,B) \arrow[d, shift left, "g"] \arrow[r, mapsto] \arrow[dr, phantom, "\longmapsto"] & \mathcal{A}(A,B) \arrow[d, "g \circ - \circ f"] \\
(A',B') \arrow[u, shift left, "f"] \arrow[r, mapsto] & \mathcal{A}(A',B').
\end{tikzcd}$$

In other words, $\text{Hom}_{\mathcal{A}}(A,B) = \mathcal{A}(A,B)$ and $(\text{Hom}_{\mathcal{A}}(f,g))(p) = g \circ p \circ f$, for $A' \xrightarrow{f} A \xrightarrow{p} B \xrightarrow{g} B'$.

Remarks 3.1.14 a. The existence of the functor $\text{Hom}_{\mathcal{A}}$ is a bit like the fact that for a metric space (X, d), the distance function is itself a continuous map $d : X \times X \longrightarrow \mathbb{R}$. (If you take two points and move each one slightly, the distance between them changes only slightly.)

b. You could also define $\mathrm{Hom}_{\mathcal{A}}$ using Sheet 1, q.5: it's the functor $\mathcal{A}^{\mathrm{op}} \times \mathcal{A} \longrightarrow \mathbf{Set}$ corresponding to the families $(H^A)_{A \in \mathcal{A}}$ and $(H_B)_{B \in \mathcal{A}}$ of functors, which satisfy the two conditions in the question.

c. In 2.1.5 we saw that for any set Y there is an adjunction $(- \times Y) \dashv (-)^Y$ of functors $\mathbf{Set} \longrightarrow \mathbf{Set}$. Similarly, for any category $\mathcal{B}$ there is an adjunction $(- \times \mathcal{B}) \dashv [\mathcal{B}, -]$ of functors $\mathbf{CAT} \longrightarrow \mathbf{CAT}$; in other words, there is a canonical bijection

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

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

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

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

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

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

The composite functor $\downarrow_{\to}$ sends (A, B) to $\mathcal{B}(F(A), B)$, and the composite $^{\to}\!\!\downarrow$ sends (A, B) to $\mathcal{A}(A, G(B))$. So the statement '$\mathcal{B}(F(A), B) \cong \mathcal{A}(A, G(B))$ naturally in A and B' now makes sense: it means that the functors $\downarrow_{\to}$ and $^{\to}\!\!\downarrow$ are naturally isomorphic.

Exercise 3.1.15 Show that this naturality requirement is equivalent to the naturality requirements (2:2) and (2:3) (p.24) that we used in Chapter 2. (You might want to use the result of Exercise 2.2.1.)

3.2 The Yoneda Lemma

What do representables see?

Recall from 1.2.6 that functors $\mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$ are sometimes called 'presheaves' on $\mathcal{A}$. So for each $A \in \mathcal{A}$ we have a representable presheaf H_A, and we're asking how the rest of the presheaf category $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ looks from H_A. In other words, if X is another presheaf, what are the maps $H_A \longrightarrow X$?

It's worth spending some time making sure that you understand this question before you try to answer it. So let's go through it again carefully.

We start by fixing a locally small category $\mathcal{A}$. We then take an object $A \in \mathcal{A}$ and a presheaf X on $\mathcal{A}$, that is, a functor $X : \mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$. We also have the presheaf H_A on $\mathcal{A}$. The question is: what are the maps $H_A \longrightarrow X$? Since H_A and X are both objects of the presheaf category $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$, the 'maps' concerned are maps in $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$. So, we're asking what natural transformations

$$\begin{tikzcd} \mathcal{A}^{\mathrm{op}} \arrow[r, bend left=50, "H_A"{name=U}] \arrow[r, bend right=50, "X"'{name=D}] & \mathbf{Set} \arrow[Rightarrow, from=U, to=D] \end{tikzcd} \tag{3:1}$$

there are. The set of such natural transformations is

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

(This expression may look scary, but it's just a special case of the notation $\mathcal{B}(B, B')$ for the set of maps $B \longrightarrow B'$ in a category $\mathcal{B}$. Here, $\mathcal{B} = [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$, $B = H_A$, and $B' = X$.) We want to know what this set is.

There's an informal principle that allows us to guess the answer. Look back at Remarks 1.1.3(a), 1.2.2(a) and 1.3.2(a) on the definitions of category, functor and natural transformation. Each remark is of the form 'input of one given type produces exactly one output of another given type'. (More snappily: 'there's only one way from A to B'.) For example, in 1.1.3(a) the input is a sequence of maps $A_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} A_n$, the output is a map $A_0 \longrightarrow A_n$, and the point is that no matter what you do with the input data $f_1, \ldots, f_n$, there's only one map $A_0 \longrightarrow A_n$ that you can produce from it.

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

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

for all $A \in \mathcal{A}$ and $X \in [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$. This turns out to be true: and that's the Yoneda Lemma.

Another way of saying this is that for any A and X, the natural transformations (3:1) correspond one-to-one with the elements of $X(A)$. We began the section with the question 'what do representables see?' More precisely put: given an object A and a presheaf X, what does H_A see in X? Yoneda answers: it sees the elements of $X(A)$.

The hardest thing about the Yoneda Lemma is understanding the statement, which may seem bewilderingly abstract and general at first. Here it is.

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

$$[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \cong X(A) \tag{3:2}$$

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

This is exactly what we had above, except that the word 'naturally' has snuck in. Recall from the bottom of page 12 that if $F, G : \mathcal{C} \longrightarrow \mathcal{D}$ are a pair of functors, 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 (3:2) must be functorial in both A and X. This means, for instance, that a map $X \longrightarrow X'$ must induce a map

$$[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \longrightarrow [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X'),$$

and that not only does the isomorphism (3:2) hold for *every* A and X, but also, the isomorphisms can be chosen in a way that's compatible with such induced maps. Precisely, the Yoneda Lemma states that the composite functor

$$\begin{array}{ccccc} \mathcal{A}^{\mathrm{op}} \times [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] & \xrightarrow{H_\bullet^{\mathrm{op}} \times 1} & [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]^{\mathrm{op}} \times [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] & \xrightarrow{\mathrm{Hom}_{[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]}} & \mathbf{Set} \\ (A, X) & \longmapsto & (H_A, X) & \longmapsto & [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X) \end{array}$$

is naturally isomorphic to the 'evaluation' functor

$$\begin{array}{ccc} \mathcal{A}^{\mathrm{op}} \times [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}] & \xrightarrow{\mathrm{ev}} & \mathbf{Set} \\ (A, X) & \longmapsto & X(A). \end{array}$$

If the Yoneda Lemma were false then the world would look very different, and much more complex. For take a presheaf $X : \mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$, and define a new presheaf X' by

$$X' = [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_\bullet, X) : \mathcal{A}^{\mathrm{op}} \longrightarrow \mathbf{Set}$$

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

We'll do the proof next week. For now, your task is to understand the statement. You can test this as follows: if you've *really* understood the statement, you should be able to work out the proof for yourself. From beginning to end, there's only one thing you can possibly do.

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

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

$$\begin{tikzcd}
{[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X)} \arrow[r, shift left, "\widehat{(\ )}"] & X(A) \arrow[l, shift left, "\widetilde{(\ )}"]
\end{tikzcd}$$

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

- Given $\alpha : H_A \longrightarrow X$, define $\widehat{\alpha} = \alpha_A(1_A) \in X(A)$. (What else could we possibly do?)

- Let $x \in X(A)$. We have to define a natural transformation $\widetilde{x} : H_A \longrightarrow X$. That is, we have to define for each $B \in \mathcal{A}$ a function

  $$\widetilde{x}_B : H_A(B) = \mathcal{A}(B, A) \longrightarrow X(B)$$

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

  Given $B \in \mathcal{A}$ and $f \in \mathcal{A}(B, A)$, put

  $$\widetilde{x}_B(f) = (Xf)(x) \in X(B),$$

  which makes sense as Xf is a map $X(A) \longrightarrow X(B)$. (What else could we possibly do?) To prove naturality, we must show that for any map $B' \xrightarrow{g} B$ in $\mathcal{A}$, the square

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

  commutes. Indeed, for all $f \in \mathcal{A}(B, A)$ we have

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

  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 $\widehat{\widetilde{x}} = x$. And indeed,

  $$\widehat{\widetilde{x}} = \widetilde{x}_A(1_A) = (X1_A)(x) = 1_{X(A)}(x) = x.$$

- Given $\alpha : H_A \longrightarrow X$, we have to show that $\widetilde{\widehat{\alpha}} = \alpha$. Two natural transformations are equal if and only if all their components are equal, so we have to show that $(\widetilde{\widehat{\alpha}})_B = \alpha_B$ for all $B \in \mathcal{A}$. Each side of this equation is a function from $H_A(B) = \mathcal{A}(B, A)$ 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

  $$(\widetilde{\widehat{\alpha}})_B(f) = \alpha_B(f)$$

  for all $B \in \mathcal{A}$ and $f : B \longrightarrow A$ in $\mathcal{A}$. The left-hand side is by definition

  $$(\widetilde{\widehat{\alpha}})_B(f) = (Xf)(\widehat{\alpha}) = (Xf)(\alpha_A(1_A)),$$

  so we have to show that

  $$(Xf)(\alpha_A(1_A)) = \alpha_B(f). \tag{3:3}$$

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

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

  commutes, which when taken at $1_A \in \mathcal{A}(A, A)$ gives equation (3:3).

(It's worth pausing to think about the significance of the fact that $\widetilde{\widehat{\alpha}} = \alpha$. Since $\widehat{\alpha}$ is the value of α at 1_A, this tells us that

$$\boxed{\textbf{A natural transformation } H_A \longrightarrow X \textbf{ is entirely determined by its value at } 1_A}$$

...which is an element of $X(A)$. Precisely *how* a natural transformation is determined by its value at 1_A is described in equation (3:3).)

This establishes the bijection for each A and X. We now show that the bijection is natural in A and X. *A priori* we have to prove naturality of both $\widehat{(\;)}$ and $\widetilde{(\;)}$, but by Lemma 1.3.6, it is enough to prove naturality of just one of them. We prove naturality of $\widehat{(\;)}$. By Sheet 5, q.5, $\widehat{(\;)}$ 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.

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

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

commutes. For $\alpha : H_A \longrightarrow X$, 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 $(\alpha \circ H_f)_B(1_B) = (Xf)(\alpha_A(1_A))$. But

$$(\alpha \circ H_f)_B(1_B) = \alpha_B((H_f)_B(1_B)) = \alpha_B(f \circ 1_B) = \alpha_B(f) = (Xf)(\alpha_A(1_A)),$$

the first step by definition of composition in $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$, the second by definition of H_f, and the last by (3:3).

- Naturality in X says that for each $A \in \mathcal{A}$ and map

$$\begin{tikzcd}
\mathcal{A}^{\mathrm{op}} \arrow[r, bend left=50, "X"{name=U}] \arrow[r, bend right=50, "X'"'{name=D}] & \mathbf{Set}
\arrow[Rightarrow, from=U, to=D, "\theta"]
\end{tikzcd}$$

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

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

commutes. For $\alpha : H_A \longrightarrow X$, 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 $[\mathcal{A}^{\text{op}}, \mathbf{Set}]$, so the diagram commutes.

This completes the proof. □

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 $\sim$, as in $A \xrightarrow{\sim} B$, denotes an isomorphism.

A representation is a universal element

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

$$\text{for each } B \in \mathcal{A} \text{ and } x \in X(B)\text{, there is a unique map } f : B \longrightarrow A \text{ such that } (Xf)(u) = x. \tag{3:4}$$

To clarify the statement, first recall that by definition, a representation of X is an object $A \in \mathcal{A}$ together with a natural isomorphism $\alpha : H_A \xrightarrow{\sim} X$. The Corollary says that such pairs (A, α) are in natural one-to-one correspondence with pairs (A, u) satisfying (3:4).

Pairs (B, x) with $B \in \mathcal{A}$ and $x \in X(B)$ are sometimes called **elements** of the presheaf X. An element u satisfying (3:4) is sometimes called a **universal** element of X.

Proof By the Yoneda Lemma, we have only to show that for $A \in \mathcal{A}$ and $u \in X(A)$, the natural transformation $\widetilde{u} : H_A \longrightarrow X$ is an isomorphism if and only if (3:4) holds. Now, $\widetilde{u}$ is an isomorphism if and only if for all $B \in \mathcal{A}$, the function

$$\widetilde{u}_B : H_A(B) = \mathcal{A}(B, A) \longrightarrow X(B)$$

is a bijection, if and only if for all $B \in \mathcal{A}$ and $x \in X(B)$, there is a unique map $f : B \longrightarrow A$ such that $\widetilde{u}_B(f) = x$. But $\widetilde{u}_B(f) = (Xf)(u)$, so this is exactly condition (3:4). □

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

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

$$\text{for each } B \in \mathcal{A} \text{ and } x \in X(B)\text{, there is a unique map } f : A \longrightarrow B \text{ such that } (Xf)(u) = x. \tag{3:5}$$

Proof Follows immediately by duality. □

Example 3.3.4 Fix a set S and consider the functor

$$\begin{array}{rccc} X = \mathbf{Set}(S, U(-)) : & \mathbf{Vect}_k & \longrightarrow & \mathbf{Set}, \\ & V & \longmapsto & \mathbf{Set}(S, U(V)). \end{array}$$

Here are two familiar statements about X:

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

$$\mathbf{Vect}_k(F(S), V) \cong \mathbf{Set}(S, U(V)) \tag{3:6}$$

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

b. there are a vector space $F(S)$ and a function $u : S \longrightarrow U(F(S))$ such that

> for each vector space V and function $f : S \longrightarrow U(V)$, there is a unique linear map $\overline{f} : F(S) \longrightarrow V$ such that
>
> $$\begin{tikzcd} S \arrow[r, "u"] \arrow[dr, "f"'] & U(F(S)) \arrow[d, "U(\overline{f})"] \\ & U(V) \end{tikzcd}$$
>
> commutes

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

These two statements both say that X is representable.

Statement (a) says that we have an isomorphism $X(V) \cong \mathbf{Set}(F(S), V)$ 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 (3:5). 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's 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 $g \longmapsto U(g) \circ u$. But it looks at first as if (b) says rather more than (a): that not only are the two things naturally isomorphic, they are naturally isomorphic in a rather special way. The Corollary tells us that this is an illusion: all natural isomorphisms (3:6) arise in this way. It's the word 'natural' in (a) that hides all the explicit detail.

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

$$X = \mathcal{A}(A, G(-)) : \mathcal{B} \longrightarrow \mathbf{Set}.$$

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

a. $\mathcal{A}(A, G(B)) \cong \mathcal{B}(F(A), B)$ naturally in B; that is, $X \cong H^{F(A)}$

b. the unit map $\eta_A : A \longrightarrow G(F(A))$ is an initial object of the comma category $(A \Rightarrow G)$; that is, $\eta_A \in X(F(A))$ satisfies condition (3:5).

This observation can be developed to give an alternative proof of Proposition 2.3.5, the reformulation of adjointness in terms of initial objects.

Example 3.3.6 In sheet 2, q.4 you saw that for any group G and element $x \in G$, there is a unique homomorphism $\phi : \mathbb{Z} \longrightarrow G$ such that $\phi(1) = x$. This says that $1 \in U(\mathbb{Z})$ is a universal element of the forgetful functor $U : \mathbf{Gp} \longrightarrow \mathbf{Set}$; in other words, condition (3:5) holds when $\mathcal{A} = \mathbf{Gp}$, $X = U$, $A = \mathbb{Z}$ 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. So -1 gives a different isomorphism $H^{\mathbb{Z}} \xrightarrow{\sim} U$. They are different because Corollary 3.3.3 provides a *one-to-one* correspondence between universal elements and representations.

There are no further universal elements $u \in U(\mathbb{Z})$ (exercise). So there are exactly two isomorphisms $H^{\mathbb{Z}} \xrightarrow{\sim} U$.

The Yoneda embedding

Here is the second consequence of the Yoneda Lemma.

Corollary 3.3.7 *For any locally small category $\mathcal{A}$, the functor*

$$H_\bullet : \mathcal{A} \longrightarrow [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$$

is full and faithful.

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

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

$$\begin{array}{ccc} \mathcal{A}(A, A') & \longrightarrow & [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, H_{A'}) \\ f & \longmapsto & H_f \end{array}$$

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

$$\widetilde{(\ )} : H_{A'}(A) \longrightarrow [\mathcal{A}^{\mathrm{op}}, \mathbf{Set}](H_A, H_{A'})$$

is bijective, so it's enough to prove that these two functions are equal. Indeed, if $f \in \mathcal{A}(A, A')$ then

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

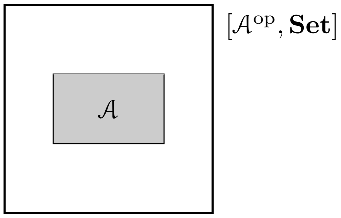
so $\widetilde{f} = \widetilde{\widehat{H_f}} = H_f$, as required. □

In lectures I wrote up the following corollary of Proposition 1.3.12:

> Let $J : \mathcal{C} \longrightarrow \mathcal{D}$ be a full and faithful functor. Then $\mathcal{C}$ is equivalent to the full subcategory of $\mathcal{D}$ whose objects are those of the form $J(C)$ for some $C \in \mathcal{C}$.

(This is true whether we take 'of the form $J(C)$ for some C' to mean 'equal to $J(C)$ for some C' or 'isomorphic to $J(C)$ for some C'.) The lesson is that a full and faithful functor $J : \mathcal{C} \longrightarrow \mathcal{D}$ can be thought of as 'embedding' $\mathcal{C}$ into $\mathcal{D}$ as a full subcategory.

The functor $H_\bullet$ is called the **Yoneda embedding** of $\mathcal{A}$. It embeds $\mathcal{A}$ into the category $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ of presheaves:



So $\mathcal{A}$ is equivalent to the full subcategory of $[\mathcal{A}^{\mathrm{op}}, \mathbf{Set}]$ whose objects are the representable functors.

Later we'll see how that every presheaf can be built up from representables, in very roughly the same way that every natural number can be built up from primes.

Isomorphism of representables

Full subcategories are the easiest subcategories to handle. For instance, given objects C and C' of the subcategory, you can talk unambiguously about 'maps from C to C'': it makes no difference whether you interpret that as meaning maps in the subcategory or maps in the whole category. Similarly, you can talk unambiguously about isomorphism of objects of the subcategory, as follows from this lemma:

Lemma 3.3.8 *Let $J : \mathcal{C} \longrightarrow \mathcal{D}$ be a full and faithful functor and $C, C' \in \mathcal{C}$. Then:*

a. for any map $f : C \longrightarrow C'$ in $\mathcal{C}$, if $J(f)$ is an isomorphism then f is an isomorphism *b.* *for any isomorphism $g : J(C) \longrightarrow J(C')$ in $\mathcal{D}$, there is a unique isomorphism $f : C \longrightarrow C'$ in $\mathcal{C}$ such that $J(f) = g$*

c. *if $J(C) \cong J(C')$ then $C \cong C'$.*

Proof Exercise. □

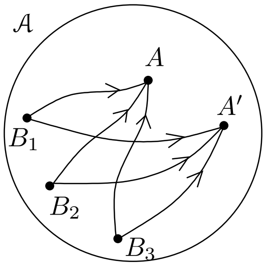
Here is the third consequence of the Yoneda Lemma.

