RUDIMENTARY CATEGORY THEORY NOTES

GREG STEVENSON

Contents

1. Categories

We'll give a somewhat informal definition of what a category is (but which is rigorous) and then give a more 'sets with structure' style definition (or perhaps just more pedantic).

Remark 1.1. There are some foundational issues involved in category theory, particularly to do with what is a set and what is 'too big' to be a set (think Russell's paradox). Eventually one has to confront these issues to some extent, but the emphasis there is on *eventually* and we will systematically avoid them here. We will say something is a *set* or *small* when it lies within the usual set theory and use words like *collection* when we don't wish to specify.

We will also make frequent and tacit use of the axiom of choice and its various equivalent statements, sometimes for 'collections' rather than just sets. This is not, ultimately, an issue for us.

Definition 1.2. A category C consists of a collection of *objects* $\operatorname{Ob}\mathsf{C}$ (we will often write $x \in \mathsf{C}$ rather than $x \in \operatorname{Ob}\mathsf{C}$), and for each pair of objects $x, y \in \operatorname{Ob}\mathsf{C}$ a set of *morphisms* $\mathsf{C}(x, y)$ (also sometimes denoted by $\operatorname{Hom}(x, y)$ or $\operatorname{Hom}_{\mathsf{C}}(x, y)$). We will often depict a morphism $f \in \mathsf{C}(x, y)$ as $f\colon x \longrightarrow y$ or $x \xrightarrow{f} y$.

In addition, for each object $x \in \mathsf{C}$ there is a distinguished morphism $1_x \in \mathsf{C}(x, x)$ called *the identity morphism of* x or *unit of* x and for each ordered triple of objects x, y, z there is a *composition* map (i.e. a function between sets)

$$\mathsf{C}(y, z) \times \mathsf{C}(x, y) \longrightarrow \mathsf{C}(x, z) \quad (g, f) \mapsto g \circ f \text{ (or just } gf \text{ for short)},$$

such that:

(1) composition is associative in the sense that given objects w, x, y, z and maps
$$w \xrightarrow{f} x \xrightarrow{g} y \xrightarrow{h} z$$
the two morphisms we can construct using composition, namely $(h \circ g) \circ f$ and $h \circ (g \circ f)$, are equal;

(2) composition is compatible with the identities in the sense that given $f \in \mathsf{C}(x, y)$ we have
$$1_y \circ f = f = f \circ 1_x.$$

In particular, if $x \in \mathsf{C}$ then $\mathsf{C}(x, x)$ (sometimes written $\mathrm{End}(x)$ where End is for endomorphism) is a *monoid*, i.e. a set equipped with a unital and associative binary operation
$$\mathsf{C}(x, x) \times \mathsf{C}(x, x) \longrightarrow \mathsf{C}(x, x).$$

As a point of notation we will call the $\mathsf{C}(x, y)$'s *hom-sets* or the *homs*; this is short hand for homomorphisms, the terminology being inspired by algebra where one speaks of homomorphisms of groups, rings, etc. We will call $f \colon x \longrightarrow y$ by many names, but usually we will say f is a *map* or *morphism*.

Now let us give the more 'formal' definition of a category.

Definition 1.3. A category C is a tuple $\mathsf{C} = (\mathrm{Ob}\,\mathsf{C}, \mathrm{Mor}\,\mathsf{C}, s, t, \mathrm{id}, \circ)$ where:

(i) $\mathrm{Ob}\,\mathsf{C}$ is a collection, called the *objects* of C

(ii) $\mathrm{Mor}\,\mathsf{C}$ is a collection, called the *morphisms* of C

(iii) s and t, the *source* and *target*, are functions $\mathrm{Mor}\,\mathsf{C} \longrightarrow \mathrm{Ob}\,\mathsf{C}$. We define subsets of $\mathrm{Mor}\,\mathsf{C}$
$$\mathsf{C}(x, y) = \{f \in \mathrm{Mor}\,\mathsf{C} \mid s(f) = x \text{ and } t(f) = y\}$$

(iv) id, the *identity* is a function $\mathrm{id} \colon \mathrm{Ob}\,\mathsf{C} \longrightarrow \mathrm{Mor}\,\mathsf{C}$ and we set $\mathrm{id}(x) = 1_x$

(v) $\circ$, the composition, is a function
$$\{(g, f) \in \mathrm{Mor}\,\mathsf{C} \times \mathrm{Mor}\,\mathsf{C} \mid t(f) = s(g)\} \longrightarrow \mathrm{Mor}\,\mathsf{C} \quad (g, f) \mapsto g \circ f$$
from composable morphisms to morphisms (we will sometimes write $\circ$ as a function on all pairs of morphisms—this is strictly nonsense since it is only partially defined, but it is convenient).

This data is subject to the following compatibility conditions:

(1) composition is associative in the sense that the two composites
$$\begin{array}{c} \{(h', g', f') \in \mathrm{Mor}\,\mathsf{C} \times \mathrm{Mor}\,\mathsf{C} \times \mathrm{Mor}\,\mathsf{C} \mid s(h') = t(g'), s(g') = t(f')\} \\ {\scriptstyle 1\times\circ}\downarrow\;\downarrow{\scriptstyle \circ\times 1} \\ \{(g, f) \in \mathrm{Mor}\,\mathsf{C} \times \mathrm{Mor}\,\mathsf{C} \mid t(f) = s(g)\} \\ \downarrow{\scriptstyle \circ} \\ \mathrm{Mor}\,\mathsf{C} \end{array}$$
are equal, i.e. given (h, g, f) with $s(h) = t(g)$ and $s(g) = t(f)$ we have $h \circ (g \circ f) = (h \circ g) \circ f$;

(2) composition is unital in the sense that given $f \in \mathrm{Mor}\,\mathsf{C}$ with $s(f) = x$ and $t(x) = y$ we have
$$1_y \circ f = f = f \circ 1_x.$$

Remark 1.4. One thing to take away from the very formal nature of this definition is that a category consists of some *objects* and some *morphisms* and some operations on them and nothing more. By nothing more we mean that there is no assumption that the objects or maps have any structure: the objects need not be sets with structure, they are just labels, and similarly the morphisms need not be functions in any traditional sense.

The guiding principle is that how objects relate to one another via morphisms is the essential thing to consider. Rather than understanding an object x by some 'internal structure' one should understand x in terms of how other objects map to it.

Remark 1.5. It is possible to see how to move from the first definition to the second as follows: first define

$$\operatorname{Mor} \mathsf{C} = \coprod_{x,y \in \operatorname{Ob} \mathsf{C}} \mathsf{C}(x,y)$$

to be the disjoint union of the hom-sets in C to get the collection of all morphisms. The composition $\circ$ is extended in the obvious way, using that the cartesian product commutes with disjoint unions. Similarly the source and target are defined by extending the constant maps

$$\mathsf{C}(x,y) \longrightarrow \{x\} \longrightarrow \operatorname{Ob} \mathsf{C} \quad \text{and} \quad \mathsf{C}(x,y) \longrightarrow \{y\} \longrightarrow \operatorname{Ob} \mathsf{C}$$

respectively. Finally, id is also defined in the evident way.

2. First examples

We now give several examples of categories, some fundamental and which will occur repeatedly, and some more of a 'hey look, this thing is a category' nature. We begin with some categories that we will, at least implicitly, deal with a lot.

Example 2.1. There is a category Set of sets: the objects are sets and the morphisms are functions between sets with the usual composition and identities. We can also consider $\mathsf{Set}^{\mathrm{f}}$ the category of finite sets, whose objects are the sets with finitely many elements. Intuitively, we see that $\mathsf{Set}^{\mathrm{f}}$ 'sits inside' Set and we will make this precise in Section 4.

Example 2.2. There is a category Grp of groups: the objects are groups and the morphisms are group homomorphisms, i.e. functions that respect group multiplication and the identity element. As above we can also consider $\mathsf{Grp}^{\mathrm{f}}$ the category of finite groups.

Example 2.3. Given a unital ring R we can consider $\mathsf{Mod}\, R$ the category of right R-modules. This is an *abelian category*, and behaves in many senses quite differently from Set and Grp.

Example 2.4. There is a category Top of topological spaces: the objects are topological spaces and the morphisms are continuous maps of spaces. We get many variants by imposing additional properties on the spaces we allow, for instance we could consider the category of compact Hausdorff spaces, or the category of metric spaces.

We can make new categories from old ones by imposing some additional structure and requiring morphisms to preserve that structure. For instance:

Example 2.5. A pointed set (X, x) is a set X with a distinguished element $x \in X$. A morphism of pointed sets $f\colon (X,x) \longrightarrow (Y,y)$ is a function $f\colon X \longrightarrow Y$ such that $f(x) = y$, i.e. it respects the distinguished elements. We denote by Set_* the corresponding category of pointed sets and morphisms of pointed sets. Similarly there is a category Top_* of pointed or based topological spaces and basepoint preserving continuous maps.

Or we can make new categories via *duality*.

Example 2.6. Let C be a category. We define a new category C^{op} called *the opposite category of* C by

$$\mathrm{Ob}\,\mathsf{C}^{\mathrm{op}} = \mathrm{Ob}\,\mathsf{C}$$
$$\mathsf{C}^{\mathrm{op}}(x, y) = \mathsf{C}(y, x)$$

that is, we think of maps as going in the opposite direction. The units in C^{op} are the same as those in C and composition is defined by

$$\mathsf{C}^{\mathrm{op}}(y, z) \times \mathsf{C}^{\mathrm{op}}(x, y) \longrightarrow \mathsf{C}^{\mathrm{op}}(x, z) \quad (g, f) \mapsto fg$$

so we remember that we can compose f and g in C (in the opposite order) and define their composite in C^{op} to be the honest composite in C but again viewed as going in the opposite direction.

We can also consider smaller categories of a rather different flavour.

Example 2.7. Suppose that G is a group. Then there is a category BG with

$$\mathrm{Ob}\,BG = \{*\}$$
$$BG(*, *) = G$$

composition $G \times G \longrightarrow G$ given by the group operation of G and identity the identity element of G. Thus we can view any group as a category with a single object.

We can also view a group as a category in another way.

Example 2.8. Suppose that G is a group. There is a category EG with

$$\mathrm{Ob}\,EG = G$$
$$EG(g, h) = \{g^{-1}h\}$$

that is, the objects of EG are the group elements and the maps from g to h are precisely the elements x such that $gx = h$ of which there is precisely one. Thus there is a unique map between any two elements of EG.

Both BG and EG have the property than any morphism is an *isomorphism* i.e., for every $f \colon x \longrightarrow y$ there is an $f^{-1} \colon y \longrightarrow x$ such that $f^{-1}f = 1_x$ and $ff^{-1} = 1_y$. We call a category with this property a *groupoid*, and one can think of such a category as a 'group with several objects'. If C is a groupoid then for every object $x \in \mathsf{C}$ we have a group $\mathsf{C}(x, x)$, and if $\mathsf{C}(x, y)$ is not empty then every $f \in \mathsf{C}(x, y)$ gives an isomorphism of groups

$$\mathsf{C}(x, x) \longrightarrow \mathsf{C}(y, y) \quad g \mapsto f \circ g \circ f^{-1}$$

Example 2.9. We can generalise the previous example as follows. Suppose that X is a set and G acts on X via

$$\rho \colon X \times G \longrightarrow X$$

i.e. X is a (right) G-set, where we write $\rho(x, g) = x \cdot g$. The *action groupoid* $X//G$ is the category with

$$\mathrm{Ob}\,X//G = X$$
$$X//G(x, x') = \{g \in G \mid x \cdot g = x'\}$$

so the maps can be represented as pairs $(x, g) \colon x \longrightarrow x \cdot g$. The identity element of G provides units, and composition is defined by

$$(x \cdot g \xrightarrow{h} x \cdot g \cdot h) \circ (x \xrightarrow{g} x \cdot g) = (x \xrightarrow{gh} x \cdot gh)$$

which is well defined by associativity of the action of G.

We can think of $X//G$ as a refined version of the quotient X/G, where we keep track of orbits and stabilizers.

Now let us give another class of examples.

Example 2.10. Let X be a topological space. We define a category $\mathcal{O}(X)$ with object set

$$\operatorname{Ob}\mathcal{O}(X) = \{U \subseteq X \mid U \text{ is open in } X\}.$$

We set

$$\mathcal{O}(U,V) = \begin{cases} \{*\} & \text{if } U \subseteq V \\ \varnothing & \text{else} \end{cases}$$

Thus $\mathcal{O}(X)$ has the open subsets of X as objects and the morphisms record inclusion of open subsets—this is precisely the poset of open sets of X.

Example 2.11. More generally, if $(P, \leq)$ is a poset then we can view P as a category with object set P and with a map from x to y precisely if $x \leq y$. Conversely, if C is a category with at most one morphism between any two objects (one says C is *thin*) then it is more or less a poset (it is a poset up to equivalence, see Section 6).

We finish with one last, somewhat innocuous but very important, example which will appear again later.

Example 2.12. We let Δ be the category with objects the sets

$$[n] = \{0, 1, \ldots, n\}$$

and morphisms the weakly order preserving functions, i.e. $\Delta([m],[n])$ consists of those functions f such that $i \leq j$ implies $f(i) \leq f(j)$.

3. Properties of morphisms

We now briefly describe three important properties that morphisms in a category can have and give some examples. Let C be some fixed category. We start with the strongest notion (which was already introduced after Example 2.8).

Definition 3.1. We say that $f\colon x \longrightarrow y$ is an *isomorphism* (or *is invertible*) if there is a morphism $f^{-1}\colon y \longrightarrow x$ such that

$$f \circ f^{-1} = 1_y \quad \text{and} \quad f^{-1} \circ f = 1_x.$$

In this case f^{-1} is unique and is called the *inverse* of f.