Corollary 3.3.9 *Let $\mathcal{A}$ be a locally small category and $A, A' \in \mathcal{A}$. 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 '$\iff$'. The '$\Leftarrow$' is trivial, since $H_\bullet$, like all functors, preserves isomorphism (Sheet 1, q.2). The '$\Rightarrow$' follows from Corollary 3.3.7 and Lemma 3.3.8(c). □

The force of this is that $H_A \cong H_{A'} \Rightarrow A \cong A'$. In other words, if $\mathcal{A}(B, A) \cong \mathcal{A}(B, A')$ naturally in B, then $A \cong A'$. Think of $\mathcal{A}(B, A)$ as 'A viewed from B'; then the corollary tells us that two objects are the same if and only if they look the same from all viewpoints. (If it walks like a duck and quacks like a duck, it probably is a duck.)



For example, take $\mathcal{A} = \mathbf{Gp}$ and two groups A and A'. 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 $H_A(1) = \mathbf{Gp}(1, A)$ 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 A'. So A and A' have isomorphic underlying sets, though perhaps quite 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 3.1.4(a), A and A' have the same number of elements of order p.

Each of these gives only partial information about the similarity of A and A', but the whole natural isomorphism $H_A \cong H_{A'}$ implies that $A \cong A'$.

The category of sets is very unusual in this respect. If A is a set then

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

so $H_A(1) \cong H_{A'}(1) \Rightarrow A \cong A'$. 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: the only thing that matters about a set is its elements!

Example 3.3.10 Let $G : \mathcal{B} \longrightarrow \mathcal{A}$ and suppose that both F and F' are left adjoint to G. Then for each $A \in \mathcal{A}$, we have

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

naturally in $B \in \mathcal{B}$, so $H^{F(A)} \cong H^{F'(A)}$, so (by the Corollary) $F(A) \cong F'(A)$. In fact this isomorphism is natural in A, so that $F \cong F'$. This shows that left adjoints are unique, as claimed in 2.1.2(c). Dually, right adjoints are unique. See also Sheet 6, q.4.

Example 3.3.11 Corollary 3.3.9 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

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

of Example 3.1.4(e). Corollary 3.3.9 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 T's are isomorphic, it is legitimate to refer to any of them as *the* tensor product of U and V.

We have just deduced this fact from the Yoneda Lemma, but we also proved it directly in Lemma 0.1.4.

Chapter 4

Limits

Limits—and the dual concept, colimits—provide our third approach to the idea of universal property.

They also provide a unified approach to many familiar constructions in mathematics. Whenever you meet a method for taking some objects and maps of a particular category and building a new object out of them, there's a good chance you're looking at either a limit or a colimit. For instance, in group theory, you can take a pair of groups and a homomorphism between them and form the kernel, which is a new group. This construction is an example of a limit in the category of groups. Or, you might take two natural numbers and form their lowest common multiple. This is an example of a colimit in the category of natural numbers ordered by divisibility.

4.1 Limits: basics

Limits are a *really* general concept. They're so general that even examples of limits, such as products and equalizers, are still general concepts, applicable in any category. And even when you look at products, equalizers, etc. in a particular category, those are still *constructions* rather than actual *things*.

Despite their generality, limits are not so hard to understand: you're already familiar with lots of examples. We'll begin with some examples to motivate the definition of product; then we'll have some different examples to motivate the definition of equalizer; and so on. Once we have these definitions, we'll be ready to see the most general definition: limit!

Products

Example 4.1.1 (Spaces) Any two topological spaces X and Y have a product, $X \times Y$, which is another topological space. It is the set of pairs (x, y) with $x \in X$ and $y \in Y$, equipped with the so-called product topology. The product topology is designed so that, for instance, a function

$$\begin{array}{ccc} \mathbb{R} & \longrightarrow & X \times Y \\ t & \longmapsto & (x(t), y(t)) \end{array}$$

is continuous if and only if it is continuous in each coordinate: in other words, both functions

$$t \longmapsto x(t), \qquad t \longmapsto y(t)$$

are continuous. So a continuous map $\mathbb{R} \longrightarrow X \times Y$ amounts to a continuous map $\mathbb{R} \longrightarrow X$ together with a continuous map $\mathbb{R} \longrightarrow Y$. The same is true when $\mathbb{R}$ is replaced by any other space A.

What lies behind that phrase 'amounts to'? To answer this, first observe that we have continuous projection maps

$$\begin{tikzcd}
& X \times Y \arrow[dl, "\mathrm{pr}_1"'] \arrow[dr, "\mathrm{pr}_2"] & & & (x,y) \arrow[dl, mapsto] \arrow[dr, mapsto] & \\
X & & Y & x & & y.
\end{tikzcd} \tag{4:1}$$

Now, given a space A and a continuous map $h : A \longrightarrow X \times Y$, we obtain a pair of continuous maps

$$\begin{tikzcd}
& A \arrow[dl, "\mathrm{pr}_1 \circ h"'] \arrow[dr, "\mathrm{pr}_2 \circ h"] & \\
X & & Y.
\end{tikzcd}$$

Conversely, given a space A and a pair of continuous maps

$$\begin{tikzcd}
& A \arrow[dl, "f"'] \arrow[dr, "g"] & \\
X & & Y
\end{tikzcd} \tag{4:2}$$

we obtain a continuous map

$$h : A \longrightarrow X \times Y$$

defined by $h(a) = (f(a), g(a))$. This map h makes the diagram

$$\begin{tikzcd}
& A \arrow[d, "h"] \arrow[ddl, "f"'] \arrow[ddr, "g"] & \\
& X \times Y \arrow[dl, "\mathrm{pr}_1"] \arrow[dr, "\mathrm{pr}_2"'] & \\
X & & Y
\end{tikzcd} \tag{4:3}$$

commute. Moreover, it's the *only* map making it commute. For suppose that $h' : A \longrightarrow X \times Y$ also works. Let $a \in A$, and write $h'(a) = (x, y)$. Then $f(a) =$

$\mathrm{pr}_1(h'(a)) = \mathrm{pr}_1(x, y) = x$; similarly $g(a) = y$. Hence $h'(a) = (f(a), g(a)) = h(a)$ for all a, and $h' = h$.

To summarize: take spaces X and Y. Let's say that a **cone** on (X, Y) is a diagram of the form (4:2). (Picture the cone with the base at the bottom and the pointed end at A.) Then the product $X \times Y$ and its projections form a cone (4:1) on (X, Y). This cone has the following special property: for any cone (4:2) on (X, Y), there is a unique map $h : A \longrightarrow X \times Y$ making (4:3) commute.

You can show that this property characterizes the product and its projections uniquely, up to isomorphism. (Compare Lemma 0.1.4.) We'll prove something much more general later.

Example 4.1.2 (Sets) The last example can be repeated with sets in place of topological spaces: simply forget about the topologies and the continuity.

Example 4.1.3 (Vector spaces) Now let X and Y be vector spaces. We can form their direct sum, $X \oplus Y$, whose elements can be written either (x, y) or $x + y$ (with $x \in X$ and $y \in Y$), according to taste. There are linear projection maps

$$\begin{tikzcd}
& X \oplus Y \arrow[dl, "\mathrm{pr}_1"'] \arrow[dr, "\mathrm{pr}_2"] & & & (x,y) \arrow[dl, mapsto] \arrow[dr, mapsto] & \\
X & & Y & x & & y.
\end{tikzcd}$$

It can easily be shown (try it!) that for any vector space A and pair (4:2) of linear maps, there is a unique linear map $h : A \longrightarrow X \oplus Y$ such that diagram (4:3) (with $\times$ changed to $\oplus$) commutes .

Examples 4.1.4 (Elements of ordered sets)

a. Let $x, y \in \mathbb{R}$. Then $\min\{x, y\}$ satisfies

$$\min\{x, y\} \leq x, \qquad \min\{x, y\} \leq y$$

and has the further property that if $a \in \mathbb{R}$ and

$$a \leq x, \qquad a \leq y$$

then $a \leq \min\{x, y\}$.

b. Fix a set S. Let $X, Y \in \mathcal{P}(S)$. Then $X \cap Y$ satisfies

$$X \cap Y \subseteq X, \qquad X \cap Y \subseteq Y$$

and has the further property that if $A \in \mathcal{P}(S)$ and

$$A \subseteq X, \qquad A \subseteq Y$$

then $A \subseteq X \cap Y$.

c. Let $x, y \in \mathbb{N}$. Their greatest common divisor (highest common factor) $\gcd(x, y)$ satisfies

$$\gcd(x,y)|x, \qquad \gcd(x,y)|y$$

(it's a common divisor!) and has the further property that if $a \in \mathbb{N}$ and

$$a|x, \qquad a|y$$

then $a|\gcd(x, y)$.

There's obviously a pattern here. In each example we're dealing with an ordered set: $(\mathbb{R}, \leq)$, $(\mathcal{P}(S), \subseteq)$ and $(\mathbb{N}, |)$. Let $(A, \leq)$ be an ordered set and $x, y \in A$. A **lower bound** for x and y is an element $a \in A$ such that $a \leq x$ and $a \leq y$. A **greatest lower bound** for x and y is a lower bound z for x and y with the further property that if a is a lower bound for x and y then $a \leq z$.

Greatest lower bounds do not always exist, but when they do, they are unique. They are also called **meets**. The greatest lower bound or meet of x and y, when it exists, is written $x \wedge y$. So in our examples, $x \wedge y = \min\{x, y\}$, $X \wedge Y = X \cap Y$, and $x \wedge y = \gcd(x, y)$.

Here is the general definition.

Definition 4.1.5 Let $\mathcal{A}$ be a category and $X, Y \in \mathcal{A}$. A **cone** on (X, Y) is a diagram (4:2) in $\mathcal{A}$. A **product** of X and Y is a cone

$$\begin{tikzcd} & P \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\ X & & Y \end{tikzcd}$$

with the property that for any cone (4:2) on (X, Y), there is a unique map $h : A \longrightarrow P$ in $\mathcal{A}$ such that

$$\begin{tikzcd} & A \arrow[d, "h"] \arrow[ddl, "f"'] \arrow[ddr, "g"] & \\ & P \arrow[dl, "p_1"] \arrow[dr, "p_2"'] & \\ X & & Y \end{tikzcd}$$

commutes. The maps p_1 and p_2 are called the **projections**.

Remarks 4.1.6 a. Products do not always exist. For example, if $\mathcal{A}$ is the discrete two-object category

$$\boxed{X\bullet \qquad \bullet Y}$$

then X and Y do not have a product—indeed, there are no cones on (X, Y) at all. But when X and Y do have a product, it is unique up to isomorphism (as we will show). This justifies talking about *the* product of X and Y.

b. Strictly speaking, the product consists of the object P *together with* the projections p_1 and p_2. But informally, we often refer to P alone as the product of X and Y. We write $P = X \times Y$.

Examples 4.1.7 In **Top** and **Set**, product is × in the usual sense, as shown above. In $\mathbf{Vect}_k$, product is direct sum, $\oplus$. In an ordered set, a cone on (x, y) is a lower bound for x and y, and the product of x and y (if it exists) is their meet, $x \wedge y$.

We've been discussing the product $X \times Y$ of two objects, but there's no reason to stick to two: we can just as well talk about products $X \times Y \times Z$ of three objects, or of infinitely many objects. The definition changes in the most obvious way.

Definition 4.1.8 Let $\mathcal{A}$ be a category, I a set, and $(X_i)_{i \in I}$ a family of objects of $\mathcal{A}$. A **cone** on $(X_i)_{i \in I}$ is an object $A \in \mathcal{A}$ together with a family

$$(A \xrightarrow{f_i} X_i)_{i \in I} \tag{4:4}$$

of maps in $\mathcal{A}$. A **product** of $(X_i)_{i \in I}$ is a cone

$$(P \xrightarrow{p_i} X_i)_{i \in I}$$

with the property that for any cone (4:4) on $(X_i)_{i \in I}$, there is a unique map $h : A \longrightarrow P$ such that

$$f_i = p_i \circ h$$

for all $i \in I$. The maps p_i are called the **projections**.

Remarks 4.1.6 apply equally to this definition. When the product P exists, we write $P = \prod_{i \in I} X_i$. Taking I to be a two-element set, we recover the special case 4.1.5 of binary products.

Examples 4.1.9 a. **Real infima** In ordered sets, the extension from binary to arbitrary products works in the obvious way: given an ordered set $(A, \leq)$, a **lower bound** for a family $(x_i)_{i \in I}$ of elements is an element $a \in A$ such that $a \leq x_i$ for all i; a **greatest lower bound** or **meet** of the family is a lower bound greater than any other, written $\bigwedge_{i \in I} x_i$. These are the products in $(A, \leq)$.

For example, in $\mathbb{R}$ with its usual ordering, the greatest lower bound of a family $(x_i)_{i \in I}$ is $\inf\{x_i \mid i \in I\}$ (and one exists if and only if the other does).

b. **Terminal objects** What happens when the indexing set I is empty? Let $\mathcal{A}$ be a category. In general, an I-indexed family $(X_i)_{i \in I}$ of objects of $\mathcal{A}$ is a function $I \longrightarrow \mathrm{ob}(\mathcal{A})$. When I is empty, there is exactly one such function; in other words, there is exactly one family $(X_i)_{i \in \emptyset}$, the 'empty family'.

> A cone on the empty family is an object $A \in \mathcal{A}$ together with, for each $i \in \emptyset$, a map $A \longrightarrow X_i$. But there are no elements $i \in \emptyset$, so a cone on the empty family is simply an object $A \in \mathcal{A}$. A product of the empty family is a cone—that is, an object $P \in \mathcal{A}$—with the property that for any cone (object) $A \in \mathcal{A}$, there is a unique map $h : A \longrightarrow P$. So a product of the empty family is exactly a terminal object.
>
> We've been writing 1 for terminal objects, which was justified by the fact that in categories such as **Set**, **Top**, **Ring** and **Gp**, the terminal object has just one element. But here we see that the terminal object is the product of no things, which in the context of elementary arithmetic is the number 1. This is a second, related, reason for the notation.

Now take a set I and an object X of a category $\mathcal{A}$. There is a constant family $(X)_{i \in I}$. Its product $\prod_{i \in I} X$, if it exists, is written X^I and called a **power** of X.

Example 4.1.10 (Function sets) We've already met powers in **Set**, in Section $2\frac{1}{2}$.1. Here X^I is the set of functions from I to X, also written $\mathbf{Set}(I, X)$.

Equalizers

Example 4.1.11 (Sets) In Section $2\frac{1}{2}$.1 we defined the equalizer E of two functions $X \overset{f}{\underset{g}{\rightrightarrows}} Y$. Write $i : E \longrightarrow X$ for the inclusion; then we have maps

$$\begin{tikzcd} E \arrow[r, "i"] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y. \end{tikzcd} \qquad (4{:}5)$$

If $e \in E$ then—by the very definition of E!—we have $f(i(e)) = g(i(e))$. So $f \circ i = g \circ i$. A diagram (4:5) of objects and maps in a category is called a **fork** if $f \circ i = g \circ i$; so this is a fork in **Set**. Moreover, it's the 'best possible fork', in the sense that for any fork

$$\begin{tikzcd} A \arrow[r, "q"] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y, \end{tikzcd} \qquad (4{:}6)$$

there's a unique function $\overline{q} : A \longrightarrow E$ such that

$$\begin{tikzcd} A \arrow[d, "\overline{q}"'] \arrow[dr, "q"] & \\ E \arrow[r, "i"'] & X \end{tikzcd} \qquad (4{:}7)$$

commutes. (Proof: exercise.)

Example 4.1.12 (Spaces) Given topological spaces X, Y and continuous maps $X \overset{f}{\underset{g}{\rightrightarrows}} Y$, we can form their equalizer E and the inclusion $i : E \longrightarrow X$

as above. Since E is a subset of the space X, it acquires the subspace topology from X, and i is then continuous. We then have a fork (4:5) in **Top**, which has the analogous universal property. (Checking this amounts to checking that for any fork (4:6) in **Top**, the induced function $\overline{q}$ is continuous. This follows from the definition of subspace topology.)

Example 4.1.13 (Groups) Let $\theta : G \longrightarrow H$ be a homomorphism of groups. With notation as in Example 0.1.5, θ gives rise to a fork

$$\begin{tikzcd}
\ker\theta \arrow[r, hook, "\iota"] & G \arrow[r, shift left, "\theta"] \arrow[r, shift right, "\varepsilon"'] & H.
\end{tikzcd}$$

It has a universal property like those in the previous two examples.

Definition 4.1.14 Let $\mathcal{A}$ be a category and let $\begin{tikzcd} X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \end{tikzcd}$ be objects and maps in $\mathcal{A}$. An **equalizer** of f and g is a map $E \xrightarrow{i} X$ in $\mathcal{A}$ such that

$$\begin{tikzcd}
E \arrow[r, "i"] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y
\end{tikzcd}$$

is a fork, and with the further property that for any fork (4:6), there is a unique map $\overline{q} : A \longrightarrow E$ making (4:7) commute.

Remarks 4.1.6 on products all apply to equalizers too.

Examples 4.1.15 a. Equalizers in **Set** and **Top** are as described above.

b. The equalizer of $\begin{tikzcd} G \arrow[r, shift left, "\theta"] \arrow[r, shift right, "\varepsilon"'] & H \end{tikzcd}$ in **Gp**, where ε is the trivial homomorphism, is $\ker\theta$. (We haven't said what the equalizer of an arbitrary parallel pair $\begin{tikzcd} G \arrow[r, shift left, "\theta"] \arrow[r, shift right, "\phi"'] & H \end{tikzcd}$ in **Gp** is.)

c. In $\mathbf{Vect}_k$, the equalizer of linear maps $\begin{tikzcd} V \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & W \end{tikzcd}$ is $\ker(f - g)$.

Pullbacks

Example 4.1.16 (Sets and inverse images) A basic construction with sets and functions is the formation of inverse images (also called preimages). Given a function $f : Y \longrightarrow X$ and a subset $X' \subseteq X$, we obtain a new set, the inverse image

$$f^{-1}X' = \{y \in Y \mid f(y) \in X'\}$$

of X' under f, and a new function,

$$\begin{array}{rccc} f' : & f^{-1}X' & \longrightarrow & X' \\ & y & \longmapsto & f(y). \end{array}$$

We also have the inclusion functions $i : X' \hookrightarrow X$ and $j : f^{-1}X' \hookrightarrow Y$. Putting everything together gives a commutative square

$$\begin{tikzcd}
f^{-1}X' \arrow[r, "f'"] \arrow[d, hook, "j"'] & X' \arrow[d, hook, "i"] \\
Y \arrow[r, "f"'] & X.
\end{tikzcd} \tag{4:8}$$

Note that the data we started with was the lower-right part of this square (X, Y, X', f and i), and from it we constructed the rest of the square ($f^{-1}X'$, f' and j). It's the 'best possible' commutative square on the starting data: for any commutative square

$$\begin{tikzcd}
A \arrow[r, "g"] \arrow[d, "k"'] & X' \arrow[d, hook, "i"] \\
Y \arrow[r, "f"'] & X,
\end{tikzcd}$$

there is a unique map $h : A \longrightarrow f^{-1}X'$ such that

$$\begin{tikzcd}
A \arrow[dr, "h"] \arrow[drr, bend left, "g"] \arrow[ddr, bend right, "k"'] & & \\
& f^{-1}X' \arrow[r, "f'"'] \arrow[d, hook, "j"] & X' \arrow[d, hook, "i"] \\
& Y \arrow[r, "f"'] & X
\end{tikzcd}$$

commutes. Proof: for uniqueness, let h be a map for which the diagram commutes. Then for all $a \in A$, we have $j(h(a)) = k(a)$, that is, $h(a) = k(a)$, and this determines h uniquely. For existence, first note that if $a \in A$ then $f(k(a)) = i(g(a)) \in X'$, so we may define $h : A \longrightarrow f^{-1}X'$ by $h(a) = k(a)$ for all a. Then for all a, we have $j(h(a)) = h(a) = k(a)$ and

$$f'(h(a)) = f(h(a)) = f(k(a)) = i(g(a)) = g(a),$$

so $j \circ h = k$ and $f' \circ h = g$, as required.

Example 4.1.17 (Groups and inverse images) Now suppose that $Y \xrightarrow{f} X$ is a homomorphism of groups and X' a subgroup of X. The inverse image of a subgroup is a subgroup, so (4:8) becomes a commutative square in **Gp**. It can be shown that it has the same universal property in the category of groups. (The extra thing to be checked is that the map 'h' is a homomorphism.)

Example 4.1.18 (Sets and intersections) If X' and X'' are two subsets of another set, X, we may form their intersection, $X' \cap X''$. We then have a square

$$\begin{tikzcd}
X' \cap X'' \arrow[r, hook, "m"] \arrow[d, hook, "j"'] & X' \arrow[d, hook, "i"] \\
X'' \arrow[r, hook, "l"'] & X,
\end{tikzcd} \tag{4:9}$$

where all the arrows are inclusions of subsets. In fact, this is a special case of the first example (4.1.16): we have a function $X'' \xrightarrow{l} X$ and a subset $X' \subseteq X$, and

$$l^{-1}X' = \{x'' \in X'' \mid l(x'') \in X'\} = \{x'' \in X'' \mid x'' \in X'\} = X' \cap X''.$$

So intersections are a special case of inverse images. In particular, (4:9) has the usual universal property.

In each of our examples, some of the maps have been injective. But there are other interesting examples where this is not the case, and the general definition places no restrictions on the maps involved.

Definition 4.1.19 Let $\mathcal{A}$ be a category, and take a diagram

$$\begin{tikzcd}
 & Z \arrow[d, "g"] \\
Y \arrow[r, "f"'] & X
\end{tikzcd} \tag{4:10}$$

of objects and maps in $\mathcal{A}$. A **pullback** of this diagram is an object $P \in \mathcal{A}$ together with maps $p_1 : P \longrightarrow Y$ and $p_2 : P \longrightarrow Z$ such that

$$\begin{tikzcd}
P \arrow[r, "p_2"] \arrow[d, "p_1"'] & Z \arrow[d, "g"] \\
Y \arrow[r, "f"'] & X
\end{tikzcd} \tag{4:11}$$

commutes, and with the further property that for any commutative square

$$\begin{tikzcd}
A \arrow[r, "q_2"] \arrow[d, "q_1"'] & Z \arrow[d, "g"] \\
Y \arrow[r, "f"'] & X
\end{tikzcd} \tag{4:12}$$

in $\mathcal{A}$, there is a unique map $\overline{q} : A \longrightarrow P$ such that

$$\begin{tikzcd}
A \arrow[rrd, "q_2"] \arrow[rd, "\overline{q}"] \arrow[rdd, "q_1"'] & & \\
& P \arrow[r, "p_2"'] \arrow[d, "p_1"] & Z \arrow[d, "g"] \\
& Y \arrow[r, "f"'] & X
\end{tikzcd}$$

commutes. The maps p_1 and p_2 are called the **projections**.

Again, Remarks 4.1.6 apply. We call (4:11) a **pullback square**.

Examples 4.1.20 a. The squares (4:8) and (4:9) are pullback squares in **Set** (and similarly in **Gp**). In the situation of (4:8), people sometimes say things like 'take a subset X' of X and pull it back along f to get a subset of Y': hence the name.

b. The pullback of a diagram (4:10) in **Set** is

$$P = \{(y, z) \in Y \times Z \mid f(y) = g(z)\}$$

with projections p_1, p_2 given by $p_1(y, z) = y$ and $p_2(y, z) = z$. You can check this now, but soon we'll prove a general formula of which this is a special case. You can also check that this formula is compatible with Example 4.1.16, where g was the inclusion of a subset.

The definition of limit

We've now looked at three general constructions: products, equalizers and pullbacks. They clearly have something in common. Each starts with some objects and (in the case of equalizers and pullbacks) some maps, and tries to construct from them a new object together with some maps from it to the original objects. Each construction is characterized by some universal property.

Let's analyze this more closely. What's the starting data in each construction? For products, it's a pair of objects

$$X, \qquad Y. \tag{4:13}$$

(Let's stick to *binary* products for now.) For equalizers, it's a diagram

$$\begin{tikzcd}
X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y.
\end{tikzcd} \tag{4:14}$$

For pullbacks, it's a diagram

$$\begin{tikzcd}
& Z \arrow[d, "g"] \\
Y \arrow[r, "f"'] & X.
\end{tikzcd} \tag{4:15}$$

We've already seen that 'shapes' or 'figures' in a geometric object can be described by maps into it. For instance, a curve in a topological space A is a map $\mathbb{R} \longrightarrow A$. More to the point, an object of a category $\mathcal{A}$ is a functor $D : \mathbf{1} \longrightarrow \mathcal{A}$; you can think of $\mathbf{1} = \boxed{\bullet}$ as an unlabelled object and D as labelling it with the name of an object of $\mathcal{A}$. Similarly, an arrow of a category $\mathcal{A}$ is a functor $\mathbf{2} \longrightarrow \mathcal{A}$, where $\mathbf{2} = \boxed{\bullet \longrightarrow \bullet}$. (To make this clear: $\mathbf{2}$ is the category with two objects, say 0 and 1, with one map $0 \longrightarrow 1$, and with no other maps except for identities.) And if we take $\mathbb{I}$ to be one of the categories

$$\mathbb{T} = \boxed{\bullet \qquad \bullet}\,, \qquad \mathbb{E} = \boxed{\bullet \rightrightarrows \bullet} \qquad \text{or} \qquad \mathbb{P} = \boxed{\begin{tikzcd} & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd}}$$

then a functor $\mathbb{I} \longrightarrow \mathcal{A}$ consists of data (4:13), (4:14) or (4:15) in $\mathbb{A}$, respectively.

(Recall from Section $2\frac{1}{2}$.2 the convention of using the typeface $\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots$ for *small* categories.)

Definition 4.1.21 Let $\mathcal{A}$ be a category and $\mathbb{I}$ a small category. A functor $\mathbb{I} \longrightarrow \mathcal{A}$ is called a **diagram** in $\mathcal{A}$ of **shape** $\mathbb{I}$.

So (4:13), (4:14) and (4:15) are diagrams of shape $\mathbb{T}$, $\mathbb{E}$ and $\mathbb{P}$.

We already have the definition of product of a diagram of shape $\mathbb{T}$, equalizer of a diagram of shape $\mathbb{E}$, and pullback of a diagram of shape $\mathbb{P}$. Now we unify them in the definition of limit.

Definition 4.1.22 Let $\mathcal{A}$ be a category, $\mathbb{I}$ a small category, and $D : \mathbb{I} \longrightarrow \mathcal{A}$ a diagram in $\mathcal{A}$. A **cone** on D is an object $A \in \mathcal{A}$ (the **vertex** of the cone) together with a family

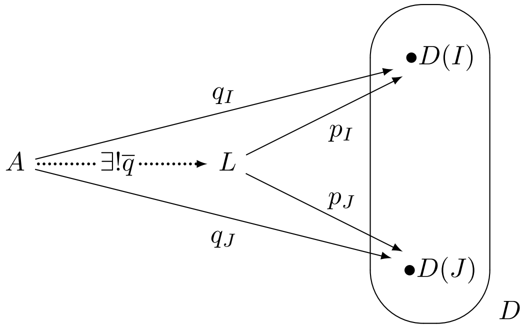
$$\left(A \xrightarrow{q_I} D(I)\right)_{I \in \mathbb{I}} \tag{4:16}$$

of maps in $\mathcal{A}$ such that for all maps $I \xrightarrow{u} J$ in $\mathbb{I}$, the diagram

$$\begin{tikzcd} & D(I) \arrow[dd, "Du"] \\ A \arrow[ur, "q_I"] \arrow[dr, "q_J"'] & \\ & D(J) \end{tikzcd}$$

commutes. A **limit** of D is a cone $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ with the property that for any cone (4:16) on D, there is a unique map $\overline{q} : A \longrightarrow L$ such that $p_I \circ \overline{q} = q_I$ for all $I \in \mathbb{I}$.

Picture:



(Compared to the pictures we drew for products, this picture has been rotated by 90°.)

Remarks 4.1.23 a. The universal property says, loosely, that for any $A \in \mathcal{A}$, maps $A \longrightarrow L$ correspond one-to-one with cones on D with vertex A. (Any map $r : A \longrightarrow L$ gives rise to a cone $(A \xrightarrow{p_I r} D(I))_{I \in \mathbb{I}}$, and the definition of limit is that this process is bijective for each A.) Later we'll use this thought to rephrase the definition of limit in terms of representability. From this it will follow that limits are unique (up to canonical isomorphism), where they exist. You can also prove this directly by the usual kind of argument (0.1.4).

b. If $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ is a limit of D, we sometimes abuse language slightly by referring to L (rather than the whole cone) as the limit of D. For emphasis, we sometimes call $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ a **limit cone**. We write $L = \lim\limits_{\leftarrow \mathbb{I}} D$.

c. We'll stick to **small limits**, that is, limits where the shape category $\mathbb{I}$ is small. It's sometimes worth considering large limits, but we shan't do so in this course. For us, 'limit' will mean small limit.

Examples 4.1.24 Let $\mathcal{A}$ be any category.

a. Let $\mathbb{I} = \mathbb{T}$. A diagram D of shape $\mathbb{T}$ in $\mathcal{A}$ is a pair (X, Y) of objects of $\mathcal{A}$. A cone on D is a cone on (X, Y) in the sense of Definition 4.1.5, and a limit of D is a product of X and Y.

   More generally, let I be a set and let $\mathbb{I}$ be the discrete category on I. As we saw at the beginning of Section 1.3, a functor $D : \mathbb{I} \longrightarrow \mathcal{A}$ is an I-indexed family $(X_i)_{i \in I}$ of objects of $\mathcal{A}$. A limit of D is exactly a product of the family $(X_i)_{i \in I}$. In particular, a limit of the unique functor $\emptyset \longrightarrow \mathcal{A}$ is a terminal object of $\mathcal{A}$.

b. Let $\mathbb{I} = \mathbb{E}$. A diagram D of shape $\mathbb{E}$ in $\mathcal{A}$ is a parallel pair $X \underset{g}{\overset{f}{\rightrightarrows}} Y$ of maps in $\mathcal{A}$. A cone on D consists of objects and maps

$$\begin{tikzcd}
& A \arrow[dl, "q"'] \arrow[dr, "r"] & \\
X \arrow[rr, shift left, "f"] \arrow[rr, shift right, "g"'] & & Y
\end{tikzcd}$$

such that $f \circ q = r$ and $g \circ q = r$. But since r is determined by q, it's equivalent to say that a cone on D consists of a map $A \xrightarrow{q} X$ into X such that

$$\begin{tikzcd}
A \arrow[r, "q"] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y
\end{tikzcd}$$

is a fork. A limit of D is a 'universal fork', that is, an equalizer of f and g.

c. Let $\mathbb{I} = \mathbb{P}$. A diagram D of shape $\mathbb{P}$ in $\mathcal{A}$ is a 'corner'

$$\begin{tikzcd}
& Z \arrow[d, "g"] \\
Y \arrow[r, "f"'] & X
\end{tikzcd}$$

in $\mathcal{A}$. A cone on D is a commutative square (4:12). (To see this, you have to perform a simplification similar to that in the last example.) A limit of D is a pullback.

A limit of D is a terminal object in the category of cones on D, and is therefore an 'extremal' example of a cone on D. The word 'limit' can be understood in the sense of 'on the boundary', rather than indicating the kind of limiting process common in analysis.

So far we haven't said much about which limits exist, except to observe (in 4.1.6(a)) that they do not exist always. We'll now show that in many familiar categories, all limits do exist—indeed, we can construct them explicitly.

Examples 4.1.25 a. Let $D : \mathbb{I} \longrightarrow \mathbf{Set}$ and, as a kind of thought experiment, let's ask ourselves what $\lim\limits_{\longleftarrow \mathbb{I}} D$ would have to be if it existed. (We don't yet know that it does.) We'd have

$$\begin{aligned}
\lim_{\longleftarrow \mathbb{I}} D &\cong \mathbf{Set}(1, \lim_{\longleftarrow \mathbb{I}} D) \\
&\cong \{\text{cones on } D \text{ with vertex } 1\}
\end{aligned}$$

$$\cong \quad \{(x_I)_{I\in\mathbb{I}} \mid x_I \in D(I) \text{ for all } I \in \mathbb{I} \text{ and } (Du)(x_I) = x_J \text{ for all } I \xrightarrow{u} J \text{ in } \mathbb{I}\} \quad (4{:}17)$$

where the second isomorphism is by definition of limit and the third is by definition of cone. In fact, (4:17) really *is* the limit of D in **Set**, with projection $p_J : \lim\limits_{\longleftarrow \mathbb{I}} D \longrightarrow D(J)$ (for $J \in \mathbb{I}$) given by $p_J((x_I)_{I\in\mathbb{I}}) = x_J$. (Proof: exercise.) So in **Set**, all limits exist.

b. The same formula gives limits in categories of algebras such as **Gp**, **Ring**, $\mathbf{Vect}_k$, ... Of course, we also have to say what the group/ring/... structure on the set (4:17) is, but this works in the most straightforward way. For instance, in $\mathbf{Vect}_k$, if $(x_I)_{I\in\mathbb{I}}, (y_I)_{I\in\mathbb{I}} \in \lim\limits_{\leftarrow\mathbb{I}} D$ then

$$(x_I)_{I\in\mathbb{I}} + (y_I)_{I\in\mathbb{I}} = (x_I + y_I)_{I\in\mathbb{I}}.$$

c. The same formula also gives limits in **Top**. The topology on the set (4:17) is the smallest for which the projections are continuous.

Definition 4.1.26 a. Let $\mathbb{I}$ be a small category. A category $\mathcal{A}$ **has limits of shape** $\mathbb{I}$ if for every diagram D of shape $\mathbb{I}$ in $\mathcal{A}$, a limit of D exists.

b. A category **has all limits** (or properly, **has small limits**) if it has limits of shape $\mathbb{I}$ for all small categories $\mathbb{I}$.

We've just shown that **Set**, **Top**, **Gp**, **Ring**, $\mathbf{Vect}_k$, ... all have all limits.

Similar terminology can be applied to special classes of limits (e.g. 'has pullbacks'). The class of finite limits is particularly important. By definition, a category is **finite** if it contains only finitely many maps (in which case it also contains only finitely many objects). A **finite limit** is a limit of shape $\mathbb{I}$ for some finite $\mathbb{I}$. For instance, binary products, terminal objects, equalizers and pullbacks are all finite limits.

The next result shows that any limit can be built up out of certain basic limits.

Proposition 4.1.27 *Let $\mathcal{A}$ be a category.*

a. If $\mathcal{A}$ has all products and equalizers then $\mathcal{A}$ has all limits.

b. If $\mathcal{A}$ has binary products, a terminal object and equalizers then $\mathcal{A}$ has finite limits.

The idea of the proof can be seen in formula (4:17) for limits in **Set**. There, the limit of a diagram D is described as the subset of the product $\prod_I D(I)$ on which certain equations hold. If it were just one equation, the limit could therefore be expressed as an equalizer of some parallel pair $\prod D(I) \rightrightarrows$ (something). Since it's more than one, things get slightly more complicated.

Proof

a. Not hard, but omitted (and non-examinable). See for instance *Categories for the Working Mathematician* or the lecture notes of Cheng.

b. By induction, any finite product $X_1 \times \cdots \times X_n$ $(n \geq 0)$ can be built from the terminal object and binary products. So we have finite products and equalizers. Then repeat the argument of (4.1.27(a)), inserting the word 'finite' in appropriate places. □

Examples 4.1.28 a. The Proposition provides a neat way of showing that the category of compact Hausdorff spaces has all limits.

b. Finite limits in $\mathbf{Vect}_k$ or **Ab** can all be built from $\oplus$ (binary direct sum), $\{0\}$, and ker.

Monics

It's often useful to talk about injective maps. However, injectivity of a map doesn't make sense in an arbitrary category. Here's a substitute that does.

Definition 4.1.29 Let $\mathcal{A}$ be a category. A map $X \xrightarrow{f} Y$ in $\mathcal{A}$ is **monic** (or a **monomorphism**) if for all objects A and maps $A \underset{x'}{\overset{x}{\rightrightarrows}} X$,

$$f \circ x = f \circ x' \Rightarrow x = x'.$$

In lectures I explained the idea of 'generalized elements'. A **generalized element** of an object X of a category is simply a map into X. (In the category of sets, the maps $1 \longrightarrow X$ are the *ordinary* elements of X.) The definition of f being monic reads: 'for all generalized elements x and x' of X, if f of x equals f of x' then x equals x''. So it's the generalized-element analogue of injectivity.

Examples 4.1.30 a. In **Set**, a map is monic if and only if it is injective. For if f is injective then certainly f is monic; to see the converse, take $A = 1$.

b. In categories of algebras such as **Gp**, $\mathbf{Vect}_k$, etc., it is also true that monic $\iff$ injective. Again, it's easy to show '$\Leftarrow$'. For '$\Rightarrow$', take $A = F(1)$, where F is the free functor (2.1.3).

Why is the definition of monic in a chapter on limits? Because of this:

Lemma 4.1.31 *A map* $X \xrightarrow{f} Y$ *is monic if and only if*

$$\begin{tikzcd}
X \arrow[r, "1"] \arrow[d, "1"'] & X \arrow[d, "f"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

is a pullback square.

Proof Harmless exercise. □

The significance of this lemma is that whenever we prove a result about limits, a result about monics will follow. For example, we will soon show that the forgetful functors from **Gp**, $\mathbf{Vect}_k$ etc. to **Set** 'preserve limits', from which it follows immediately that they also 'preserve monics', and this in turn gives an alternative proof of the less easy half of 4.1.30(b).

4.2 Colimits: basics

Now that we have seen how examples of limits occur throughout mathematics, it makes sense to look at the dual concept—colimits—and ask whether it is similarly ubiquitous. There's no need to build up coyly to the definition by a sequence of examples, as we did for limits: we can write it down immediately. So the structure of this section will be something like a backwards version of the previous section. We first define colimit, then define sum, coequalizer and pushout (dual to product, equalizer and pullback), giving examples in particular categories.

Dual concepts are often named by adding or subtracting the prefix 'co'. A different convention is to use 'left' and 'right', as in left/right adjoint. Sometimes the names are just irregular, as in initial/terminal object and pullback/pushout.

Definition 4.2.1 Let $\mathcal{A}$ be a category, $\mathbb{I}$ a small category, and $D : \mathbb{I} \longrightarrow \mathcal{A}$ a diagram in $\mathcal{A}$. A **cocone** on D is an object $A \in \mathcal{A}$ (the **vertex** of the cone) together with a family

$$\left(D(I) \xrightarrow{q_I} A\right)_{I \in \mathbb{I}} \tag{4:18}$$

of maps in $\mathcal{A}$ such that for all maps $I \xrightarrow{u} J$ in $\mathbb{I}$, the diagram

$$\begin{tikzcd}
D(I) \arrow[dd, "Du"'] \arrow[dr, "q_I"] & \\
 & A \\
D(J) \arrow[ur, "q_J"'] &
\end{tikzcd}$$

commutes. A **colimit** of D is a cocone $(D(I) \xrightarrow{p_I} C)_{I \in \mathbb{I}}$ with the property that for any cocone (4:18) on D, there is a unique map $\overline{q} : C \longrightarrow A$ such that $\overline{q} \circ p_I = q_I$ for all $I \in \mathbb{I}$.

The associated picture is the mirror image of the picture on page 76.