Now let us consider a generalization of injective and surjective functions to arbitrary categories. We begin with the former.

Definition 3.2. We say that $f\colon x \longrightarrow y$ is a *monomorphism* (or *mono* or is *monic*) if it is *left cancellable* in the sense that given any diagram

$$w \underset{g_2}{\overset{g_1}{\rightrightarrows}} x \xrightarrow{f} y$$

i.e. a pair of maps $w \longrightarrow x$, then $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$.

Dually we say that $f\colon x \longrightarrow y$ is an *epimorphism* (or *epi* or is *epic*) if it is *right cancellable* in the sense that given any diagram

$$x \xrightarrow{f} y \underset{h_2}{\overset{h_1}{\rightrightarrows}} z$$

with $h_1 \circ f = h_2 \circ f$ we have $h_1 = h_2$.

Remark 3.3. What we mean by 'dually' above is the following: one obtains the definition of an epimorphism by reversing all the arrows in the definition of a monomorphism. Put another way, an epimorphism in C is precisely a monomorphism in C^{op}. This will be a recurring theme—concepts in category theory have dual concepts. The nomenclature often reflects this by declaring some gadget which satisfies the properties of a *thing* in C^{op} to be a *co*thing; unfortunately this is not such a case, but we will see plenty of them.

Note that isomorphisms are automatically monomorphisms and epimorphisms, by virtue of having a left and right inverse.

Let us give some examples, harking back to the categories we have seen in the last section.

Example 3.4. In Set the monomorphisms are precisely the injective maps of sets, the epimorphisms are the surjective set maps, and the isomorphisms are the bijections.

Example 3.5. In Top the monomorphisms and epimorphisms are the continuous injective and surjective maps respectively. The isomorphisms are the homeomorphisms. Note that a map can be monic and epic without being an isomorphism: a continuous bijection is epic and monic, but can fail to have a continuous inverse.

4. Functors

We now discuss the natural notion of map between categories. Let us fix categories C and D.

Definition 4.1. A *functor* $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ consists of a function $F\colon \operatorname{Ob}\mathsf{C} \longrightarrow \operatorname{Ob}\mathsf{D}$ and for each $x, y \in \mathsf{C}$ a map of sets

$$F_{x,y}\colon \mathsf{C}(x,y) \longrightarrow \mathsf{D}(Fx, Fy)$$

such that:

(i) F is compatible with units in the sense that $F_{x,x}(1_x) = 1_{Fx}$;

(ii) F is compatible with composition in the sense that the square

$$\begin{array}{ccc} \mathsf{C}(y,z) \times \mathsf{C}(x,y) & \xrightarrow{\circ_{\mathsf{C}}} & \mathsf{C}(x,z) \\ {\scriptstyle F_{y,z}\times F_{x,y}}\downarrow & & \downarrow{\scriptstyle F_{x,z}} \\ \mathsf{D}(Fy,Fz) \times \mathsf{D}(Fx,Fy) & \xrightarrow[\circ_{\mathsf{D}}]{} & \mathsf{D}(Fx,Fz) \end{array}$$

commutes, i.e. $F_{x,z}\circ_{\mathsf{C}} = \circ_{\mathsf{D}}(F_{y,z} \times F_{x,y})$. More concretely, this says that given composable maps

$$x \xrightarrow{f} y \xrightarrow{g} z$$

we have $F(g \circ_{\mathsf{C}} f) = F(g) \circ_{\mathsf{D}} F(f)$.

Remark 4.2. For the sake of clarity we have distinguished the compositions in C and D in the definition by using $\circ_{\mathsf{C}}$ and $\circ_{\mathsf{D}}$ respectively. This will not be the norm, and we will usually omit all notation and just use juxtaposition to denote composition.

Remark 4.3. The notion of a diagram *commuting*, which we have informally introduced during the above definition, means that any two ways of navigating the diagram between a fixed start and end point agree. It is of utmost importance and, for instance, gives an incredibly convenient way of packaging equations that might otherwise be difficult to read.

Remark 4.4. The statement that F preserves units can also be expressed diagrammatically by commutativity of

$$\begin{CD} \mathrm{Ob}\,\mathsf{C} @>{\mathrm{id}_{\mathsf{C}}}>> \mathrm{Mor}\,\mathsf{C} \\ @V{F}VV @VV{F}V \\ \mathrm{Ob}\,\mathsf{D} @>>{\mathrm{id}_{\mathsf{D}}}> \mathrm{Mor}\,\mathsf{D} \end{CD}$$

Remark 4.5. Since F preserves composition and units it must send isomorphisms to isomorphisms: indeed, if $f\colon x \longrightarrow y$ is an isomorphism with inverse f^{-1} we have

$$F(f)F(f^{-1}) = F(ff^{-1}) = F(1_y) = 1_{Fy}$$

and similarly for the other composite.

To summarise: a functor is just an assignment on objects and maps which preserves all the structure defining a category. Let us now give some examples and further definitions.

Example 4.6. If C is a category then we can consider the identity functor $\mathrm{id}_{\mathsf{C}}\colon \mathsf{C} \longrightarrow \mathsf{C}$ which is the identity on both objects and morphisms. This is a surprisingly important functor.

Example 4.7. There is a canonical functor $\mathsf{Set}^{\mathrm{f}} \longrightarrow \mathsf{Set}$, from the category of finite sets to the category of sets, which is the identity on objects and morphisms. Evidently doing nothing preserves identities and compositions. Similarly, there is a canonical functor $\mathsf{Grp}^{\mathrm{f}} \longrightarrow \mathsf{Grp}$.

Definition 4.8. We say a functor $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ is *faithful* if the maps on morphisms $F_{x,y}$ are injective for all $x, y \in \mathsf{C}$. We say F is *full* if the $F_{x,y}$ are all surjective. Finally, we say F is fully faithful if it is both full and faithful, i.e. all the $F_{x,y}$ are bijective.

Example 4.9. In the last example both $\mathsf{Set}^{\mathrm{f}} \longrightarrow \mathsf{Set}$ and $\mathsf{Grp}^{\mathrm{f}} \longrightarrow \mathsf{Grp}$ are fully faithful. We say that they are *full subcategories*.

Example 4.10. We can consider the category $\mathsf{Set}^{\mathrm{m}}$ of sets where the morphisms are the injective maps, i.e. the monomorphisms. A composite of monomorphisms is a monomorphism and identity maps are monomorphisms, so this is indeed a category and we see the inclusion $\mathsf{Set}^{\mathrm{m}} \longrightarrow \mathsf{Set}$ is a faithful functor. We say it is a *subcategory* of Set.

Similarly, we can regard the category Δ from Example 2.12 as a subcategory of $\mathsf{Set}^{\mathrm{f}}$.

Now let us give a couple of more exciting examples.

Example 4.11. There is a *forgetful functor* $U\colon \mathsf{Grp} \longrightarrow \mathsf{Set}$ which sends a group G to its underlying set UG and a map to the underlying function; in other words, U takes a group and forgets that it has a group structure. This functor is faithful, but it is not full—not every map of the underlying sets is a map of groups.