Of course, Remarks 4.1.23 apply equally here. We write (the vertex of) the colimit as $\varinjlim_{\mathbb{I}} D$, and call the maps p_I the **coprojections**.

Sums

Definition 4.2.2 A **sum** or **coproduct** is a colimit over a discrete category. (That is, it is a colimit of shape $\mathbb{I}$ for some discrete category $\mathbb{I}$.)

If $(X_i)_{i \in I}$ is a family of objects of a category then their sum (if it exists) is written $\sum_{i \in I} X_i$ or $\coprod_{i \in I} X_i$. When I is a finite set $\{1, \ldots, n\}$, we write $\sum_{i \in I} X_i$ as $X_1 + \cdots + X_n$, or as 0 if $n = 0$.

Examples 4.2.3 a. **Initial objects** By the dual of 4.1.9(b), a sum of the empty family is exactly an initial object. So the convention is to write initial objects as 0.

b. **Sets** Sums in **Set** are as described in Section $2\frac{1}{2}.1$. Let's look in detail at what the universal property says in the case of binary sums. Take two sets, X_1 and X_2. Form their sum, $X_1 + X_2$, and consider the inclusions

$$\begin{tikzcd} X_1 \arrow[r, "p_1"] & X_1 + X_2 & X_2. \arrow[l, "p_2"'] \end{tikzcd}$$

This is a colimit cocone. To prove this, we have to prove the following universal property: for any diagram

$$\begin{tikzcd} X_1 \arrow[r, "q_1"] & A & X_2 \arrow[l, "q_2"'] \end{tikzcd}$$

of sets and functions, there is a unique function $\overline{q} : X_1 + X_2 \longrightarrow A$ making

$$\begin{tikzcd} X_1 \arrow[dr, "p_1"'] \arrow[drr, "q_1"] & & \\ & X_1 + X_2 \arrow[r, dotted, "\overline{q}"] & A \\ X_2 \arrow[ur, "p_2"] \arrow[urr, "q_2"'] & & \end{tikzcd}$$

commute. As noted previously, p_1 and p_2 are injections and their images partition $X_1 + X_2$; in other words, every element x of $X_1 + X_2$ is *either* equal to $p_1(x_1)$ for some $x_1 \in X_1$ (and this x_1 is then unique), *or* equal to $p_2(x_2)$ for some $x_2 \in X_2$ (and this x_2 is then unique), but not both. So we may define $\overline{q}(x)$ to be equal to $q_1(x_1)$ in the first case, or $q_2(x_2)$ in the second. This defines a function $\overline{q}$ making the diagram commute, and it is clearly the unique such function.

c. **Vector spaces** Let X_1 and X_2 be vector spaces. There are linear maps

$$\begin{tikzcd} X_1 \arrow[r, "i_1"] & X_1 \oplus X_2 & X_2 \arrow[l, "i_2"'] \end{tikzcd} \tag{4:19}$$

defined by $i_1(x_1) = (x_1, 0)$ and $i_2(x_2) = (0, x_2)$ ($x_1 \in X_1$, $x_2 \in X_2$), and (4:19) is a colimit cocone in $\mathbf{Vect}_k$. Hence binary direct sum is sum in the categorical sense. This is surprising: we saw in 4.1.3 that $X_1 \oplus X_2$ is also the *product* of X_1 and X_2! Contrast this with the category of sets (or almost any other category), where sum and product are completely different.

d. **Elements of ordered sets** Let $(A, \leq)$ be an ordered set. **Upper bounds** and **least upper bounds** (**joins**) in A are defined by dualizing the definitions in Example 4.1.4. The join of a family $(x_i)_{i\in I}$ is written $\bigvee_{i\in I} x_i$. In the binary case (where I has two elements), the join of x_1 and x_2 is written $x_1 \vee x_2$. A join of the empty family (when $I = \emptyset$) is an initial object of the category A, as in (a). Equivalently, it is a **least element** of A: an element $0 \in A$ such that for all $a \in A$ we have $0 \leq a$.

For instance, in $(\mathbb{R}, \leq)$, join is max or (in the infinite case) sup, and there is no least element. In a powerset $(\mathcal{P}(S), \subseteq)$, join is union and the least element is $\emptyset$. In $(\mathbb{N}, |)$, join is lowest common multiple and the least element is 1 (since 1 divides everything).

Coequalizers

We continue to write $\mathbb{E}$ for the category $\boxed{\bullet \rightrightarrows \bullet}$.

Definition 4.2.4 A **coequalizer** is a colimit of shape $\mathbb{E}$.

In other words, given a diagram $X \overset{f}{\underset{g}{\rightrightarrows}} Y$, a coequalizer of f and g is a map $Y \xrightarrow{e} C$ such that

$$\begin{tikzcd} X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[r, "e"] & C \end{tikzcd}$$

is a **cofork** ($e \circ f = e \circ g$) and is universal with this property.

Coequalizers are something like quotients, as shown by the examples below. But first:

Digression 4.2.5 A binary relation R on a set A can be regarded as a subset $R \subseteq A \times A$. Think of $(a, a') \in R$ as meaning 'a and a' are related'. It makes sense to speak of one relation S on A 'containing' another such relation, R. This means that $R \subseteq S$: whenever a and a' are R-related, they are also S-related.

For any binary relation R on a set A, there is a smallest equivalence relation $\sim$ containing R. This is called the equivalence relation **generated** by R. 'Smallest' means that any equivalence relation containing R also contains $\sim$. We can construct $\sim$ as the intersection of all equivalence relations on A containing R, using the fact that the intersection of any family of equivalence relations is again an equivalence relation. There is also an explicit construction. The rough idea is as follows: writing $x \to y$ to mean $(x, y) \in R$, we should have $a \sim a'$ if and only if there is a zigzag like

$$a \to b \leftarrow c \leftarrow d \to e \leftarrow a'$$

between a and a'. Formally, we first define a relation S on A by

$$S = \{(a, a') \in A \times A \mid (a, a') \in R \text{ or } (a', a) \in R\}$$

(which enlarges R to a symmetric relation), then define $\sim$ by declaring that $a \sim a'$ if and only if there exist $n \geq 0$ and $a_0, \ldots, a_n \in A$ such that

$$a = a_0,\ (a_0, a_1) \in S,\ (a_1, a_2) \in S,\ \ldots,\ (a_{n-1}, a_n) \in S,\ a_n = a'$$

(which also forces reflexivity and transitivity).

In Section $2\frac{1}{2}$.1 we recalled that for any equivalence relation $\sim$ on a set A, we may construct the set $A/{\sim}$ of equivalence classes and the quotient map

$p : A \longrightarrow A/{\sim}$, which is surjective and has the property that $p(a) = p(a') \iff a \sim a'$, for $a, a' \in A$. For any set B, the maps $A/{\sim} \longrightarrow B$ correspond one-to-one (via composition with p) with the maps $q : A \longrightarrow B$ such that

$$\forall a, a' \in A, \qquad a \sim a' \Rightarrow q(a) = q(a'). \tag{4:20}$$

If $\sim$ is the equivalence relation generated by some relation R, this condition is equivalent to

$$\forall a, a' \in A, \qquad (a, a') \in R \Rightarrow q(a) = q(a'). \tag{4:21}$$

(Proof: there is an equivalence relation $\approx$ on A defined by $a \approx a' \iff q(a) = q(a')$. Condition (4:20) says that $\sim \subseteq \approx$, and condition (4:21) that $R \subseteq \approx$. But by definition of 'generates', these statements are equivalent.) So for any set B, the maps $A/{\sim} \longrightarrow B$ correspond one-to-one with the maps $q : A \longrightarrow B$ satisfying (4:21).

Examples 4.2.6 a. **Sets** Take sets and functions $X \overset{f}{\underset{g}{\rightrightarrows}} Y$. To find the coequalizer of f and g, we must construct in some kind of canonical way a set C and a function $e : Y \longrightarrow C$ such that $e(f(x)) = e(g(x))$ for all $x \in X$. So, let $\sim$ be the equivalence relation on Y generated by $f(x) \sim g(x)$ for all $x \in X$ (that is, generated by the relation $R = \{(f(x), g(x)) \mid x \in X\}$). Take the quotient map $e : Y \longrightarrow Y/{\sim}$. By the correspondence discussed in the Digression, this is indeed the coequalizer of f and g.

b. **Abelian groups** Given $A \overset{f}{\underset{g}{\rightrightarrows}} B$ in **Ab**, there is a homomorphism $g - f : A \longrightarrow B$, hence a subgroup $\mathrm{im}(g - f)$ of B. The coequalizer of f and g is the canonical homomorphism $B \longrightarrow B/\mathrm{im}(g - f)$.

Pushouts

Definition 4.2.7 A **pushout** is a colimit of shape

$$\mathbb{P}^{\mathrm{op}} = \boxed{\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \\ \bullet & \end{tikzcd}}\,.$$

In other words, the pushout of a diagram

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "g"'] & Y \\ Z & \end{tikzcd} \tag{4:22}$$

(if it exists) is a commutative square

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d] \\
Z \arrow[r] & \cdot
\end{tikzcd}$$

that is 'universal as such'.

Examples 4.2.8 a. **Sets** Take a diagram (4:22) in **Set**. Its pushout P is $(Y+Z)/\sim$, where $\sim$ is the equivalence relation on $Y+Z$ generated by $f(x) \sim g(x)$ for all $x \in X$. The coprojection $Y \longrightarrow P$ sends $y \in Y$ to its equivalence class in P, and the coprojection $Z \longrightarrow P$ is described similarly.

For example, if Y and Z are subsets of a set A then

$$\begin{tikzcd}
Y \cap Z \arrow[r, hook] \arrow[d, hook] & Y \arrow[d, hook] \\
Z \arrow[r, hook] & Y \cup Z
\end{tikzcd}$$

is a pushout square in **Set**. (It's also a pullback square!) You can check this by verifying the universal property or by using the formula just stated. In this case, the formula takes the two sets Y and Z, places them side by side (giving $Y+Z$), then glues the subset $Y \cap Z$ of Y to the subset $Y \cap Z$ of Z (giving $(Y+Z)/\sim \, = Y \cup Z$).

b. **Sums as pushouts** If $\mathcal{A}$ is a category with an initial object 0, and if $Y, Z \in \mathcal{A}$, then a pushout of the unique diagram

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & Y \\
Z &
\end{tikzcd}$$

is exactly a sum of Y and Z.

c. **Spaces and groups** The van Kampen Theorem (Example 0.1.6) says that given a pushout square in **Top** satisfying certain further hypotheses, the square in **Gp** obtained by taking fundamental groups is also a pushout.

With all these examples in mind, we now write down a general formula for colimits in **Set**.

Example 4.2.9 (Colimits in Set) The colimit of a diagram $D : \mathbb{I} \longrightarrow$ **Set** is given by

$$\lim_{\to \mathbb{I}} D = \left(\sum_{I \in \mathbb{I}} D(I) \right) /\sim$$

where $\sim$ is the equivalence relation on $\sum D(I)$ generated by

$$x \sim (Du)(x)$$

for all $I \xrightarrow{u} J$ in $\mathbb{I}$ and $x \in D(I)$. To see this, note that for any set A, maps $(\sum D(I))/\sim \longrightarrow A$ correspond to maps $q : \sum D(I) \longrightarrow A$ such that $q(x) = q((Du)(x))$ for all u and x as above, which in turn correspond to families of maps $(D(I) \xrightarrow{q_I} A)_{I \in \mathbb{I}}$ such that $q_I(x) = q_J((Du)(x))$ for all u and x as above—but these are exactly the cocones on D with vertex A.

There is a kind of duality between the formulas for limits in **Set** (4:17) and colimits in **Set**. The limit is constructed as a *subset* of a *product*, and the colimit as a *quotient* of a *sum*.

Epics

Definition 4.2.10 Let $\mathcal{A}$ be a category. A map $X \xrightarrow{f} Y$ in $\mathcal{A}$ is **epic** (or an **epimorphism**) if for all objects Z and maps $Y \underset{g'}{\overset{g}{\rightrightarrows}} Z$,

$$g \circ f = g' \circ f \;\Rightarrow\; g = g'.$$

This is the formal dual of the definition of monic. It is in some sense the categorical version of surjectivity. However, the definition of epic looks much less like the definition of surjective than the definition of monic looks like the definition of injective. The following examples confirm that in categories where 'surjective' makes sense, it only sometimes coincides with 'epic'.

Examples 4.2.11 a. In **Set**, a map is epic if and only if it is surjective. If f is surjective then certainly f is epic. To see the converse, take $Z = \{\mathtt{true}, \mathtt{false}\}$, take g to be χ_S (as defined in in Section $2\frac{1}{2}.1$) where S is the image of f, and take g' to be the function with constant value `true`.

Any isomorphism in any category is both monic and epic. In **Set** the converse also holds, since any injective surjective function is invertible.

b. In categories of algebras, any surjective map is certainly epic. In some such categories (including **Ab** and **Vect**$_k$), the converse also holds. However, there are other categories of algebras where it fails: for instance, in **Ring**, the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is epic but not surjective. (Proof: exercise.) This is also an example of a map that is monic and epic but not an isomorphism.

c. In the category of Hausdorff spaces and continuous maps, any map with dense image is epic.

Of course, there is a dual of Lemma 4.1.31, saying that a map is epic if and only if a certain square is a pushout.

4.3 Interaction of (co)limits with functors

We saw in 4.1.25(b) that limits in categories such as **Gp**, **Ring** and $\mathbf{Vect}_k$ can be computed by first taking the limit in the category of sets, then equipping the result with a suitable algebraic structure. On the other hand, colimits in these categories look unlike colimits in **Set**. For example, the underlying set of the initial object of **Gp** (which has one element) is not the initial object of **Set** (which has no elements), and the underlying set of the direct sum $X \oplus Y$ of two vector spaces is not the sum of the underlying sets of X and Y. So these forgetful functors interact well with limits and badly with colimits. In this section we develop terminology that will, among other things, enable us to express these thoughts precisely.

Definition 4.3.1 a. A functor $F : \mathcal{A} \longrightarrow \mathcal{B}$ **preserves limits (of shape $\mathbb{I}$)** if for all diagrams $D : \mathbb{I} \longrightarrow \mathcal{A}$ and cones $(A \xrightarrow{q_I} D(I))_{I\in\mathbb{I}}$ on D,

$$\begin{aligned} &\Big(A \xrightarrow{q_I} D(I)\Big)_{I\in\mathbb{I}} \text{ is a limit cone on } D \text{ in } \mathcal{A} \\ \Rightarrow\quad &\Big(F(A) \xrightarrow{Fq_I} FD(I)\Big)_{I\in\mathbb{I}} \text{ is a limit cone on } F \circ D \text{ in } \mathcal{B}. \end{aligned}$$

b. **Reflection** of limits is defined as in (a), but with $\Leftarrow$ in place of $\Rightarrow$.

Of course, the same terminology can be applied to colimits.

Examples 4.3.2 a. The forgetful functor $U : \mathbf{Top} \longrightarrow \mathbf{Set}$ preserves both limits and colimits. (As we will see, this follows from the fact that U has adjoints on both sides.) It doesn't reflect all limits or all colimits. For instance, take any non-discrete spaces X and Y and write $X\Box Y$ for the set $U(X) \times U(Y)$ equipped with the discrete topology. Then we have a cone

$$X \longleftarrow X\Box Y \longrightarrow Y \tag{4:23}$$

in **Top** whose image in **Set** is the product cone

$$U(X) \longleftarrow U(X) \times U(Y) \longrightarrow U(Y).$$

But (4:23) is not a product cone in **Top**, since the discrete topology on $U(X) \times U(Y)$ is not the product topology.

b. In the first paragraph of this section, we observed that the forgetful functor $U : \mathbf{Gp} \longrightarrow \mathbf{Set}$ does not preserve initial objects and that the forgetful functor $\mathbf{Vect}_k \longrightarrow \mathbf{Set}$ does not preserve binary sums. Forgetful functors on categories of algebras very seldom preserve all colimits.

c. We also saw that forgetful functors on categories of algebras do preserve limits. In fact, something stronger is true. Let's examine the case of binary products in **Gp**, although all of the following can be said for any limits in any of the categories **Gp**, **Ab**, $\mathbf{Vect}_k$, **Ring**, etc.

Take groups X_1 and X_2. We can form the product set $U(X_1) \times U(X_2)$, which comes equipped with projections

$$\begin{tikzcd} U(X_1) & U(X_1) \times U(X_2) \arrow[l, "p_1"'] \arrow[r, "p_2"] & U(X_2). \end{tikzcd}$$

I claim that there is exactly one group structure on the set $U(X_1) \times U(X_2)$ with the property that p_1 and p_2 are homomorphisms. To prove uniqueness, suppose that we have a group structure on $U(X_1) \times U(X_2)$ with this property. Take elements (x_1, x_2) and (x'_1, x'_2) of $U(X_1) \times U(X_2)$ and write $(x_1, x_2) \cdot (x'_1, x'_2) = (y_1, y_2)$. Since p_1 is a homomorphism,

$$y_1 = p_1(y_1, y_2) = p_1((x_1, x_2) \cdot (x'_1, x'_2)) = p_1(x_1, x_2) \cdot p_1(x'_1, x'_2) = x_1 \cdot x'_1,$$

and similarly $y_2 = x_2 \cdot x'_2$. Hence $(x_1, x_2) \cdot (x'_1, x'_2) = (x_1x'_1, x_2x'_2)$. A similar argument shows that $(x_1, x_2)^{-1} = (x_1^{-1}, x_2^{-1})$ and that the identity element 1 of the group is $(1, 1)$. For existence, define $\cdot$, $(\ )^{-1}$ and 1 by the formulas just given; it can then be checked that the group axioms are satisfied and that p_1 and p_2 are group homomorphisms. This proves the claim.

Write L for the set $U(X_1) \times U(X_2)$ equipped with this group structure. Then we have a cone

$$\begin{tikzcd} X_1 & L \arrow[l, "p_1"'] \arrow[r, "p_2"] & X_2 \end{tikzcd}$$

in **Gp**, where p_1 and p_2 are the evident projections. It's easy to check that this is, in fact, a *product* cone in **Gp**. (Compare 4.1.3.)

We can summarize this in language that isn't tied to group theory:

- any product cone on $(U(X_1), U(X_2))$ in **Set** lifts uniquely to a cone on (X_1, X_2) in **Gp** (where 'lifts' means that the image under U of the cone in **Gp** is the given cone in **Set**), and
- this cone on (X_1, X_2) is a product cone.

Exercise 4.3.3 State and prove analogous statements for arbitrary limits in **Gp**, using the formula (4:17) for limits in **Set**. Satisfy yourself that the same could be done in other categories of algebras.

Definition 4.3.4 A functor $F : \mathcal{A} \longrightarrow \mathcal{B}$ **strictly creates limits (of shape $\mathbb{I}$)** if whenever $D : \mathbb{I} \longrightarrow \mathcal{A}$ is a diagram in $\mathcal{A}$,

- for any limit cone $\left(B \xrightarrow{q_I} FD(I)\right)_{I \in \mathbb{I}}$ on $F \circ D$, there is a unique cone $\left(A \xrightarrow{p_I} D(I)\right)_{I \in \mathbb{I}}$ on D such that $F(A) = B$ and $F(p_I) = q_I$ for all $I \in \mathbb{I}$, and
- this cone $\left(A \xrightarrow{p_I} D(I)\right)_{I \in \mathbb{I}}$ is a limit cone on D.

So the forgetful functors from **Gp**, **Ring**, ... to **Set** all strictly create limits.

Exercise 4.3.5 Prove that if $F : \mathcal{A} \longrightarrow \mathcal{B}$ strictly creates limits of shape $\mathbb{I}$ and $\mathcal{B}$ has them then $\mathcal{A}$ has them and F preserves them.