Example 4.12. There is a *free functor* $F\colon \mathsf{Set} \longrightarrow \mathsf{Grp}$ which sends a set X to the free group FX with generating set X and a map $f\colon X \longrightarrow Y$ to the unique map $FX \longrightarrow FY$ which is obtained by extending f to words in X. The functor F is also faithful, but again it is not full. We shall see in Section 7 that the forgetful and free functors are intimately related.

In a similar vein, there are forgetful and free functors between Set and Set_*, Set and Top, and Top and Top_*, given, for instance, by forgetting the basepoint or the topological space structure as appropriate.

Example 4.13. Let G and H be groups. Then any homomorphism $f\colon G \longrightarrow H$ defines a functor $Bf\colon BG \longrightarrow BH$. Conversely, any functor $BG \longrightarrow BH$ is uniquely determined by a group homomorphism $G \longrightarrow H$. Moreover, sending f to Bf is compatible with identities and compositions. In fact this gives an isomorphism of categories from Grp to the category of one object groupoids and functors between them.

Here are two extremely important (dual) examples.

Example 4.14. Let C be a category and x an object of C. The *functor represented by x* is

$$\mathsf{C}(-,x)\colon \mathsf{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}, \quad y \mapsto \mathsf{C}(y,x).$$

Let us unwind what this does: it acts on objects as indicated and sends a morphism $f\colon y \longrightarrow y'$ in C to

$$\mathsf{C}(f,x)\colon \mathsf{C}(y',x) \longrightarrow \mathsf{C}(y,x), \quad (y' \xrightarrow{g} x) \mapsto (y \xrightarrow{f} y' \xrightarrow{g} x)$$

i.e. $\mathsf{C}(f,x)$ acts on morphisms by precomposition with f

$$\mathsf{C}(-,x)_{y,y'}\colon \mathsf{C}(y,y') \longrightarrow \mathsf{Set}(\mathsf{C}(y',x),\mathsf{C}(y,x)), \quad f \mapsto (-)\circ f$$

Dually, there is a *functor corepresented by x*

$$\mathsf{C}(x,-)\colon \mathsf{C} \longrightarrow \mathsf{Set}$$

which sends $f\colon y \longrightarrow y'$ to the map $\mathsf{C}(x,y) \longrightarrow \mathsf{C}(x,y')$ given by postcomposition with f, i.e. $g \in \mathsf{C}(x,y)$ is sent to fg.

Thus we can think of $\mathsf{C}(-,-)$ as a *functor of two variables* or *bifunctor*. In fact, it is a functor

$$\mathsf{C}^{\mathrm{op}} \times \mathsf{C} \longrightarrow \mathsf{Set}$$

where $\times$ means we take the cartesian product of the collections of objects and take the maps from (x,y) to (x',y') to be

$$\mathsf{C}(x,x') \times \mathsf{C}(y,y').$$

And as an important special case of the last example:

Example 4.15. Let k be a field and let $\mathsf{Mod}\,k$ be the category of k-vector spaces. Given any vector space V the dual

$$V^* = \mathrm{Hom}_k(V,k) = \mathsf{Mod}\,k(V,k)$$

again has the structure of a vector space. Thus the functor represented by k can be viewed as again landing in vector spaces:

$$(-)^* = \mathrm{Hom}_k(-,k)\colon\ \mathsf{Mod}\,k^{\mathrm{op}} \longrightarrow \mathsf{Mod}\,k.$$

We end this section with two examples of particular relevance to us.

Example 4.16. Recall that the fundamental group of a pointed space $(X,x_0) \in \mathsf{Top}_*$, denoted $\pi_1(X,x_0)$, is the group of homotopy classes of loops in X with base point x_0, i.e. homotopy classes of maps $(S^1,*) \longrightarrow (X,x_0)$. Taking the fundamental group defines a functor $\mathsf{Top}_* \longrightarrow \mathsf{Grp}$, the action on maps being given by postcomposition: given $f\colon (X,x_0) \longrightarrow (Y,y_0)$ the map $\pi_1(f)$ sends the class of a loop $\gamma\colon (S^1,*) \longrightarrow (X,x_0)$ to the class $[f\gamma]$.

In fact, this is an example of a corepresentable functor. We could define $\mathrm{Ho}(\mathsf{Top}_*)$ to be the category whose objects are pointed spaces and whose morphisms are (based) homotopy classes of continuous base point preserving maps. Then $\pi_1(-) = \mathrm{Ho}(\mathsf{Top}_*)((S^1, *), -)$.

Remark 4.17. It is worth pointing out that $\mathrm{Ho}(\mathsf{Top}_*)$ is, in many senses (which we won't make precise here), a fairly grim example of a category. In practice one restricts to some better behaved subcategory of topological spaces.

Example 4.18. Let Δ^n denote the standard n-simplex which is defined to be

$$\Delta^n = \{x \in \mathbb{R}^{n+1} \mid \sum_{i=0}^{n} x_i = 1 \text{ and } \forall i \; x_i \geq 0\} \subseteq \mathbb{R}^{n+1},$$

equipped with the subspace topology, where $\mathbb{R}^{n+1}$ has the usual Euclidean topology. Thus $\Delta^0 = \{1\}$ is a point, Δ^1 is the straight line segment between $(0,1)$ and $(1,0)$ i.e. a closed interval of length 1, Δ^2 is a filled triangle, Δ^3 a tetrahedron, and so on. We note that Δ^n has dimension n.

We can define a functor $\Delta \longrightarrow \mathsf{Top}$, where Δ is the (suggestively named) category from Example 2.12. The object assignment is given by sending $[n]$ to Δ^n and a map $f\colon [m] \longrightarrow [n]$ (which we recall is a weakly increasing function) is sent to the map $\Delta(f)\colon \Delta^m \longrightarrow \Delta^n$ defined by

$$\Delta(f)(x_0, \ldots, x_m) = (y_0, \ldots, y_n), \text{ where } y_i = \sum_{j \in f^{-1}(i)} x_j.$$

One can check that this functor is faithful, and so exhibits Δ (strictly speaking its image under our functor) as a subcategory of Top.

5. Natural transformations

So far we have considered categories, and functors, which are the maps between categories. We next consider maps between functors i.e. the maps between maps.

Definition 5.1. Given a pair of functors $F, G\colon \mathsf{C} \longrightarrow \mathsf{D}$ a *natural transformation* $\alpha\colon F \longrightarrow G$ is a collection

$$\{\alpha_x\colon Fx \longrightarrow Gx \mid x \in \mathrm{Ob}\,\mathsf{C}\}$$

of maps in D, one for each $x \in \mathrm{Ob}\,\mathsf{C}$, such that for each $f\colon x \longrightarrow y$ in C the square

$$\begin{array}{ccc} Fx & \xrightarrow{F(f)} & Fy \\ {\scriptstyle\alpha_x}\downarrow & & \downarrow{\scriptstyle\alpha_y} \\ Gx & \xrightarrow[G(f)]{} & Gy \end{array}$$

commutes. We refer to this condition as *naturality* and we call the α_x the *components* of α.