Since **Set** has all limits, it follows that all our categories of algebras have all limits and that the forgetful functors preserve them.

Remark 4.3.6 There's something fishy about Definition 4.3.4: it refers to equality of objects of categories, a notion that, as we saw on page 12, is usually too strict to be appropriate. It's almost always better to replace it with isomorphism. If we replace equality with isomorphism throughout the definition of 'creates limits strictly', we obtain the more healthy and inclusive notion, 'creates limits'. Although this more inclusive notion is really the right one, it's slightly more complicated, and in this course we'll be able to get away with using just the strict notion.

4.4 The definition of (co)limit, revisited

There's more than one way to present the definition of limit. So far we've been using a form of the definition that's particularly convenient for examples. But we'll soon be developing the *theory* of limits and colimits, and for that it will be useful to rephrase the definition. In fact, we rephrase it in two different ways: once in terms of representability, and once in terms of adjoints.

Notation 4.4.1 Given categories $\mathbb{I}$ and $\mathcal{A}$ and an object $A \in \mathcal{A}$, there is a functor $\Delta A : \mathbb{I} \longrightarrow \mathcal{A}$ with constant value A on objects and 1_A on arrows. This defines, for each $\mathbb{I}$ and $\mathcal{A}$, the **diagonal functor** $\Delta : \mathcal{A} \longrightarrow [\mathbb{I}, \mathcal{A}]$.

Here is the first rephrasing.

Proposition 4.4.2 *Let $\mathbb{I}$ be a small category, $\mathcal{A}$ a category, and $D : \mathbb{I} \longrightarrow \mathcal{A}$ a diagram. Then a limit of D is the same thing as a representation of the functor*

$$\begin{array}{rccc} [\mathbb{I}, \mathcal{A}](\Delta -, D) : & \mathcal{A}^{\mathrm{op}} & \longrightarrow & \mathbf{Set} \\ & A & \longmapsto & [\mathbb{I}, \mathcal{A}](\Delta A, D). \end{array}$$

Proof Given $A \in \mathcal{A}$, a natural transformation $\Delta A \longrightarrow D$ is exactly a cone on D with vertex A. So by Corollary 3.3.2, a representation of $[\mathbb{I}, \mathcal{A}](\Delta -, D)$ consists of a cone on D with a certain universal property, and this is exactly the universal property in the definition of limit cone. □

From Corollary 3.3.9 we deduce:

Corollary 4.4.3 *Limits are unique up to isomorphism.* □

To reach the second rephrasing, we have to prove a basic lemma. So far we've considered limits on individual diagrams D. What happens if we vary the diagram? In other words, given a map $D \longrightarrow D'$ between diagrams, is there an induced map between the limits of D and D'? The answer is yes:

Lemma 4.4.4 *Let $\mathbb{I}$ be a small category and* $\begin{tikzcd} \mathbb{I} \arrow[r, bend left=50, "D"{name=U}] \arrow[r, bend right=50, "D'"'{name=L}] & \mathcal{A} \arrow[Rightarrow, from=U, to=L, "\alpha"] \end{tikzcd}$ *a natural transformation. Let* $(\lim\limits_{\leftarrow \mathbb{I}} D \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ *and* $(\lim\limits_{\leftarrow \mathbb{I}} D' \xrightarrow{p'_I} D'(I))_{I \in \mathbb{I}}$ *be limit cones. Then:*

a. *There is a unique map* $\lim\limits_{\leftarrow \mathbb{I}} \alpha : \lim\limits_{\leftarrow \mathbb{I}} D \longrightarrow \lim\limits_{\leftarrow \mathbb{I}} D'$ *such that*

$$\begin{tikzcd}
\lim\limits_{\leftarrow \mathbb{I}} D \arrow[r, "p_I"] \arrow[d, "\lim\limits_{\leftarrow \mathbb{I}} \alpha"'] & D(I) \arrow[d, "\alpha_I"] \\
\lim\limits_{\leftarrow \mathbb{I}} D' \arrow[r, "p'_I"'] & D'(I)
\end{tikzcd}$$

commutes for all $I \in \mathbb{I}$.

b. *If* $(A \xrightarrow{q_I} D(I))_{I \in \mathbb{I}}$ *and* $(A' \xrightarrow{q'_I} D'(I))_{I \in \mathbb{I}}$ *are cones and $f : A \longrightarrow A'$ is a map such that*

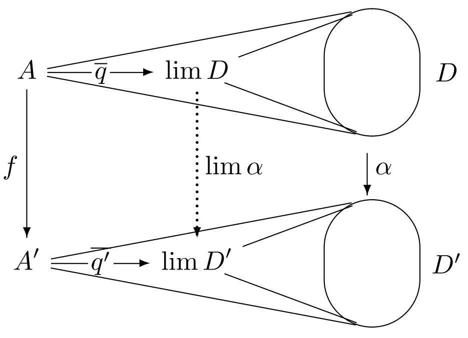
$$\begin{tikzcd}
A \arrow[r, "q_I"] \arrow[d, "f"'] & D(I) \arrow[d, "\alpha_I"] \\
A' \arrow[r, "q'_I"'] & D'(I)
\end{tikzcd}$$

commutes for all $I \in \mathbb{I}$, then

$$\begin{tikzcd}
A \arrow[r, "\overline{q}"] \arrow[d, "f"'] & \lim\limits_{\leftarrow \mathbb{I}} D \arrow[d, "\lim\limits_{\leftarrow \mathbb{I}} \alpha"] \\
A' \arrow[r, "\overline{q'}"'] & \lim\limits_{\leftarrow \mathbb{I}} D'
\end{tikzcd}$$

also commutes.

Picture:



Proof

a. This follows immediately from the fact that $(\lim\limits_{\longleftarrow \mathbb{I}} D \xrightarrow{\alpha_I p_I} D'(I))_{I\in\mathbb{I}}$ is a cone on D'.

b. For each $I \in \mathbb{I}$ we have

$$p'_I \circ (\lim_{\longleftarrow \mathbb{I}} \alpha) \circ \overline{q} = \alpha_I \circ p_I \circ \overline{q} = \alpha_I \circ q_I = q'_I \circ f = p'_I \circ \overline{q'} \circ f,$$

so $(\lim\limits_{\longleftarrow \mathbb{I}} \alpha) \circ \overline{q} = \overline{q'} \circ f$ by the uniqueness part of the definition of limit. □

We can now give the second rephrasing of the notion of limit. It only applies when the category has *all* limits of a given shape.

Proposition 4.4.5 *Let $\mathbb{I}$ be a small category and $\mathcal{A}$ a category with all limits of shape $\mathbb{I}$. Then $\lim\limits_{\longleftarrow \mathbb{I}}$ defines a functor $[\mathbb{I}, \mathcal{A}] \longrightarrow \mathcal{A}$, and this functor is right adjoint to the diagonal functor.*

(Compare Sheet 4, q.3, and sheet 7, q.2.)

Proof Choose for each $D \in [\mathbb{I}, \mathcal{A}]$ a limit cone on D, and call its vertex $\lim\limits_{\longleftarrow \mathbb{I}} D$. For each map $\alpha : D \longrightarrow D'$ in $[\mathbb{I}, \mathcal{A}]$, we have a map $\lim\limits_{\longleftarrow \mathbb{I}} \alpha : \lim\limits_{\longleftarrow \mathbb{I}} D \longrightarrow \lim\limits_{\longleftarrow \mathbb{I}} D'$ defined as in Lemma 4.4.4(a). This makes $\lim\limits_{\longleftarrow \mathbb{I}}$ into a functor. Proposition 4.4.2 implies that

$$[\mathbb{I}, \mathcal{A}](\Delta A, D) \cong \mathcal{A}(A, \lim_{\longleftarrow \mathbb{I}} D)$$

naturally in $A \in \mathcal{A}$, and taking $f = 1_A$ in Lemma 4.4.4(b) tells us that the isomorphism is also natural in D. □

To define the functor $\lim\limits_{\longleftarrow \mathbb{I}}$ we had to *choose* for each D a limit cone on D. This is a non-canonical choice. Nevertheless, different choices only affect the functor $\lim\limits_{\longleftarrow \mathbb{I}}$ up to natural isomorphism, by uniqueness of limits or of adjoints.

We have now made a start on the subject of the final chapter: bringing together the notions of adjoint, representability and limit.

Chapter 5

Adjoints, Representables and Limits

We have approached the idea of universal property in three different ways, and so far we have studied the different approaches more or less in isolation. In this final chapter, we look at the relationships between them.

Along the way, we'll discover many fundamental results of category theory. Here are some of the highlights.

- Limits and colimits of presheaves work in the simplest possible way.
- The embedding of a category $\mathbb{A}$ into its presheaf category $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ preserves limits (but not colimits).
- The representables are the 'prime numbers' of presheaves: every presheaf can be expressed as a colimit of representables.
- Roughly, a functor has a left adjoint if and only if it preserves limits.
- Categories of presheaves $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ behave very much like the category of sets—the beginning of an incredible story that unifies logic with geometry.

In terms of exam questions, this chapter is a gold mine.

5.1 Limits and colimits of presheaves

The main questions of this section are these: what do limits and colimits of presheaves look like? In particular, what about limits and colimit of representables—for instance, are they still representable? How much of this extends from presheaves (set-valued functors) to functors in general?

We begin with the observation that for a category $\mathcal{A}$ with binary products, there is a bijection

$$\mathcal{A}(A, X \times Y) \cong \mathcal{A}(A, X) \times \mathcal{A}(A, Y) \tag{5:1}$$

natural in $A, X, Y \in \mathcal{A}$. This basically says that a map $A \longrightarrow X \times Y$ amounts to a pair of maps ($A \longrightarrow X, A \longrightarrow Y$). Is this a special feature of products, or does some analogous statement hold for every kind of limit? Let's try equalizers. Suppose that $\mathcal{A}$ has equalizers, and write $\mathrm{Eq}(X \underset{g}{\overset{f}{\rightrightarrows}} Y)$ for the equalizer of maps f and g. Then maps from A to $\mathrm{Eq}(X \underset{g}{\overset{f}{\rightrightarrows}} Y)$ correspond one-to-one with maps $q : A \longrightarrow X$ such that $f \circ q = g \circ q$. In other words, there is a canonical bijection

$$\mathcal{A}\left(A, \mathrm{Eq}(X \underset{g}{\overset{f}{\rightrightarrows}} Y)\right) \cong \mathrm{Eq}\left(\mathcal{A}(A, X) \underset{\mathcal{A}(A,g)}{\overset{\mathcal{A}(A,f)}{\rightrightarrows}} \mathcal{A}(A, Y)\right) \qquad (5{:}2)$$

where $\mathcal{A}(A, f) = H^A(f) = f \circ -$ and similarly $\mathcal{A}(A, g)$. This looks very much like (5:1). And indeed, (5:1) and (5:2) can be generalized to arbitrary limit shapes. We now set about proving this.

Lemma 5.1.1 *Let $\mathbb{I}$ be a small category, $\mathcal{A}$ a locally small category, $D : \mathbb{I} \longrightarrow \mathcal{A}$ a diagram, and $A \in \mathcal{A}$. Then*

$$[\mathbb{I}, \mathcal{A}](\Delta A, D) \cong \lim_{\longleftarrow \mathbb{I}} (H^A \circ D).$$

Remark In the proof of Proposition 4.4.2 we met the important fact that $[\mathbb{I}, \mathcal{A}](\Delta A, D)$ is the set of cones on D with vertex A. We will use this without mention. Also note that $H^A \circ D$ is the functor

$$\begin{array}{rccc} \mathcal{A}(A, D) : & \mathbb{I} & \longrightarrow & \mathbf{Set} \\ & I & \longmapsto & \mathcal{A}(A, D(I)), \end{array}$$

which, like all diagrams in **Set**, does have a limit.

Proof We have

$$\begin{array}{rcl} \lim\limits_{\longleftarrow \mathbb{I}} (H^A \circ D) & \cong & \{(q_I)_{I \in \mathbb{I}} \mid \; q_I \in \mathcal{A}(A, D(I)) \text{ for all } I \in \mathbb{I} \text{ and} \\ & & \qquad (Du) \circ q_I = q_J \text{ for all } I \xrightarrow{u} J \text{ in } \mathbb{I}\} \\ & \cong & [\mathbb{I}, \mathcal{A}](\Delta A, D), \end{array}$$

the first isomorphism by the explicit formula (4:17) for limits in **Set**, and the second by definition of natural transformation. □

Proposition 5.1.2 (Representables preserve limits) *Let $\mathcal{A}$ be a locally small category and $A \in \mathcal{A}$. Then $H^A : \mathcal{A} \longrightarrow$ **Set** preserves limits.*

Proof Let $\mathbb{I}$ be a small category and $D : \mathbb{I} \longrightarrow \mathcal{A}$. Then

$$\begin{array}{rcl} H^A(\lim\limits_{\longleftarrow \mathbb{I}} D) & = & \mathcal{A}(A, \lim\limits_{\longleftarrow \mathbb{I}} D) \\ & \cong & [\mathbb{I}, \mathcal{A}](\Delta A, D) \\ & \cong & \lim\limits_{\longleftarrow \mathbb{I}} (H^A \circ D), \end{array}$$

the equality by definition of H^A, the first isomorphism by Proposition 4.4.2, and the second by Lemma 5.1.1. $\Box$

Remarks 5.1.3 a. This proof is a bit sloppy. To understand why, we need to go back to the definition of preservation of limits.

Let $F : \mathcal{A} \longrightarrow \mathcal{B}$ be a functor and $D : \mathbb{I} \longrightarrow \mathcal{A}$ a diagram in $\mathcal{A}$. If both D and $F \circ D$ have a limit, there is a canonical map

$$j : F(\lim_{\leftarrow \mathbb{I}} D) \longrightarrow \lim_{\leftarrow \mathbb{I}}(F \circ D). \tag{5:3}$$

This is the unique map j making the diagram

$$\begin{tikzcd}
 & & FD(I) \\
F(\lim_{\leftarrow \mathbb{I}} D) \arrow[rru, "F(\mathrm{pr}_I)"] \arrow[r, "j"'] & \lim_{\leftarrow \mathbb{I}}(F \circ D) \arrow[ru, "\mathrm{pr}_I"'] &
\end{tikzcd}$$

commute for all $I \in \mathbb{I}$. (Here 'pr_I' denotes the Ith projection in both limit cones, and there is a unique such j because $(F(\lim_{\leftarrow \mathbb{I}} D) \xrightarrow{F(\mathrm{pr}_I)} FD(I))_{I \in \mathbb{I}}$ is a cone on $F \circ D$.)

It is straightforward to show that F preserves limits of shape $\mathbb{I}$ if and only if the following condition holds: for all diagrams $D : \mathbb{I} \longrightarrow \mathcal{A}$ that have a limit in $\mathcal{A}$, the diagram $F \circ D$ has a limit in $\mathcal{B}$ and the canonical map (5:3) is an isomorphism. So if F preserves limits then certainly $F(\lim_{\leftarrow \mathbb{I}} D) \cong \lim_{\leftarrow \mathbb{I}}(F \circ D)$. But saying that there exists an isomorphism $F(\lim_{\leftarrow \mathbb{I}} D) \cong \lim_{\leftarrow \mathbb{I}}(F \circ D)$ is not enough to guarantee that j itself is an isomorphism, which is what limit-preservation means.

That's why our proof is sloppy: we constructed an isomorphism $k : H^A(\lim_{\leftarrow \mathbb{I}} D) \xrightarrow{\sim} \lim_{\leftarrow \mathbb{I}}(H^A \circ D)$, but we didn't show that k was the canonical map. To complete the proof, we'd have to show that

$$\begin{tikzcd}
 & & H^A(D(I)) \\
H^A(\lim_{\leftarrow \mathbb{I}} D) \arrow[rru, "H^A(\mathrm{pr}_I)"] \arrow[r, "k"'] & \lim_{\leftarrow \mathbb{I}}(H^A \circ D) \arrow[ru, "\mathrm{pr}_I"'] &
\end{tikzcd}$$

commutes for all $I \in \mathbb{I}$. This isn't hard, but to avoid getting bogged down in detail, I've omitted this check and other similar ones later. When we prove results of the form 'such-and-such a functor preserves limits' later, we'll often be equally sloppy. In other words, if F is the functor concerned, we'll just show that for all diagrams D, the limit of $F \circ D$ exists and is isomorphic to $F(\lim_{\leftarrow \mathbb{I}} D)$.

b. Rephrased, the Proposition says that

$$\mathcal{A}(A, \lim_{\leftarrow \mathbb{I}} D) \cong \lim_{\leftarrow \mathbb{I}} \mathcal{A}(A, D) \tag{5:4}$$

for all diagrams $D : \mathbb{I} \longrightarrow \mathcal{A}$ that have a limit. This is the form that most resembles our motivating examples, (5:1) and (5:2).

c. The dual of the Proposition (replacing $\mathcal{A}$ by $\mathcal{A}^{\mathrm{op}}$) is that $H_A : \mathcal{A}^{\mathrm{op}} \longrightarrow$ **Set** preserves limits. A limit in $\mathcal{A}^{\mathrm{op}}$ is a colimit in $\mathcal{A}$, so H_A turns colimits in $\mathcal{A}$ into limits in **Set**:

$$\mathcal{A}(\lim_{\rightarrow \mathbb{I}} D, A) \cong \lim_{\leftarrow \mathbb{I}} \mathcal{A}(D, A). \tag{5:5}$$

The right-hand side is a *limit*, not a colimit! So even though (5:4) and (5:5) are dual statements, there are more limits than colimits involved. Somehow, limits have the upper hand.

If you find this bizarre, perhaps an example will help. Let $X, Y, A \in \mathcal{A}$ and suppose that the sum $X + Y$ exists. Then a map $X + Y \longrightarrow A$ amounts to a pair of maps $(X \longrightarrow A, Y \longrightarrow A)$. In other words, there is a canonical isomorphism

$$\mathcal{A}(X + Y, A) \cong \mathcal{A}(X, A) \times \mathcal{A}(Y, A).$$

This is (5:5) in the case where $\mathbb{I}$ is the discrete category with two objects.

Earlier in the course we learned that it's sometimes useful to view functors as objects in their own right, rather than as maps of categories. For instance, if G is a group then functors $G \longrightarrow$ **Set** are G-sets. This point of view leads to the definition of functor category (Section 1.3).

We now begin an analysis of limits and colimits in functor categories. We'll study functor categories $[\mathbb{A}, \mathbb{S}]$ where $\mathbb{A}$ is small and $\mathbb{S}$ is locally small. We're especially interested in the case $\mathbb{S} =$ **Set**. In fact, you won't lose much if you think of $\mathbb{S}$ as standing for **Set** in everything that follows.

Here is the first of this week's two main theorems. It says that limits and colimits in $[\mathbb{A}, \mathbb{S}]$ work in the simplest way possible, provided that $\mathbb{S}$ has all (co)limits of whatever type we're concerned with. For instance, if $\mathbb{S}$ has binary products then so does $[\mathbb{A}, \mathbb{S}]$, and the product of two functors $X, Y : \mathbb{A} \longrightarrow \mathbb{S}$ is the functor $X \times Y : \mathbb{A} \longrightarrow \mathbb{S}$ defined by

$$(X \times Y)(A) = X(A) \times Y(A)$$

for all $A \in \mathbb{A}$.