Again we illustrate this concept with several examples, some more exotic than others.

Example 5.2. The first example of a natural transformation is the identity natural transformation. Given any functor $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ there is a natural transformation $\mathrm{id}_F\colon F \longrightarrow F$ with component at x the map 1_{Fx}.

Definition 5.3. We say $\alpha\colon F \longrightarrow G$ is a *natural isomorphism* and write $\alpha\colon F \xrightarrow{\sim} G$ (or just $F \cong G$ if α is clear or we only care F and G are isomorphic) if α is an isomorphism in the sense that it has a two-sided inverse α^{-1} (which makes sense, as we have seen above there are identity natural transformations).

One can check this is equivalent to each component of α being an isomorphism.

Example 5.4. Consider the forgetful and free functors $U\colon \mathsf{Grp} \longrightarrow \mathsf{Set}$ and $F\colon \mathsf{Set} \longrightarrow \mathsf{Grp}$ from Example 4.11 and 4.12. Given a group G we can consider the map

$$\varepsilon_G\colon FU(G) \longrightarrow G$$

determined by extending the assignment $g \mapsto g$ to arbitrary elements of $FU(G)$, i.e. sending words in G (which are the elements of $FU(G)$) to the corresponding product of elements in G. One can check (it is close to tautological) that the ε_G assemble as the components of a natural transformation $\varepsilon\colon FU \longrightarrow \mathrm{id}_{\mathsf{Grp}}$.

Given a set X we can consider the map

$$\eta_X\colon X \longrightarrow UF(X)$$

which includes X into the free group it generates in the obvious way. Again, more or less tautologically this determines a natural transformation $\eta\colon \mathrm{id}_{\mathsf{Set}} \longrightarrow UF$.

Example 5.5. Let R be a ring and consider the identity functor $\mathrm{id}_{\mathsf{Mod}\,R}$ of $\mathsf{Mod}\,R$, the category of (right) R-modules. If $r \in R$ is central in R, i.e.

$$sr = rs \quad \forall s \in R,$$

then r determines a natural transformation ρ^r from $\mathrm{id}_{\mathsf{Mod}\,R}$ to itself. Indeed, for $M \in \mathsf{Mod}\,R$ we define $\rho^r_M\colon M \longrightarrow M$ by right multiplication by r, which is R-linear since for $s \in R$ and $m \in M$ we have

$$\rho^r_M(m) \cdot s = (m \cdot r) \cdot s = m \cdot (rs) = m \cdot (sr) = (m \cdot s) \cdot r = \rho^r_M(m \cdot s).$$

To check naturality, given $f\colon M \longrightarrow N$ we need to check that

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle \rho^r_M}\downarrow & & \downarrow{\scriptstyle \rho^r_N} \\ M & \xrightarrow[f]{} & N \end{array}$$

commutes, i.e. that

$$f(m) \cdot r = f(m \cdot r)$$

for all $m \in M$ which is just the fact that f is R-linear.

Example 5.6. Given a small category C and another category D we can form the *functor category* $[\mathsf{C}, \mathsf{D}]$ whose objects are the functors $\mathsf{C} \longrightarrow \mathsf{D}$ and whose morphisms are the natural transformations, i.e. for $F, G\colon \mathsf{C} \longrightarrow \mathsf{D}$

$$[\mathsf{C}, \mathsf{D}](F, G) = \{\alpha\colon F \longrightarrow G \mid \alpha \text{ is a natural transformation}\}.$$

The reason we emphasize that C should be *small*, i.e. have a set of objects, is that otherwise we run into 'size issues', for instance there may be too many natural transformations between two given functors to form a set.

Let us consider a couple of special cases of the last example.

Example 5.7. Let G be a group and consider the category BG of Example 2.7. We form, as in the previous example, the category $[BG^{\mathrm{op}}, \mathsf{Set}]$ of functors from BG to Set. Let us unravel what this category is. A functor $X\colon BG^{\mathrm{op}} \longrightarrow \mathsf{Set}$ picks out a set $X(*)$ and for each $g \in G$ an automorphism $X(g)$ of $X(*)$. This is subject to the compatibility conditions that $X(1_G)$ is the identity on $X(*)$ and for $g, h \in G$ we have

$$X(gh) = X(h)X(g).$$

Thus X determines a (right) G-set structure on $X(*)$. Conversely, one can check that any G-set determines a functor $BG^{\mathrm{op}} \longrightarrow X$ in the obvious way.

Now let us consider the morphisms. Given $X, Y \in [BG^{\mathrm{op}}, \mathsf{Set}]$ a natural transformation $\phi\colon X \longrightarrow Y$ consists of a single map of sets, which we also call ϕ, $X(*) \longrightarrow Y(*)$ such that for every $g \in G$ the square

$$\begin{array}{ccc} X(*) & \xrightarrow{X(g)} & X(*) \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\phi} \\ Y(*) & \xrightarrow[Y(g)]{} & Y(*) \end{array}$$

commutes. This is precisely the condition that ϕ commutes with the action of G on $X(*)$ and $Y(*)$, i.e. that it is a map of G-sets. Thus, we see that $[BG^{\mathrm{op}}, \mathsf{Set}]$ can be identified with the category of G-sets (in the sense that these categories are isomorphic, we'll treat this more carefully in the next section).

Example 5.8. Let us again consider the category Δ of Example 2.12. A functor $\Delta^{\mathrm{op}} \longrightarrow \mathsf{C}$ is called a *simplicial object in* C. For instance, we can consider the functor category $[\Delta^{\mathrm{op}}, \mathsf{Set}]$ which is called the *category of simplicial sets* and denoted sSet. This category is of great importance in homotopy theory and homological algebra and will make an appearance (or at least loom menacingly in the background) when we come to defining (co)homology of spaces.

Example 5.9. Recall from Example 4.14 that given a category C and $x \in \mathsf{C}$ we get a functor represented by x

$$\mathsf{C}(-, x)\colon \mathsf{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}.$$

Given $f\colon x \longrightarrow y$ we can define a natural transformation

$$\mathsf{C}(-, f)\colon \mathsf{C}(-, x) \longrightarrow \mathsf{C}(-, y)$$

by specifying that the component at v is

$$\mathsf{C}(v, f)\colon \mathsf{C}(v, x) \longrightarrow \mathsf{C}(v, y) \quad (v \xrightarrow{g} x) \mapsto (v \xrightarrow{g} x \xrightarrow{f} y),$$

i.e. it is given by postcomposition with f, sending g to fg. Naturality can be seen by considering, for $e\colon v \longrightarrow w$, the square

$$\begin{array}{ccc} \mathsf{C}(w, x) & \xrightarrow{\mathsf{C}(w,f)} & \mathsf{C}(w, y) \\ {\scriptstyle\mathsf{C}(e,x)}\downarrow & & \downarrow{\scriptstyle\mathsf{C}(e,y)} \\ \mathsf{C}(v, x) & \xrightarrow[\mathsf{C}(v,f)]{} & \mathsf{C}(v, y) \end{array}$$

which sends a map $g\colon w \longrightarrow x$ to

$$\mathsf{C}(e, y)\mathsf{C}(w, f)(g) = (fg)e \text{ and } \mathsf{C}(v, f)\mathsf{C}(e, x)(g) = f(ge)$$

respectively.

Thus we have assignments

$$\operatorname{Ob}\mathsf{C} \longrightarrow \operatorname{Ob}[\mathsf{C}^{\mathrm{op}}, \mathsf{Set}] \quad x \mapsto \mathsf{C}(-,x)$$
$$\mathsf{C}(x,y) \longrightarrow [\mathsf{C}^{\mathrm{op}}, \mathsf{Set}](\mathsf{C}(-,x), \mathsf{C}(-,y)) \quad f \mapsto \mathsf{C}(-,f)$$

defining a functor $\mathsf{C} \longrightarrow [\mathsf{C}^{\mathrm{op}}, \mathsf{Set}]$. This functor is often called the *Yoneda functor* and it is fully faithful. This latter claim is the statement of the *Yoneda lemma*, which is not (for the moment) discussed in these notes.

6. Equivalences

Having discussed categories, functors, and natural transformations, we are now in a position to begin discussing, in a precise sense, relations between various categories. We begin, in this section, with a discussion of when two categories are 'the same'. This turns out, due to the added flexibility inherent in categories, to be more subtle than the notion of isomorphism of groups or homeomorphism of spaces (it is more akin to homotopies of spaces). However, there is a notion of isomorphism of categories (using the abstract definition of isomorphism in the category whose objects are categories and whose maps are functors) and we begin with that.

Definition 6.1. We say categories C and D are *isomorphic* if there are functors $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ and $F^{-1}\colon \mathsf{D} \longrightarrow \mathsf{C}$ such that $F^{-1}F = 1_{\mathsf{C}}$ and $FF^{-1} = 1_D$, i.e. the two composites are the respective identity functors. In particular, this says that F and F^{-1} are inverse bijections on the collections of objects and on each hom-set.

This, at first glance, probably seems like a pretty reasonable notion. However, in practice it is much too strong. Let us try and illustrate this.

Example 6.2. Consider the category of sets Set and pick a set X. Define a full subcategory (a subcategory whose inclusion functor is fully faithful) Set' consisting of all sets not *equal* to X, i.e. we just remove the set X from Set. Does this really change the category of sets? There are many sets isomorphic to X (hence they have the same cardinality, same automorphism group, and map the same way to other sets) and they are objects of Set'. Thus Set' contains essentially the same information as Set, provided we don't want to refer literally to X we could do set theory in Set' with no issues at all.

Example 6.3. Let us consider a more extreme continuation of the above idea. Consider $\mathbb{R}$ the real numbers and $\mathsf{mod}\,\mathbb{R}$ the category of finite dimensional real vector spaces. Let C denote the full subcategory with objects $\mathbb{R}^n$ for $n \geq 1$. Then we know, since every vector space has a basis, that every object of $\mathsf{mod}\,\mathbb{R}$, i.e. every finite dimensional vector space, is isomorphic to some $\mathbb{R}^n$. So the category C 'determines' $\mathsf{mod}\,\mathbb{R}$ in the sense that all of the information about objects and maps is already contained in $\mathbb{R}^n$ up to isomorphism; indeed sticking to C is the approach one takes when first learning linear algebra and talking about matrices instead of abstract linear maps.

We would like a notion that captures this 'sameness' in the above examples and gives us an abstract way of conflating such categories.

Definition 6.4. A functor $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ is an *equivalence* if there is a functor F^{-1} (called a *quasi-inverse* for F) such that $F^{-1}F \cong 1_{\mathsf{C}}$ and $FF^{-1} \cong 1_{\mathsf{D}}$, i.e. we only ask for natural isomorphisms to the identities not equalities. In this case we will write $\mathsf{C} \cong \mathsf{D}$ and say C and D are equivalent.

Remark 6.5. This is traditionally the notation used for isomorphisms, but we use it for equivalences of categories since this is the more important notion. Isomorphisms of categories don't tend to arise naturally and it's not essential to distinguish isomorphisms from the more general concept of equivalence.

Thus for an equivalence we only require that there is an inverse up to natural isomorphism. In both of the above examples the inclusion of the full subcategory is an equivalence. However, it's not always possible to construct an explicit quasi-inverse—doing so in Example 6.3 requires the axiom of choice. Let us give a naturally occurring example where one can write down a quasi-inverse.

Example 6.6. Again consider the category of finite dimensional $\mathbb{R}$-vector spaces and the functor

$$(-)^* = \mathrm{Hom}_{\mathbb{R}}(-,\mathbb{R})\colon \mathsf{mod}\,\mathbb{R}^{\mathrm{op}} \longrightarrow \mathsf{mod}\,\mathbb{R}$$

given by taking the vector space dual. Then $(-)^*$ is an equivalence with inverse $(-)^*$ i.e. it squares to the identity up to isomorphism. Indeed, there is for each $V \in \mathsf{mod}\,\mathbb{R}$ an isomorphism

$$V \longrightarrow V^{**} \quad v \mapsto ((V \xrightarrow{f} k) \mapsto f(v))$$

which is natural and shows $1_{\mathsf{mod}\,\mathbb{R}} \cong (-)^{**}$. Thus $\mathsf{mod}\,\mathbb{R}$ is equivalent to $\mathsf{mod}\,\mathbb{R}^{\mathrm{op}}$.

Remark 6.7. An important warning is that quasi-inverses are not unique—they are only unique up to a unique natural isomorphism (this is often as unique as things can get in category theory).

In practice it can be very difficult to construct a quasi-inverse (due to the number of choices that need to be made). Moreover, one often doesn't always care about the quasi-inverse, but only that a given functor has one. There is a standard abstract criterion describing when a functor is an equivalence which is very useful in such situations. In order to state it we first need a definition.

Definition 6.8. Let $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ be a functor. We say F is *essentially surjective* if for every $d \in \mathsf{D}$ there is a $c \in \mathsf{C}$ with $Fc \cong d$, i.e. every object of D is in the image of F up to isomorphism.

The *essential image* of F is

$$\mathrm{im}\, F = \{d \in \mathsf{D} \mid \exists c \in \mathsf{C} \text{ with } Fc \cong d\},$$

which can be viewed as a full subcategory of D. Thus F is essentially surjective if and only if the essential image of F is D.

Proposition 6.9. *Let $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ be a functor. Then F is an equivalence if and only if F is fully faithful and essentially surjective.*

Proof. Suppose F is an equivalence with a quasi-inverse F^{-1} and $\alpha\colon F^{-1}F \xrightarrow{\sim} 1_{\mathsf{C}}$. If $d \in \mathsf{D}$ then, since we know $FF^{-1}d \cong d$, we see that d is in the essential image of F.