Notation 5.1.4 Let $\mathbb{A}$ and $\mathbb{S}$ be categories. For each $A \in \mathbb{A}$, there is a functor

$$\begin{array}{rccc} \mathrm{ev}_A : & [\mathbb{A}, \mathbb{S}] & \longrightarrow & \mathbb{S} \\ & X & \longmapsto & X(A) \end{array}$$

(**evaluation at** A). So given a diagram $D : \mathbb{I} \longrightarrow [\mathbb{A}, \mathcal{S}]$, we have for each $A \in \mathbb{A}$ a functor

$$\begin{array}{rccc} \mathrm{ev}_A \circ D : & \mathbb{I} & \longrightarrow & \mathcal{S} \\ & I & \longmapsto & D(I)(A). \end{array}$$

We sometimes write $\mathrm{ev}_A \circ D$ as $D(-)(A)$.

Theorem 5.1.5 ((Co)limits in functor categories are computed pointwise) *Let $\mathbb{A}$ and $\mathbb{I}$ be small categories and $\mathcal{S}$ a locally small category. Let $D : \mathbb{I} \longrightarrow [\mathbb{A}, \mathcal{S}]$ be a diagram, and suppose that for each $A \in \mathcal{A}$, the diagram $D(-)(A) : \mathbb{I} \longrightarrow \mathcal{S}$ has a limit. Then there is a cone on D whose image under ev_A is a limit cone on $D(-)(A)$ for each $A \in \mathbb{A}$. Moreover, any such cone on D is a limit cone.*

Dual statements hold for colimits.

Proof Take for each $A \in \mathbb{A}$ a limit cone $(L(A) \xrightarrow{p_{I,A}} D(I)(A))_{I \in \mathbb{I}}$ on $D(-)(A)$. We will prove two things:

a. there is a unique way of defining L on maps in $\mathbb{A}$ such that L becomes a functor and $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ a cone on D

b. $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ is then a limit cone on D.

The result on limits follows immediately, and the result on colimits then holds by duality.

For (a), take a map $f : A \longrightarrow A'$ in $\mathbb{A}$. Lemma 4.4.4(a) applied to the natural transformation

$$\begin{tikzcd} \mathbb{I} \arrow[r, bend left=50, "D(-)(A)"{name=U}] \arrow[r, bend right=50, "D(-)(A')"'{name=B}] & \mathcal{S} \arrow[Rightarrow, from=U, to=B, "D(-)(f)"] \end{tikzcd}$$

tells us that there is a unique map $L(f) : L(A) \longrightarrow L(A')$ such that for all $I \in \mathbb{I}$, the square

$$\begin{tikzcd} L(A) \arrow[r, "p_{I,A}"] \arrow[d, "L(f)"'] & D(I)(A) \arrow[d, "D(I)(f)"] \\ L(A') \arrow[r, "p_{I,A'}"'] & D(I)(A') \end{tikzcd} \qquad (5{:}6)$$

commutes. We've now defined L on objects and maps of $\mathbb{A}$, and it's easy to check that L preserves composition and identities and therefore defines a functor $L : \mathbb{A} \longrightarrow \mathcal{S}$. Moreover, (5:6) says exactly that for each $I \in \mathbb{I}$, the family $(L(A) \xrightarrow{p_{I,A}} D(I)(A))_{A \in \mathbb{A}}$ is a natural transformation

$$\begin{tikzcd} \mathbb{A} \arrow[r, bend left=50, "L"{name=U}] \arrow[r, bend right=50, "D(I)"'{name=B}] & \mathcal{S}. \arrow[Rightarrow, from=U, to=B, "p_I"] \end{tikzcd}$$

So we have a family $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ of maps in $[\mathbb{A}, \mathcal{S}]$, and from the fact that $(L(A) \xrightarrow{p_{I,A}} D(I)(A))_{I \in \mathbb{I}}$ is a cone on $D(-)(A)$ for each $A \in \mathbb{A}$, it follows immediately that $(L \xrightarrow{p_I} D(I))_{I \in \mathbb{I}}$ is a cone on D.

For (b), let $X \in [\mathbb{A}, \mathcal{S}]$ and let $(X \xrightarrow{q_I} D(I))_{I \in \mathbb{I}}$ be a cone on D in $[\mathbb{A}, \mathcal{S}]$. For each $A \in \mathbb{A}$, we have a cone $(X(A) \xrightarrow{q_{I,A}} D(I)(A))_{I \in \mathbb{I}}$ on $D(-)(A)$ in $\mathcal{S}$, so there is a unique map $\overline{q}_A : X(A) \longrightarrow L(A)$ such that $p_{I,A} \circ \overline{q}_A = q_{I,A}$ for all $I \in \mathbb{I}$. We need only prove that $\overline{q}_A$ is natural in A, and this follows from Lemma 4.4.4(b). □

Corollary 5.1.6 *Let $\mathbb{A}$ be a small category and $\mathcal{S}$ a locally small category. If $\mathcal{S}$ has all (co)limits (or all (co)limits of a particular shape) then so does $[\mathbb{A}, \mathcal{S}]$, and for each $A \in \mathbb{A}$, the evaluation functor $\mathrm{ev}_A : [\mathbb{A}, \mathcal{S}] \longrightarrow \mathcal{S}$ preserves them.* □

Warning 5.1.7 If $\mathcal{S}$ does *not* have all limits of shape $\mathbb{I}$ then $[\mathbb{A}, \mathcal{S}]$ may contain certain limits of shape $\mathbb{I}$ that are not 'computed pointwise'—in other words, are not preserved by all the evaluation functors. Examples can be constructed.

Exercise 5.1.8 Let $\mathbb{A}$ be a small category and $\mathcal{S}$ a locally small category with pullbacks. Show that a natural transformation

$$\begin{tikzcd}
\mathbb{A} \arrow[r, bend left=50, "X", ""{name=U, below}] \arrow[r, bend right=50, "Y"', ""{name=D}] & \mathcal{S} \arrow[Rightarrow, from=U, to=D, "\alpha"]
\end{tikzcd}$$

is monic (as a map in $[\mathbb{A}, \mathcal{S}]$) if and only if α_A is monic for all $A \in \mathbb{A}$. (Hint: use Lemma 4.1.31.)

The Theorem has several important consequences. For example, it will help us prove that limits commute with limits. Take categories $\mathbb{I}$, $\mathbb{J}$ and $\mathcal{S}$. There are isomorphisms of categories

$$[\mathbb{I}, [\mathbb{J}, \mathcal{S}]] \cong [\mathbb{I} \times \mathbb{J}, \mathcal{S}] \cong [\mathbb{J}, [\mathbb{I}, \mathcal{S}]]$$

(compare 3.1.14(c) and Sheet 1, q.5), with $D : \mathbb{I} \times \mathbb{J} \longrightarrow \mathcal{S}$ corresponding to

$$\begin{array}{rccc} D^\bullet : & \mathbb{I} & \longrightarrow & [\mathbb{J}, \mathcal{S}], \\ & I & \longmapsto & D(I, -) \end{array} \quad \text{and} \quad \begin{array}{rccc} D_\bullet : & \mathbb{J} & \longrightarrow & [\mathbb{I}, \mathcal{S}], \\ & J & \longmapsto & D(-, J). \end{array}$$

Now suppose that $\mathcal{S}$ has all the limits that we might want. Then there is an object $\lim\limits_{\leftarrow \mathbb{I}} D^\bullet$ of $[\mathbb{J}, \mathcal{S}]$. This is itself a diagram, so we obtain an object $\lim\limits_{\leftarrow \mathbb{J}} \lim\limits_{\leftarrow \mathbb{I}} D^\bullet$ of $\mathcal{S}$. Might this also be the limit of $D : \mathbb{I} \times \mathbb{J} \longrightarrow \mathcal{S}$? And might it be the same as $\lim\limits_{\leftarrow \mathbb{I}} \lim\limits_{\leftarrow \mathbb{J}} D_\bullet$? Here's the answer:

Corollary 5.1.9 (Limits commute with limits) *Let $\mathbb{I}$ and $\mathbb{J}$ be small categories. Let $\mathcal{S}$ be a locally small category with limits of shape $\mathbb{I}$ and of shape $\mathbb{J}$. Then for all $D : \mathbb{I} \times \mathbb{J} \longrightarrow \mathcal{S}$, we have*

$$\lim_{\longleftarrow \mathbb{J}} \lim_{\longleftarrow \mathbb{I}} D^{\bullet} \cong \lim_{\longleftarrow \mathbb{I}\times\mathbb{J}} D \cong \lim_{\longleftarrow \mathbb{I}} \lim_{\longleftarrow \mathbb{J}} D_{\bullet}$$

and all these limits exist. In particular, $\mathcal{S}$ has limits of shape $\mathbb{I} \times \mathbb{J}$.

This is sometimes half-jokingly called Fubini's Theorem: it's like the interchange of integrals.

Proof Since $\mathcal{S}$ has limits of shape $\mathbb{I}$, so does $[\mathbb{J}, \mathcal{S}]$ (by the last Corollary). So $\lim_{\longleftarrow \mathbb{I}} D^{\bullet}$ exists; it is an object of $[\mathbb{J}, \mathcal{S}]$. Since $\mathcal{S}$ has limits of shape $\mathbb{J}$, $\lim_{\longleftarrow \mathbb{J}} \lim_{\longleftarrow \mathbb{I}} D^{\bullet}$ exists; it is an object of $\mathcal{S}$. Then for $S \in \mathcal{S}$,

$$\begin{aligned} \mathcal{S}(S, \lim_{\longleftarrow \mathbb{J}} \lim_{\longleftarrow \mathbb{I}} D^{\bullet}) &\cong [\mathbb{J}, \mathcal{S}](\Delta S, \lim_{\longleftarrow \mathbb{I}} D^{\bullet}) \\ &\cong [\mathbb{I}, [\mathbb{J}, \mathcal{S}]](\Delta(\Delta S), D^{\bullet}) \\ &\cong [\mathbb{I} \times \mathbb{J}, \mathcal{S}](\Delta S, D) \end{aligned}$$

naturally in S. (The first two steps follow from Proposition 4.4.2, which rephrases the definition of limit in terms of representability. The third uses the isomorphism $[\mathbb{I}, [\mathbb{J}, \mathcal{S}]] \cong [\mathbb{I} \times \mathbb{J}, \mathcal{S}]$, under which $\Delta(\Delta S)$ corresponds to ΔS and $D^{\bullet}$ corresponds to D.) So $\lim_{\longleftarrow \mathbb{J}} \lim_{\longleftarrow \mathbb{I}} D^{\bullet}$ is a representing object for the functor $[\mathbb{I} \times \mathbb{J}, \mathcal{S}](\Delta -, D)$, and by Proposition 4.4.2 again, this says that $\lim_{\longleftarrow \mathbb{I}\times\mathbb{J}} D$ exists and is isomorphic to $\lim_{\longleftarrow \mathbb{J}} \lim_{\longleftarrow \mathbb{I}} D^{\bullet}$.

The other isomorphism follows by symmetry. □

Examples 5.1.10 a. Suppose that $\mathbb{I} = \mathbb{J} = \boxed{\bullet \qquad \bullet}$. Then the Corollary says that binary products commute with binary products: if $\mathcal{S}$ has binary products and $S_{11}, S_{12}, S_{21}, S_{22} \in \mathcal{S}$ then the 4-fold product $\prod_{i,j\in\{1,2\}} S_{ij}$ exists and satisfies

$$(S_{11} \times S_{21}) \times (S_{12} \times S_{22}) \cong \prod_{i,j\in\{1,2\}} S_{ij} \cong (S_{11} \times S_{12}) \times (S_{21} \times S_{22}).$$

More generally, it doesn't matter what order you write products in or where you put the brackets: there are canonical isomorphisms

$$\begin{aligned} S \times T &\cong T \times S \\ (S \times T) \times U &\cong S \times (T \times U) \end{aligned}$$

in any category with binary products. If there is also a terminal object 1, there are further canonical isomorphisms

$$S \times 1 \cong S \cong 1 \times S.$$

b. The dual of the corollary says that colimits commute with colimits; for instance,

$$(S_{11} + S_{21}) + (S_{12} + S_{22}) \cong (S_{11} + S_{12}) + (S_{21} + S_{22})$$

in any category $\mathcal{S}$ with binary sums. But limits do *not* in general commute with colimits. For instance, in general,

$$(S_{11} + S_{21}) \times (S_{12} + S_{22}) \not\cong (S_{11} \times S_{12}) + (S_{21} \times S_{22}).$$

A counterexample is given by taking $\mathcal{S} = \mathbf{Set}$ and each S_{ij} to be a one-element set. Then the left-hand side has $(1 + 1) \times (1 + 1) = 4$ elements, but the right-hand side has $(1 \times 1) + (1 \times 1) = 2$ elements.

Here are some further consequences of Theorem 5.1.5.

Corollary 5.1.11 *Let $\mathbb{A}$ be a small category. Then $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ has all limits and colimits, and for each $A \in \mathbb{A}$, the evaluation functor $\mathrm{ev}_A : [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}] \longrightarrow \mathbf{Set}$ preserves them.*

Proof This follows from Corollary 5.1.6, using the fact that **Set** has all limits and colimits. □

Corollary 5.1.12 *For every small category $\mathbb{A}$, the Yoneda embedding $H_\bullet : \mathbb{A} \longrightarrow [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ preserves limits.*

Warning 5.1.13 The Yoneda embedding does *not* preserve colimits. For instance, if $\mathbb{A}$ has an initial object 0 then H_0 is not initial, since $H_0 = \mathbb{A}(0, 0)$ is a one-element set but the initial object of $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ is the constant presheaf $\Delta\emptyset$.

Proof of Corollary 5.1.12 Let $\mathbb{I}$ be a small category and $D : \mathbb{I} \longrightarrow \mathbb{A}$ a diagram in $\mathbb{A}$. For each $A \in \mathbb{A}$, the composite functor

$$\mathbb{A} \xrightarrow{H_\bullet} [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}] \xrightarrow{\mathrm{ev}_A} \mathbf{Set}$$

is H_A, which preserves limits (5.1.2). So for each $A \in \mathbb{A}$, the image under $\mathrm{ev}_A \circ H_\bullet$ of any limit cone on D is a limit cone on $\mathrm{ev}_A \circ H_\bullet \circ D$. By the 'moreover' of Theorem 5.1.5, the image under $H_\bullet$ of any limit cone on D is a limit cone on $H_\bullet \circ D$. In other words, $H_\bullet$ preserves limits of D. □

We've now seen that limits of representables are 'easy': the Yoneda embedding $\mathbb{A} \hookrightarrow [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ preserves them. For instance, suppose that $\mathbb{A}$ has binary products. Then for $A, B \in \mathbb{A}$, we have $H_A \times H_B \cong H_{A \times B}$. Let's view $\mathbb{A}$ as a subcategory of $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ (as in the picture on page 62), identifying $A \in \mathbb{A}$ with the representable $H_A \in [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$. Then the isomorphism means that given two objects of $\mathbb{A}$ that we want to form the product of, it doesn't matter whether we think of the product as taking place in $\mathbb{A}$ or in $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$. Similarly, if $\mathbb{A}$ has all limits, then taking limits doesn't help you escape from $\mathbb{A}$ into the rest of $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$: any limit of representable presheaves is again representable.

Colimits are a different matter. The Warning above shows that the Yoneda embedding doesn't preserve them. Similarly, colimits of representable presheaves are often not representable. In fact, the situation for colimits is at the opposite extreme from the situation for limits: by taking colimits of representable presheaves, you can get any presheaf you like! That's the second of this week's main theorems.

I've already mentioned the analogy with prime numbers. Every positive integer can be expressed as a product of primes in an essentially unique way. Every presheaf can be expressed as a colimit of representables in a canonical (though not unique) way. The representables are the 'building blocks' of presheaves.

A different analogy: any complex function holomorphic in a neighbourhood of 0 has a power series expansion, such as

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .$$

So the power functions $z \longmapsto z^n$ are the 'building blocks' of holomorphic functions. (You can take this further: $(\ )^n$ is like $\mathrm{Hom}(n, -)$, and quotients and sums are, in the categorical context, colimits.)

Before we state and prove the theorem, let's do an easy special case.

Example 5.1.14 Let $\mathbb{A}$ be the discrete category with two objects, 1 and 2. A presheaf X on $\mathbb{A}$ is just a pair $(X(1), X(2))$ of sets, and $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}] \cong \mathbf{Set} \times \mathbf{Set}$. There are two representables, H_1 and H_2, given by

$$H_i(j) = \mathbb{A}(j, i) \cong \begin{cases} 1 & \text{if } i = j \\ \emptyset & \text{if } i \neq j \end{cases}$$

($i, j \in \{1, 2\}$). Identifying $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ with $\mathbf{Set} \times \mathbf{Set}$, we have $H_1 \cong (1, \emptyset)$ and $H_2 \cong (\emptyset, 1)$. Every object of $\mathbf{Set} \times \mathbf{Set}$ is a sum of copies of $(1, \emptyset)$ and $(\emptyset, 1)$. Suppose, for instance, that $X(1)$ has 4 elements and $X(2)$ has 3 elements: then

$$(X(1), X(2)) \cong (1, \emptyset) + (1, \emptyset) + (1, \emptyset) + (1, \emptyset) + (\emptyset, 1) + (\emptyset, 1) + (\emptyset, 1)$$

in $\mathbf{Set} \times \mathbf{Set}$. Equivalently,

$$X \cong H_1 + H_1 + H_1 + H_1 + H_2 + H_2 + H_2$$

in $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$, which exhibits X as a sum of representables.

In this example, X is expressed as a sum of 7 representables—that is, as a sum indexed by the set $X(1) + X(2)$ of 'elements' of X. (Recall that a sum is a colimit over a discrete category. That we're using sums here is a consquence of $\mathbb{A}$ being discrete.) In the general case, a presheaf X on a category $\mathbb{A}$ will be expressed as a colimit over a category whose objects can be thought of as the 'elements' of X, as in the following definition.

Definition 5.1.15 Let $\mathbb{A}$ be a category and X a presheaf on $\mathbb{A}$. The **category of elements** $\mathbb{E}(X)$ of X is the category with

- objects: pairs (A, x) where $A \in \mathbb{A}$ and $x \in X(A)$
- maps $(A', x') \longrightarrow (A, x)$: maps $f : A' \longrightarrow A$ in $\mathbb{A}$ such that $(Xf)(x) = x'$.

There is a projection functor $P : \mathbb{E}(X) \longrightarrow \mathbb{A}$ defined by $P(A, x) = A$ and $P(f) = f$.

Theorem 5.1.16 (Every presheaf is a colimit of representables) *Let $\mathbb{A}$ be a small category and X a presheaf on $\mathbb{A}$. Then X is the colimit of the diagram*

$$\mathbb{E}(X) \xrightarrow{P} \mathbb{A} \xrightarrow{H_\bullet} [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$$

of shape $\mathbb{E}(X)$ in $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$.

Proof First note that since $\mathbb{A}$ is small, so too is $\mathbb{E}(X)$. Hence $H_\bullet \circ P$ is a diagram in our usual sense.

Now let $Y \in [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$. A cocone on $H_\bullet \circ P$ with vertex Y is a family

$$(H_A \xrightarrow{\alpha_{A,x}} Y)_{A \in \mathbb{A}, x \in X(A)}$$

of natural transformations with the property that for all maps $A' \xrightarrow{f} A$ in $\mathbb{A}$ and all $x \in X(A)$, the diagram

$$\begin{tikzcd}
H_{A'} \arrow[d, "H_f"'] \arrow[dr, "\alpha_{A',(Xf)(x)}"] & \\
H_A \arrow[r, "\alpha_{A,x}"'] & Y
\end{tikzcd}$$

commutes. Equivalently (by the Yoneda Lemma), such a cocone is a family $(y_{A,x})_{A \in \mathbb{A}, x \in X(A)}$ with the property that for all maps $A' \xrightarrow{f} A$ in $\mathbb{A}$ and all $x \in X(A)$,

$$(Yf)(y_{A,x}) = y_{A',(Xf)(x)}$$

(since if $\alpha_{A,x} \in [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}](H_A, Y)$ corresponds to $y_{A,x} \in Y(A)$ then $\alpha_{A,x} \circ H_f \in [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}](H_{A'}, Y)$ corresponds to $(Yf)(y_{A,x}) \in Y(A')$). Equivalently (writing $y_{A,x}$ as $\overline{\alpha}_A(x)$), it's a family

$$(X(A) \xrightarrow{\overline{\alpha}_A} Y(A))_{A \in \mathbb{A}}$$

of functions with the property that for all maps $A' \xrightarrow{f} A$ in $\mathbb{A}$ and all $x \in X(A)$,

$$(Yf)(\overline{\alpha}_A(x)) = \overline{\alpha}_{A'}((Xf)(x)).$$

Equivalently, it's a natural transformation $\overline{\alpha} : X \longrightarrow Y$. So we have for each $Y \in [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ a canonical bijection

$$[\mathbb{E}(X), [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]](H_\bullet \circ P, \Delta Y) \cong [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}](X, Y).$$

Hence X is the colimit of $H_\bullet \circ P$. □

Remarks 5.1.17 a. You can go through Example 5.1.14 and show that the way of expressing X as a sum of representables described there is a special case of the general construction in the Theorem. Since $\mathbb{A}$ is discrete, the category of elements $\mathbb{E}(X)$ is also discrete; it's the set $X(1) + X(2)$ with seven elements. The projection $P : \mathbb{E}(X) \longrightarrow \mathbb{A}$ sends four of the elements to 1 and the other three to 2, so the diagram $H_\bullet \circ P : \mathbb{E}(X) \longrightarrow [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ sends four of the elements to H_1 and three to H_2. The colimit is the sum of these seven representables, which is X.

b. The terminology 'category of elements' is compatible with the 'generalized element' terminology introduced earlier (page 79). A generalized element of an object X is just a map into X, say $Z \longrightarrow X$, but we're often interested in certain special values of Z. In **Set**, for instance, taking $Z = 1$ gives ordinary elements; in **Cat**, taking $Z = \mathbf{1} = \boxed{\bullet}$ gives objects and taking $Z = \mathbf{2} = \boxed{\bullet \longrightarrow \bullet}$ gives arrows. Further examples are in the introduction to Chapter 3. Now suppose we're in a presheaf category $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$, and consider the generalized elements $Z \longrightarrow X$ where Z is representable. By Yoneda, these generalized elements correspond to the pairs (A, x) with $A \in \mathbb{A}$ and $x \in X(A)$. In other words, they're the objects of the category of elements.

c. The fact that every presheaf is a colimit of representables is sometimes called 'density'. A subspace A of a metric space B is called dense if every point in B can be obtained as a limit of points in A. Similarly, $\mathbb{A}$ is 'dense' in $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ because every object of $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ can be obtained as a colimit of objects of $\mathbb{A}$.

5.2 Interaction of (co)limits with adjunctions

> *This week's reading contains a lot of discussion. As far as the exam is concerned, the most important results are Theorems 5.2.1 and 5.2.16.*

We saw in 3.1.5 that any **Set**-valued functor with a left adjoint is representable, and in 5.1.2 that any representable preserves limits. Hence, any **Set**-valued functor with a left adjoint preserves limits. In fact, this conclusion doesn't just hold for functors into **Set**: it's true in complete generality.

Theorem 5.2.1 *Let* $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathcal{B} \arrow[l, shift left, "G"] \end{tikzcd}$ *be an adjunction. Then* F *preserves colimits and* G *preserves limits.*

Proof By duality, it is enough to prove that G preserves limits. Let $D : \mathbb{I} \longrightarrow \mathcal{B}$ be a diagram that has a limit. Then

$$\begin{aligned} \mathcal{A}(A, G(\lim_{\leftarrow \mathbb{I}} D)) &\cong \mathcal{B}(F(A), \lim_{\leftarrow \mathbb{I}} D) \\ &\cong \lim_{\leftarrow \mathbb{I}} \mathcal{B}(F(A), D) \\ &\cong \lim_{\leftarrow \mathbb{I}} \mathcal{A}(A, G \circ D) \\ &\cong [\mathbb{I}, \mathcal{A}](\Delta A, G \circ D) \end{aligned}$$

naturally in $A \in \mathcal{A}$. (The first and third isomorphisms follow from adjointness, the second from Remark 5.1.3(b), and the last from Lemma 5.1.1.) So $G(\lim_{\leftarrow \mathbb{I}} D)$ represents $[\mathbb{I}, \mathcal{A}](\Delta -, G \circ D)$; that is, it is a limit of $G \circ D$. $\square$

Examples 5.2.2 a. Forgetful functors from categories of algebras to **Set** have left adjoints, but hardly ever right adjoints. Correspondingly, they preserve all limits, but hardly ever preserve all colimits.

b. For every set B there is an adjunction $(- \times B) \dashv (-)^B$ of functors from **Set** to **Set** (Example 2.1.5). So $- \times B$ preserves colimits and $(-)^B$ preserves limits. In particular, $- \times B$ preserves finite sums and $(-)^B$ preserves finite products, giving isomorphisms

$$0 \times B \cong 0, \qquad (A_1 + A_2) \times B \cong (A_1 \times B) + (A_2 \times B) \qquad (5{:}7)$$

$$1^B \cong 1, \qquad (A_1 \times A_2)^B \cong {A_1}^B \times {A_2}^B \qquad (5{:}8)$$

—the analogues of standard rules of arithmetic. (Compare Examples 5.1.10 and the 'Digression on arithmetic' on page 38.) You can also prove the isomorphisms directly, but this proof is easier.

c. If $\mathcal{A}$ is a category with all limits of some shape $\mathbb{I}$ then there is an adjunction $\begin{tikzcd} \mathcal{A} \arrow[r, shift left, "\Delta"] \arrow[r, phantom, "\perp"] & {[\mathbb{I}, \mathcal{A}]} \arrow[l, shift left, "\lim_{\leftarrow \mathbb{I}}"] \end{tikzcd}$ (Proposition 4.4.5). Hence $\lim_{\leftarrow \mathbb{I}}$ preserves limits, or equivalently, limits of shape $\mathbb{I}$ commute with (all) limits. This gives another proof that limits commute with limits, at least in the case when enough limits exist.

d. The Theorem is often used to prove that a functor *doesn't* have an adjoint. For instance, I claimed in Example 2.1.3(e) that the forgetful functor $\mathbf{Field} \longrightarrow \mathbf{Set}$ does not have a left adjoint. We can now prove this: if it had a left adjoint $F : \mathbf{Set} \longrightarrow \mathbf{Field}$, then F would preserve colimits and in particular initial objects. Hence $F(\emptyset)$ would be an initial object of **Field**. But **Field** has no initial object, since there are no maps between fields of different characteristics.

   Further examples of non-existence of adjoints are in Sheet 10, q.2.

Adjoint Functor Theorems

Any functor with a left adjoint preserves limits, but limit-preservation alone does not guarantee the existence of a left adjoint. For example, let $\mathcal{B}$ be any category; then the unique functor $\mathcal{B} \longrightarrow \mathbf{1}$ always preserves limits, but only has a left adjoint if $\mathcal{B}$ has an initial object.

On the other hand, if you've got a limit-preserving functor $G : \mathcal{B} \longrightarrow \mathcal{A}$ *where* $\mathcal{B}$ *has all limits* then there's an excellent chance that G will have a left adjoint. It's still not always true, but it's quite hard to find exceptions: for instance (taking $\mathcal{A} = \mathbf{1}$), can you find a category $\mathcal{B}$ that has all limits but no initial object?

This condition is so important that it has its own word:

Definition 5.2.3 A category is **complete** if it has all limits.

There are various results called Adjoint Functor Theorems, all of the following form:

> *Let $\mathcal{A}$ be a category, $\mathcal{B}$ a complete category, and $G : \mathcal{B} \longrightarrow \mathcal{A}$ a functor. Suppose that $\mathcal{A}$, $\mathcal{B}$ and G satisfy some further conditions. Then*
>
> $$G \text{ has a left adjoint} \iff G \text{ preserves limits.}$$

The '$\Rightarrow$' part of this statement is immediate (5.2.1). Our concern here is '$\Leftarrow$'.

Typically the 'further conditions' are rather fiddly, and we won't go into them much. But there is a special case in which the complications disappear, and I will use this case to explain the main idea behind the proof of the Adjoint Functor Theorems. It is the case where the categories $\mathcal{A}$ and $\mathcal{B}$ are ordered sets.

Limits in ordered sets are greatest lower bounds (GLBs). (Precisely, if $D : \mathbb{I} \longrightarrow \mathbb{B}$ is a diagram in an ordered set $\mathbb{B}$ then $\lim\limits_{\leftarrow \mathbb{I}} D = \bigwedge_{I \in \mathbb{I}} D(I)$, with one side defined if and only if the other is.) An ordered set is therefore complete if and only if every subset has a GLB. A map $G : \mathbb{B} \longrightarrow \mathbb{A}$ of ordered sets preserves limits if and only if $G(\bigwedge_{i \in I} B_i) = \bigwedge_{i \in I} G(B_i)$ for all families $(B_i)_{i \in I}$ of elements of $\mathbb{B}$ possessing a GLB.

We will need to use a simple observation. Consider a subset S of an ordered set $\mathbb{B}$, and let us ask whether S has a least element (that is, an element $l \in S$ such that $l \leq s$ for all $s \in S$). If it does, it is the GLB $\bigwedge_{s \in S} s$ in $\mathbb{B}$. Now, it may be that $\bigwedge_{s \in S} s$ exists but is not in S: for instance, the subset $(0, 1]$ of $\mathbb{R}$ has GLB (infimum) 0, but $0 \notin (0, 1]$. However, if $\bigwedge_{s \in S} s$ exists and *is* in S then it is the least element of S. So deciding whether S has a least element is equivalent to deciding whether $\bigwedge_{s \in S} s$ exists and is in S.

Exercise 5.2.4 Prove the statements in the last paragraph. This is very easy once you've understood the definition of greatest lower bound.

We now show that for ordered sets, there is an Adjoint Functor Theorem of the simplest possible kind: there are no 'further conditions' at all.

Proposition 5.2.5 (Adjoint Functor Theorem for ordered sets) *Let $\mathbb{A}$ be an ordered set and $\mathbb{B}$ a complete ordered set. Let $G : \mathbb{B} \longrightarrow \mathbb{A}$ be an order-preserving map. Then*

$$G \text{ has a left adjoint} \iff G \text{ preserves greatest lower bounds.}$$

Proof Suppose that G preserves GLBs. By Corollary 2.3.6, it's enough to show that for each $A \in \mathbb{A}$, the comma category $(A \Rightarrow G)$ has an initial object. Let $A \in \mathbb{A}$. Then $(A \Rightarrow G)$ is also an ordered set: it's $\{B \in \mathbb{B} \mid A \leq G(B)\}$, with order inherited from $\mathbb{B}$. We have to show that $(A \Rightarrow G)$ has an initial object, that is, a least element. By the observation above, this is equivalent to showing that $\bigwedge_{B \in (A \Rightarrow G)} B$ exists and is an element of $(A \Rightarrow G)$. It exists since $\mathbb{B}$ is complete, and is an element of $(A \Rightarrow G)$ since

$$G\Big(\bigwedge_{B \in (A \Rightarrow G)} B\Big) = \bigwedge_{B \in (A \Rightarrow G)} G(B) \ \geq A,$$

where the first step holds because G preserves GLBs and the second because A is a lower bound for the family $(G(B))_{B \in (A \Rightarrow G)}$. □

In the general setting of Corollary 2.3.6, the initial object of $(A \Rightarrow G)$ is the pair $(F(A),\, A \xrightarrow{\eta_A} GF(A))$ where F is the left adjoint and η the unit map. So in the Proposition, the left adjoint F is given by $F(A) = \bigwedge_{B \in (A \Rightarrow G)} B$.

Let's now try to extend the Proposition from ordered sets to categories, starting with a functor $G : \mathcal{B} \longrightarrow \mathcal{A}$. Greatest lower bounds become limits. In the case of ordered sets, we had for each A an inclusion map $(A \Rightarrow G) \hookrightarrow \mathbb{B}$, and $F(A)$ was the limit of this inclusion (understood as a diagram in $\mathbb{B}$). In the general case, the inclusion becomes the projection functor

$$\begin{array}{rccc} D: & (A \Rightarrow G) & \longrightarrow & \mathbb{B} \\ & (B,\, A \xrightarrow{f} G(B)) & \longmapsto & B \end{array}$$

and so we would like to define $F(A) = \lim\limits_{\longleftarrow (A \Rightarrow G)} D$. It can be shown that if this limit in $\mathcal{B}$ exists and is preserved by G then this formula does indeed define a left adjoint F to G.

But doesn't this mean that our Adjoint Functor Theorem generalizes smoothly to arbitrary categories, with no need for 'further conditions'?

The answer is no, and the reason is quite subtle. When we defined limit, we built in the condition that the shape category was small. (See Remark 4.1.23(c).) Similarly, we defined 'has all limits' to mean 'has all small limits' (4.1.26), and 'complete' to mean the same thing. So our assumptions are that $\mathbb{B}$ has, and G preserves, *small* limits. Now, if $\mathcal{B}$ is a large category then $(A \Rightarrow G)$ may also be large, since to specify an object or map in $(A \Rightarrow G)$ we have to specify among other things an object or map in $\mathcal{B}$. So $\lim\limits_{\leftarrow (A\Rightarrow G)} D$ may be a large limit, and there is no guarantee that it exists in $\mathcal{B}$ or is preserved by G.

(If you're having trouble with the small/large distinction, compare finite/infinite. For instance, if $\mathcal{B}$ is a finite category then $(A \Rightarrow G)$ is also finite, but if $\mathcal{B}$ is infinite then $(A \Rightarrow G)$ may be infinite too.)

Proposition 5.2.5 still stands, since there we were dealing with ordered *sets*, which are small. You might hope to extend 5.2.5 to arbitrary small categories, since the problem just described only affects large categories. But in fact, the only small categories with all small limits are ordered sets. (For a proof, see V.2 of *Categories for the Working Mathematician.*) So this get us nowhere.

We could also salvage the argument if we assumed that $\mathcal{B}$ had, and G preserved, *all* (possibly large) limits. But again this is not helpful: there are almost no such categories $\mathcal{B}$.

So the situation becomes more complicated. The Adjoint Functor Theorems impose further conditions implying that the large limit $\lim\limits_{\leftarrow (A\Rightarrow G)} D$ can be replaced by a small limit in some clever way, and this allows the argument sketched above to proceed.

I will say a little about the two most famous Adjoint Functor Theorems: the General and the Special. The statements are non-examinable, and the proofs are omitted. However, some of their implications are important general knowledge.

Definition 5.2.6 Let $\mathcal{C}$ be a category. A **weakly initial set** in $\mathcal{C}$ is a set $\mathbb{S}$ of objects with the property that for each $C \in \mathcal{C}$, there exist an element $S \in \mathbb{S}$ and a map $S \longrightarrow C$.

Note that $\mathbb{S}$ must be a set—that is, small. So the existence of a weakly initial set is some kind of size restriction.

Theorem 5.2.7 (General Adjoint Functor Theorem (GAFT)) *Let $\mathcal{A}$ be a category and $\mathcal{B}$ a complete, locally small category. Let $G : \mathcal{B} \longrightarrow \mathcal{A}$ be a functor such that for each $A \in \mathcal{A}$, the category $(A \Rightarrow G)$ has a weakly initial set. Then*

$$G \text{ has a left adjoint} \iff G \text{ preserves limits.}$$

Proof See V.6 of *Categories for the Working Mathematician.* □

Examples 5.2.8 a. The GAFT guarantees that for any category $\mathcal{B}$ of algebras (**Gp**, **Vect**$_k$, ...), the forgetful functor $U : \mathcal{B} \longrightarrow$ **Set** has a left adjoint. For we have already seen in Section 4.3 that $\mathcal{B}$ has all limits and U preserves them. Also, $\mathcal{B}$ is locally small. To apply the GAFT, we now just have to check that for every $A \in \mathbf{Set}$, the comma category $(A \Rightarrow U)$ has a weakly initial set. I'll omit this check, though if you know some cardinal arithmetic it's fairly straightforward; see 5.2.9 below.

So the GAFT tells us, for instance, that the free group functor exists. In 1.2.3(b) and 2.1.3(b) I hinted at the trickiness of constructing explicitly the free group on a generating set A; you first have to define the set of 'formal expressions' such as $x^{-1}yx^2zy^{-3}$ with $x, y, z \in A$, then say what it means for two such expressions to be equivalent (e.g. $x^{-2}x^5y$ should be equivalent to x^3y), then define $F(A)$ to be the set of all equivalence classes, then give it a group structure, then prove that it has the required universal property. But by using the GAFT, we avoid these complications entirely.

The price to be paid is that the GAFT doesn't give us an explicit description of free groups (or more generally, left adjoints). When people talk about knowing something 'explicitly', they usually mean knowing its elements. An element of an object is a map *into* it, and we've got no handle on maps into $F(A)$: since F is a left adjoint, it's maps *out* of $F(A)$ that we know about. This is why 'explicit' descriptions of left adjoints are often hard to come by.

b. More generally, the GAFT guarantees that forgetful functors between categories of algebras, such as

$$\mathbf{Ab} \longrightarrow \mathbf{Gp}, \quad \mathbf{Gp} \longrightarrow \mathbf{Mon}, \quad \mathbf{Ring} \longrightarrow \mathbf{Mon}, \quad \mathbf{Vect}_{\mathbb{C}} \longrightarrow \mathbf{Vect}_{\mathbb{R}},$$

always have left adjoints. (Some of them are described in Examples 2.1.3.) This is 'more generally' because you can view **Set** as a degenerate example of a category of algebras: a group, ring, etc. is a set equipped with some operations obeying some equations, and a set is a set equipped with no operations obeying no equations!

Digression 5.2.9 For those who know about cardinals, here is a sketch of the check that we omitted in 5.2.8(a). Take $U : \mathbf{Gp} \longrightarrow \mathbf{Set}$, for instance, and let A be a set. We have to show that $(A \Rightarrow U)$ has a weakly initial set. This reduces to two facts:

- if $(g_a)_{a \in A}$ is an A-indexed family of elements of a group G then the subgroup H that it generates can't be too big; specifically, $\text{cardinality}(H) \leq \max\{\aleph_0, \text{cardinality}(A)\}$
- for any cardinal κ, the collection of isomorphism classes of groups of cardinality $\leq \kappa$ is small.

Very similar statements hold for other forgetful functors between categories of algebras.

The Special Adjoint Functor Theorem (SAFT) operates under much tighter hypotheses than the GAFT, and is much less widely applicable. Its main advantage is that it removes the condition on weakly initial sets—indeed, it removes *all* further conditions on the functor G.

Theorem 5.2.10 (Special Adjoint Functor Theorem) *Let $\mathcal{A}$ be a locally small category and $\mathcal{B}$ a complete, locally small category satisfying certain further conditions. Then for any functor $G : \mathcal{B} \longrightarrow \mathcal{A}$,*

$$G \textit{ has a left adjoint} \iff G \textit{ preserves limits.}$$

Proof Omitted; again, see *Categories for the Working Mathematician.* □

Example 5.2.11 Here is the example *par excellence* of the SAFT. Let **CptHff** be the category of compact Hausdorff spaces and $U : \mathbf{CptHff} \longrightarrow \mathbf{Top}$ the forgetful functor. Then the SAFT tells us that U has a left adjoint F, turning any space into a compact Hausdorff space. The existence of this left adjoint is far from obvious, and verifying the conditions of the SAFT (or constructing F in any other way) requires some deep theorems of topology. $F(X)$ is called the **Stone–Čech compactification** of a space X, and contains X as a subspace provided that X satisfies some fairly mild conditions.