If $c, c' \in \mathsf{C}$ then we have a commutative diagram

$$\begin{array}{ccccc}
\mathsf{C}(c,c') & \longrightarrow & \mathsf{C}(Fc,Fc') & \longrightarrow & \mathsf{C}(F^{-1}Fc,F^{-1}Fc') \\
\Vert & & & & \downarrow{\scriptstyle \mathsf{C}(F^{-1}Fc,\alpha_{c'})} \\
\mathsf{C}(c,c') & & \xleftarrow[\mathsf{C}(\alpha_c^{-1},c')]{} & & \mathsf{C}(F^{-1}Fc,c')
\end{array}$$

where the labelled arrows are isomorphisms (since α is). Indeed, this commutes by naturality of α^{-1} which tells us that for $f\colon c \longrightarrow c'$ the square

$$\begin{CD} c @>{f}>> c' \\ @V{\alpha^{-1}}VV @VV{\alpha^{-1}}V \\ F^{-1}Fc @>>{F^{-1}F(f)}> F^{-1}Fc' \end{CD}$$

commutes. Equivalently, this says that $F^{-1}F(f) = \alpha f \alpha^{-1}$. It's then clear that F is fully faithful, the first commutative diagram giving an inverse to $F_{c,c'}$.

Now suppose that F is fully faithful and essentially surjective. Since F is essentially surjective we can, for each $d \in \mathsf{D}$, choose an object $F^{-1}d$ in C and an isomorphism $\phi_d\colon FF^{-1}d \longrightarrow d$. Given $f\colon d \longrightarrow d'$ there is a unique dashed map making the square

$$\begin{CD} d @>{f}>> d' \\ @V{\phi_d^{-1}}VV @VV{\phi_{d'}^{-1}}V \\ FF^{-1}d @>>> FF^{-1}d' \end{CD}$$

namely $\phi_{d'}^{-1} f \phi_d$. Since F is fully faithful there is a unique map $F^{-1}(f)\colon F^{-1}d \longrightarrow F^{-1}d'$ in C which F sends to $\phi_{d'}^{-1} f \phi_d$ and this defines the action of F^{-1} on morphisms. It's a good exercise to verify that F^{-1} respects units and composition. $\square$

Remark 6.10. The proof of the above involves a choice function on the collection of objects of D which some people find rather offensive from a set theory perspective.

It follows from the proposition that if C is a category and C' is a full subcategory containing at least one object from each isomorphism class in C, i.e. such that $\mathsf{C}' \longrightarrow \mathsf{C}$ is essentially surjective, then $\mathsf{C}' \cong \mathsf{C}$. For instance, we could pick one representative from each isomorphism class in C to obtain a category $\mathsf{sk}\,\mathsf{C}$ called a *skeleton* for C. The category $\mathsf{sk}\,\mathsf{C}$ is *skeletal* in the sense that if two objects of $\mathsf{sk}\,\mathsf{C}$ are isomorphic then they are equal. For instance $\{\mathbb{R}^n \mid n \geq 0\}$ is a skeleton for $\mathsf{mod}\,\mathbb{R}$.

7. Adjunctions

There are many situations in which a pair of functors

$$\begin{CD} \mathsf{C} \\ @V{F}VV @AA{G}A \\ \mathsf{D} \end{CD}$$

is not an equivalence but provides a very strong relation between the categories C and D. Now that we have the appropriate terminology let us develop this by more closely examining the functors introduced in Examples 4.11 and 4.12. Recall, from said examples, that we have a pair of functors

$$\begin{CD} \mathsf{Set} \\ @V{F}VV @AA{U}A \\ \mathsf{Grp} \end{CD}$$

called the free and forgetful functor. Given a set X and a group G let us think about how FX and UG behave with respect to morphisms.

To give a map $f\colon X \longrightarrow UG$ is just to pick for each element x of X a corresponding element $f(x) \in UG$. But UG is, as a set, just G and so f is just a collection of elements of G indexed by X. This defines a unique map of groups $f^\sharp\colon FX \longrightarrow G$, by sending a word $x_1 \cdots x_n$ in FX to the product $f(x_1)\cdots f(x_n)$. This construction is fairly easily seen to give a bijection: distinct maps of sets $X \longrightarrow UG$ give distinct maps of groups and since a map $FX \longrightarrow G$ is totally determined by its action on a generating set, say X, we see that every map is realised. Thus, we have uncovered isomorphisms

$$\mathsf{Set}(X, UG) \cong \mathsf{Grp}(FX, G)$$

for every set X and every group G. Moreover, one can easily confirm these isomorphisms are natural in both variables.

Now let us give another point of view on these isomorphisms. Given a set X we can consider the identity map $1_{FX}\colon FX \longrightarrow FX$. By the above discussion this corresponds to a morphism $\eta_X\colon X \longrightarrow UFX$, and this is a very reasonable morphism—it's just the inclusion of X into FX viewed as a set, identifying the canonical set of generators. It satisfies a *universal property* in the sense that any map $f\colon X \longrightarrow UG$ factors via UFX, i.e. we can always fill in to a commutative triangle

$$\begin{array}{ccc} X & \xrightarrow{\eta_X} & UFX \\ {\scriptstyle f}\downarrow & \swarrow{\scriptstyle Uf^\sharp} & \\ UG & & \end{array}$$

Thus we can think of UFX as the 'closest underlying set of a group to X' in some sense. This is one way to make precise that FX is 'free' on X. One can check that η is natural in X and so we have defined a natural transformation $\mathrm{id}_{\mathsf{Set}} \longrightarrow UF$.

Given a group G there is also a very appealing map $\varepsilon_G\colon FUG \longrightarrow G$, which just sends a word in G to its product in G. One can check that this corresponds, under our isomorphism $\mathsf{Set}(UG, UG) \cong \mathsf{Grp}(FUG, G)$, to the identity map on UG. One can think of this as a 'universal presentation' of G where FUG is G without the relations and the projection $FUG \longrightarrow G$ is killing the relations. Again this can be made precise in the sense that FUG is the 'best approximation of G by a free group'. Given any set X and a map $\pi\colon FX \longrightarrow G$ we can fill in to a commutative triangle

$$\begin{array}{ccc} FX & \xrightarrow{F\pi^\flat} & FUG \\ {\scriptstyle \pi}\downarrow & \swarrow{\scriptstyle \varepsilon_G} & \\ G & & \end{array}$$

where $\pi^\flat\colon X \longrightarrow UG$ is the map of sets underlying π restricted to X. Again ε is natural and so we have a natural transformation $FU \longrightarrow \mathrm{id}_{\mathsf{Grp}}$.