Another advantage of the SAFT is that it *does* give an explicit formula for left adjoints. In this case, it tells us that $F(X)$ is the closure of the image of the canonical map $X \longrightarrow [0,1]^{\mathbf{Top}(X,[0,1])}$, where the codomain is a power of $[0,1]$ in **Top**. I'll let you figure out what this canonical map must be.

Cartesian closed categories

We have seen that for every set B there is an adjunction $(- \times B) \dashv (-)^B$, and that for every small category $\mathbb{B}$ there is an adjunction $(- \times \mathbb{B}) \dashv [\mathbb{B}, -]$ (3.1.14(c)).

Definition 5.2.12 A category $\mathcal{A}$ is **cartesian closed** if it has finite products and for each $B \in \mathcal{A}$, the functor $- \times B : \mathcal{A} \longrightarrow \mathcal{A}$ has a right adjoint.

We write the right adjoint as $(-)^B$ and call C^B an **exponential**. You can think of it as the space of maps from B to C. The adjointness says that for all $A, B, C \in \mathcal{A}$ we have

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

naturally in A and C. (In fact, it's natural in B too: that comes for free.)

Examples 5.2.13

a. **Set** is cartesian closed; C^B is the function set $\mathbf{Set}(B, C)$.

b. **Cat** is cartesian closed; $\mathbb{C}^{\mathbb{B}}$ is the functor category $[\mathbb{B}, \mathbb{C}]$.

In any cartesian closed category, there are isomorphisms (5:7) and (5:8) (page 102). So the objects of a cartesian closed category possess an arithmetic very like that of the natural numbers. This thought leads in all sorts of interesting directions. Here, we'll just note that these isomorphisms provide a way of proving that a category is *not* cartesian closed.

Example 5.2.14 $\mathbf{Vect}_k$ is not cartesian closed, for any field k. It does have finite products, as we saw in 4.1.3: binary product is direct sum, $\oplus$, and the terminal object is the trivial vector space $\{0\}$ (which is also initial). But if $\mathbf{Vect}_k$ were cartesian closed then (5:7) would hold, so that $\{0\} \oplus B \cong \{0\}$ for all vector spaces B; in other words, $B \cong \{0\}$ for all B, a contradiction.

Digression 5.2.15 For any vector spaces V and W, the set $\mathbf{Vect}_k(V, W)$ of linear maps can itself be given the structure of a vector space. (When V and W are finite-dimensional, this amounts to saying that you can add together matrices of the same size, and multiply them by scalars.) Write $[V, W]$ for this vector space. Given that exponentials are meant to be 'function spaces', you might expect $\mathbf{Vect}_k$ to be cartesian closed with exponential given by $[-, -]$. We've just seen that this can't be so. In fact, the linear maps $U \longrightarrow [V, W]$ correspond to the *bi*linear maps $U \times V \longrightarrow W$, or equivalently to the linear maps $U \otimes V \longrightarrow W$. So $\mathbf{Vect}_k$ is an example of a 'monoidal closed category'—it's like a cartesian closed category, but the cartesian product $\times$ has been replaced by the tensor product $\otimes$.

It's fairly clear that $\mathbf{Set}^I$ is cartesian closed for any set I, just because $\mathbf{Set}$ is. (The exponential, as well as the product, is defined pointwise.) We now show that for any small category $\mathbb{A}$—not necessarily discrete!—the presheaf category $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ is cartesian closed. We write $\widehat{\mathbb{A}}$ for $[\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$.

Before proving it, let's conduct a thought experiment. If $\widehat{\mathbb{A}}$ *is* cartesian closed, what must exponentials be? In other words, given presheaves Y and Z, what must Z^Y be in order that

$$\widehat{\mathbb{A}}(X \times Y, Z) \cong \widehat{\mathbb{A}}(X, Z^Y)$$

for all presheaves X? Well, if it's true for all presheaves X then in particular it's true when X is representable, so

$$\widehat{\mathbb{A}}(H_A \times Y, Z) \cong \widehat{\mathbb{A}}(H_A, Z^Y) \cong Z^Y(A)$$

for all $A \in \mathbb{A}$, the last step by Yoneda. This tells us what Z^Y must be.

Theorem 5.2.16 *Let $\mathbb{A}$ be a small category. Then the presheaf category $\widehat{\mathbb{A}}$ is cartesian closed.*

Here is the strategy of the proof. The argument in the thought experiment gives us the isomorphism when X is representable. In general, X is not representable, but it is a *colimit* of representables, and that will allow us to deduce the result.

Proof $\widehat{\mathbb{A}}$ has all limits, and in particular, finite products.

Fix $Y \in \widehat{\mathbb{A}}$.

First, $- \times Y : \widehat{\mathbb{A}} \longrightarrow \widehat{\mathbb{A}}$ preserves colimits. For since products and colimits in $\widehat{\mathbb{A}}$ are computed pointwise, it's enough to prove that for any set S the functor $- \times S : \mathbf{Set} \longrightarrow \mathbf{Set}$ preserves colimits; and this follows from **Set** being cartesian closed.

For each presheaf Z on $\mathbb{A}$, define a presheaf Z^Y by

$$Z^Y(A) = \widehat{\mathbb{A}}(H_A \times Y, Z)$$

for all $A \in \mathbb{A}$. This is clearly functorial in Z, so that $(\ )^Y$ is a functor $\widehat{\mathbb{A}} \longrightarrow \widehat{\mathbb{A}}$.

I claim that $(- \times Y) \dashv (-)^Y$. For let $X, Z \in \widehat{\mathbb{A}}$ and write $P : \mathbb{E}(X) \longrightarrow \mathbb{A}$ for the projection, as in Definition 5.1.15. Then

$$\begin{aligned}
\widehat{\mathbb{A}}(X, Z^Y) &\cong \widehat{\mathbb{A}}(\lim_{\rightarrow \mathbb{E}(X)} H_P, Z^Y) && (5{:}9)\\
&\cong \lim_{\leftarrow \mathbb{E}(X)} \widehat{\mathbb{A}}(H_P, Z^Y) && (5{:}10)\\
&\cong \lim_{\leftarrow \mathbb{E}(X)} Z^Y(P) && (5{:}11)\\
&\cong \lim_{\leftarrow \mathbb{E}(X)} \widehat{\mathbb{A}}(H_P \times Y, Z) && (5{:}12)\\
&\cong \widehat{\mathbb{A}}(\lim_{\rightarrow \mathbb{E}(X)} (H_P \times Y), Z) && (5{:}13)\\
&\cong \widehat{\mathbb{A}}((\lim_{\rightarrow \mathbb{E}(X)} H_P) \times Y, Z) && (5{:}14)\\
&\cong \widehat{\mathbb{A}}(X \times Y, Z) && (5{:}15)
\end{aligned}$$

naturally in X and Z, as required. Here (5:9) and (5:15) are by Theorem 5.1.16; (5:10) and (5:13) are because representables preserve limits (as rephrased in Remark 5.1.3(c)); (5:11) is by Yoneda; (5:12) is the definition of Z^Y; and (5:14) is because $- \times Y$ preserves colimits. □

So there is an 'arithmetic of presheaves' bearing a strong resemblance to the arithmetic of natural numbers. For instance, if G is any group then there is a nice arithmetic of G-sets, since the category of (right) G-sets is the presheaf category $[G^{\mathrm{op}}, \mathbf{Set}]$.

Goodbye

There are many active research areas within category theory, and more still on the borders between category theory and other subjects—especially algebra and topology, and increasingly geometry and physics. The interplay between category theory and computer science has been particularly fertile; in fact, many of the world's category theorists work in computer labs and use category theory to advance our understanding of computer science. (These people are applied mathematicians. Don't fall into the trap of thinking that applied mathematics has to mean the application of differential equations to physics problems.)

If you want to read more on category theory, there are plenty of directions you can go in. Probably the most important topics that we haven't covered are monads and monoidal categories, described briefly in Section 0.2. You can read about them in most introductory books on category theory. If you plan to do research in algebra or algebraic topology, then sooner or later you'll probably want to learn some of the theory of abelian categories—otherwise known as homological algebra. (An abelian category is one that behaves something like $\mathbf{Vect}_k$, or more generally like the category of modules over a ring.) One good book is *An Introduction to Homological Algebra* by Weibel.

I will finish with a long (and obviously non-examinable) digression on another categorical topic: topos theory. This may help to explain the significance of the fact that presheaf categories are cartesian closed. It may also clarify some of the more mystical comments that I have made during the course.

Digression 5.2.17 A topos is a 'category of variable sets'. What can this possibly mean?

Some sets vary through time. For instance, the set of members of the University of Glasgow is not constant: it varies from year to year. Suppose that we consider all years from the year 0 to eternity. Then the 'variable set' S of members of the university is a sequence $(S_0, S_1, \ldots, S_{2007}, \ldots)$ of ordinary sets. In other words, it is an object of the category $\mathbf{Set}^{\mathbb{N}}$. And $\mathbf{Set}^{\mathbb{N}}$ is an example of a topos.

Some sets vary through space. Suppose you drive all over the world with a radio in your car. The set of radio stations that you can pick up varies as you move. (For simplicity, let's say that a station is either received or not: no halfway.) So to each point on the surface of the earth there is associated an (ordinary) set. Moreover, the way that this set varies as you move is in some sense continuous. In the jargon, this variable set is an example of a **sheaf** on the sphere; the category of sheaves on the sphere is an example of a topos. We will return to sheaves later.

Sets that vary through space can be looked at in two different ways. Just now, we took the point of view of someone who *listens* to the radio. The variable set says what radio stations are available where. But if you worked in *broadcasting*, you might view the same information in a different way: as a map of the globe covered with patches (one per radio station) showing the area to which each station is transmitted. The patches are probably roughly disk-shaped, and in any case are not just arbitrary subsets of the sphere; it's most convenient if we say that they're open sets. So they form an open cover of the sphere. The information on what radio stations are available where is perceived by listeners as something set-theoretic (a variable set) and by broadcasters as something geometric (an open cover).

The lesson is that set theory and geometry are more closely related than you might expect. Topos theory provides a way of bringing them together.

The definition of topos is obtained by writing down some of the properties of the category of (ordinary, non-varying) sets, and defining a topos to be a category that satisfies those properties. Experience shows that the right definition is this: a **topos** is a cartesian closed category with finite limits and a subobject classifier.

We've met all of these terms except 'subobject classifier'. This can be explained as follows. Given a set A, there is a natural correspondence between subsets of A and maps $A \longrightarrow 2 = \{\texttt{true}, \texttt{false}\}$ (as on page 39). We may speak of 'subobjects' of an object of an arbitrary category, generalizing the notion of subset of a set. A **subobject classifier** for a category $\mathcal{A}$ is an object $\Omega \in \mathcal{A}$ such that for $A \in \mathcal{A}$, there is a natural correspondence between subobjects of A and maps $A \longrightarrow \Omega$. So 2 is the subobject classifier of **Set**. See Sheet 9, q.5, and Sheet 10, q.4, for details.

Certainly **Set** is a topos. In a way it's a degenerate example, as a topos is a category of sets that may vary in some way, but **Set** consists of sets that don't vary at all. Every presheaf category $\widehat{\mathbb{A}} = [\mathbb{A}^{\mathrm{op}}, \mathbf{Set}]$ is a topos: we've proved in this chapter that $\widehat{\mathbb{A}}$ is cartesian closed and has (finite) limits, and it can be shown that it has a subobject classifier.

Exercise By conducting a thought experiment similar to the one before Theorem 5.2.16, work out what the subobject classifier of $\widehat{\mathbb{A}}$ must be.

Here is a further, very important, class of toposes. As we saw on page 8, there is a notion of 'presheaf' on a topological space X. **Sheaves** on X are presheaves satisfying a certain condition. The category $\mathbf{Sh}(X)$ of sheaves on X is a topos. Moreover, if you know what the topos $\mathbf{Sh}(X)$ is, you can reconstruct the space X. (I'm ignoring one subtlety here.) So in passing from X to $\mathbf{Sh}(X)$, you don't lose information; indeed, topos theorists like to view $\mathbf{Sh}(X)$ as being the *same thing* as X. Much of our intuition about toposes comes from this class of examples, and you can regard a topos as a 'generalized space' as well as a 'category of variable sets'.

Although topos theory is about variable sets, it also illuminates the world of 'constant' (ordinary, non-varying) sets. The properties of the category of sets listed informally in Section $2\frac{1}{2}$.1 can be condensed into the following: **Set** is a topos in which

- there are exactly two 'truth values' (maps from 1 to the subobject classifier)
- the Axiom of Choice holds (every epic is split)

- there is a natural numbers object (the definition of which you may be able to guess from the 'crucial property' mentioned at the bottom of page 39).

I'll call such a topos a 'category of constant sets'. The three extra conditions somehow encode the idea of constancy, or non-variation; for example, they are not satisfied by the category $\mathbf{Set}^{\mathbb{N}}$ of sets varying through time.

Here is an application of topos theory to a question about (constant) sets orginally posed by Cantor. Given sets A and B, we write $A < B$ if there exists an injection, but no bijection, from A to B. Cantor proved that $A < 2^A$ for all sets A (Lemma $2\frac{1}{2}$.2.1). He also asked: is there a set B such that $\mathbb{N} < B < 2^{\mathbb{N}}$? The **Continuum Hypothesis** is the statement that there is no such B. Now, it has been shown that there is a category of constant sets in which the Continuum Hypothesis is true, and that there is also a category of constant sets in which the Continuum Hypothesis is false. (I'm working on the foundational assumption that there is a category of constant sets at all; otherwise we can't get anywhere.) So the answer to Cantor's question is sometimes yes, sometimes no, depending on what category of sets we choose to work with.

(Incidentally, this makes it clear that I've been slightly dishonest in talking about *the* category **Set** of (constant) sets, since there are many different categories of constant sets. But it doesn't matter, because every property of **Set** that I've used can be deduced from the topos axioms plus the three properties above, and is therefore satisfied by any category of constant sets.)

We've now seen that topos theory combines aspects of geometry (or at least, topology) and set theory. It also has important connections with logic. I will explain one part of the topos/logic story, which is all about truth.

In the simplest scenario, a statement is either true or false. However, we all know that life is rarely this easy. Consider the statement 'it's raining'. Its truth certainly varies through time and space. So truth values, like sets, can be variable. For the sake of simplicity, you might insist that at any particular point in time and space, it's either true or false that it's raining. But anyone who's lived in Scotland knows that there's a third state: if someone asks us whether it's raining, we might reply 'hard to say'. (Taken literally, this sounds like a confession of ignorance—we're unable to discern which of two states the weather is in. But what we really *mean* is that the weather is in an in-between state, where the air's wet but there's not much actual falling going on.) Or you might want to have more than three options, answering the question 'is it raining?' with something like 'yes, at a rate of 5mm/hour'. Your truth values would then be real numbers. You might even want to allow answers like 'I am 95% confident that the rate of rainfall is between 4 and 6mm/hour', which is what someone with a weather station and statistical training might say. This too is a kind of truth value.

What has this got to do with topos theory? The first hint is that in the most basic topos, **Set**, the elements of the subobject classifier can usefully be called `true` and `false` (as in Section $2\frac{1}{2}$.1). In general, the subobject classifier Ω of a topos can be thought of as consisting of the 'truth values' appropriate to that topos. For example, take $\mathbf{Set}^{\mathbb{N}}$, the category of sets varying through time.

You can ask

> is Tom Leinster a member of the University of Glasgow?

but the answer varies through time. The appropriate question is really

> when is/was/will be Tom Leinster a member of the University of Glasgow?

and the answer is a subset of $\mathbb{N}$. So in $\mathbf{Set}^{\mathbb{N}}$, the truth values are the subsets of $\mathbb{N}$. Formally, the subobject classifier is $\Omega = (2, 2, \ldots)$, the terminal object is $1 = (1, 1, \ldots)$, and the maps $1 \longrightarrow \Omega$ (the 'elements of Ω') correspond one-to-one with the subsets of $\mathbb{N}$.

For another example, take sets varying on the sphere, as in the radio example. The answer to 'are you receiving Radio Moscow?' depends on your position; the question is really 'where can you receive Radio Moscow?'. So the truth values are (open) subsets of the sphere. For a general space X, the truth values in the topos of sheaves on X are the open subsets of X.

You can see some of the connections between topology and logic even if you're not comfortable with the topos idea. Let X be a set and $\mathcal{P}(X)$ its powerset. The complement operation

$$\begin{array}{rccc} \neg : & \mathcal{P}(X) & \longrightarrow & \mathcal{P}(X) \\ & U & \longmapsto & X \setminus U \end{array}$$

can be called 'not': if U consists of the elements of X satisfying some property then $\neg U$ ('not U') consists of the elements not satisfying the property. We have the identity $U \cup \neg U = X$. This is what justifies proof by contradiction: to prove that some object lies in the set U of things satisfying some property, we show that it couldn't possibly lie in $\neg U$ (for that would lead to a contradiction); and since $U \cup \neg U = X$, it must be in U. This is the classical way of setting up logic—but it's not always appropriate, as we shall see.

Now let X be a topological space and $\mathcal{O}(X)$ its set of open subsets. The nearest we can get to a complement operation on $\mathcal{O}(X)$ is

$$\begin{array}{rccc} \neg : & \mathcal{O}(X) & \longrightarrow & \mathcal{O}(X) \\ & U & \longmapsto & \text{interior of } X \setminus U. \end{array}$$

(We can't take $\neg U$ to be the set of points not in U, since that mightn't be an open set.) The law $U \cup \neg U = X$ now fails: for instance, if X is a sphere and U is the northern hemisphere then $\neg U$ is the southern hemisphere and $U \cup \neg U$ misses out the equator. This is how truth values in the 'radio topos' (sheaves on a sphere) behave. Grandly put, logic in a topos is non-classical.

What's the idea? Think of being in U as *definitely* being able to receive Radio Moscow: you can receive it, and everyone in some small region around you can receive it too. (You're not on the edge of the transmission zone.) Then $\neg U$ is where you definitely *can't* receive it: you can't receive it, and nor can anyone near you. But unless Radio Moscow is transmitted either nowhere or everywhere, the transmission zone has an edge, and so $U \cup \neg U$ is not the whole sphere. The identity $U \cup \neg U = X$ fails because there are places where you're not definitely in or out of the zone. Not definitely being able to receive it isn't the same as definitely not being able to receive it!

Scottish criminal trials work a bit like this. Juries return a guilty (or technically, 'proven') verdict if they are sure beyond reasonable doubt of the defendant's guilt. Certainty beyond reasonable doubt is like being in an open set: a small change in the evidence wouldn't change your verdict. They return a 'not guilty' verdict if they are sure of the defendant's innocence. But if they are not sure that the defendant is guilty *or* sure that s/he is innocent, they return a third verdict: 'not proven'. This is like being on the boundary between U and $\neg U$. Here too, we see the coming-together of topology and logic.

> *Further reading: For a coffee-break introduction to topos theory, try 'Topos theory in a nutshell' from the great John Baez:*
>
> *http://math.ucr.edu/home/baez/topos.html*
>
> *Lawvere and Rosebrugh's book* Sets for Mathematics *has lots on variable sets, and Lawvere and Schanuel's* Conceptual Mathematics *introduces toposes right at the end. For something more advanced, but well-designed and user-friendly, try Mac Lane and Moerdijk's* Sheaves in Geometry and Logic: a first introduction to topos theory.

⋆ ⋆ ⋆ ⋆ ⋆