Our two maps also satisfy a pair of relations expressing their good behaviour under applications of F and G and making clearer the connection between η and ε. Consider the pair of composites

$$FX \xrightarrow{F(\eta_X)} FUFX \xrightarrow{\varepsilon_{FX}} FX \quad \text{and} \quad UG \xrightarrow{\eta_{UG}} UFUG \xrightarrow{U(\varepsilon_G)} UG$$

We claim they are both equal to their respective identity maps. For the first, one sees this by noting that the first map is just the inclusion of FX as a free generating set for $FUFX$

followed by the canonical projection, with the net effect of doing nothing. We leave a more detailed check and the verification of the second identity as an exercise. These identities are called the *triangle identities*.

Now let us show that these maps recover our bijections on hom-sets. We define a pair of maps as follows:

$$\mathsf{Set}(X, UG) \xrightarrow{\phi} \mathsf{Grp}(FX, G) \qquad f \mapsto \varepsilon_G F(f)$$

and

$$\mathsf{Grp}(FX, G) \xrightarrow{\psi} \mathsf{Set}(X, UG) \qquad g \mapsto U(g)\eta_X$$

which we claim are inverse to one another. One checks this by computing that for $f\colon X \longrightarrow UG$ we have

$$\begin{aligned} U(\varepsilon_G F(f))\eta_X &= U(\varepsilon_G)UF(f)\eta_X \\ &= U(\varepsilon_G)\eta_{UG} f \\ &= f \end{aligned}$$

where the first equality is functoriality, the second is naturality of η, and the third is one of the triangle identities.

To summarise, we have shown that the free and forgetful functor induce natural bijections

$$\mathsf{Set}(X, UG) \cong \mathsf{Grp}(FX, G)$$

and that the data of these bijections is equivalent to the natural transformations $\eta\colon \mathrm{id}_{\mathsf{Set}} \longrightarrow UF$ and $\varepsilon\colon FU \longrightarrow \mathrm{id}_{\mathsf{Grp}}$. Such a situation is called an *adjunction* and we will now give the formal definition.

Definition 7.1. Let C and D be categories and suppose we have a pair of functors

$$\begin{array}{c} \mathsf{C} \\ {\scriptstyle F}\big\downarrow \big\uparrow{\scriptstyle G} \\ \mathsf{D} \end{array}$$

We say that F *is left adjoint to* G if there are natural isomorphisms

$$\phi_{c,d}\colon \mathsf{D}(Fc, d) \xrightarrow{\sim} \mathsf{C}(c, Gd)$$

for all $c \in \mathsf{C}$ and $d \in \mathsf{D}$. In this situation we also say that G *is right adjoint to* F or that F and G are *adjoint*. We will sometimes write $F \dashv G$ to indicate that F is left adjoint to G and G is right adjoint to F.

Remark 7.2. To recapitulate: by the adjunction isomorphisms $\phi_{c,d}$ being natural we mean that given $f\colon c' \longrightarrow c$ and $g\colon d \longrightarrow d'$ the squares

$$\begin{array}{ccc} \mathsf{D}(Fc, d) & \xrightarrow{\phi_{c,d}} & \mathsf{C}(c, Gd) \\ {\scriptstyle \mathsf{D}(Ff', d)}\big\downarrow & & \big\downarrow{\scriptstyle \mathsf{C}(f, Gd)} \\ \mathsf{D}(Fc', d) & \xrightarrow[\phi_{c',d}]{} & \mathsf{C}(c', Gd) \end{array} \qquad \begin{array}{ccc} \mathsf{D}(Fc, d) & \xrightarrow{\phi_{c,d}} & \mathsf{C}(c, Gd) \\ {\scriptstyle \mathsf{D}(Fc, g)}\big\downarrow & & \big\downarrow{\scriptstyle \mathsf{C}(c, Gg)} \\ \mathsf{D}(Fc, d') & \xrightarrow[\phi_{c,d'}]{} & \mathsf{C}(c, Gd') \end{array}$$

commute. In other words ϕ gives a natural isomorphism of bifunctors $\mathsf{D}(F-, ?) \longrightarrow \mathsf{C}(-, G?)$.

Remark 7.3. A right adjoint to F, if one exists, is unique up to unique isomorphism. Thus we can relatively safely speak of *the* right adjoint to F.

We saw in the example of the forgetful and free functors between Set and Grp that there was a bunch of additional gadgets that naturally appeared, for instance certain natural transformations in and out of identity functors. This phenomenon is general.

Definition 7.4. Given an adjoint pair of functors $F \dashv G$ as above we can consider, for $c \in \mathsf{C}$ and $d \in \mathsf{D}$, the distinguished morphisms

$$\eta_c = \phi_{c,Fc}(1_{Fc}) \in \mathsf{C}(c, GFc)$$

and

$$\varepsilon_d = \phi^{-1}_{Gd,d}(1_{Gd}) \in \mathsf{D}(FGd, d).$$

These assemble, by naturality of ϕ, to natural transformations

$$\eta\colon \mathrm{id}_{\mathsf{C}} \longrightarrow GF \quad \text{and} \quad \varepsilon\colon FG \longrightarrow \mathrm{id}_{\mathsf{D}}$$

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

Lemma 7.5. *Given $F \dashv G$ as above the unit and counit satisfy the triangle identities, i.e. the composites*

$$F \xrightarrow{F(\eta)} FGF \xrightarrow{\varepsilon_F} F \quad \text{and} \quad G \xrightarrow{\eta_G} GFG \xrightarrow{G(\varepsilon)} G$$

are the identity.

Proof. We leave the proof as an exercise; it's instructive, but rather difficult in some sense. ☐

The fact that the data of the natural isomorphisms $\phi_{c,d}$ and the unit and counit are equivalent, as was observed for the free and forgetful functors is also a general fact.

Theorem 7.6. *The functors $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ and $G\colon \mathsf{D} \longrightarrow \mathsf{C}$ are adjoint if and only if there are natural transformations $\eta\colon \mathrm{id}_{\mathsf{C}} \longrightarrow GF$ and $\varepsilon\colon FG \longrightarrow \mathrm{id}_{\mathsf{D}}$ satisfying the triangle identities.*

Proof. We sketch the relevant constructions. A more complete proof, together with several other chunks of data that determine an adjunction, can be found in [ML98, Theorem IV.1.2].

We have already seen that if $F \dashv G$ we can construct η and ε from applying ϕ and ϕ^{-1} to appropriate identity morphisms in C and D and we left it as an exercise to verify the triangle identities (Lemma 7.5).

Given the unit and counit we define $\phi_{c,d}$ by the composite

$$\mathsf{D}(Fc, d) \xrightarrow{G_{Fc,d}} \mathsf{C}(GFc, Gd) \xrightarrow{\mathsf{C}(\eta_c, Gd)} \mathsf{C}(c, Gd)$$

and $\phi^{-1}_{c,d}$ by the composite

$$\mathsf{C}(c, Gd) \xrightarrow{F_{c,Gd}} \mathsf{D}(Fc, FGd) \xrightarrow{\mathsf{D}(Fc,\varepsilon_d)} \mathsf{D}(Fc, d)$$

We again leave it as an exercise (in using the triangle identities and naturality) to verify that these are actually inverse. ☐

We now come to examples; it is a good exercise to check the details in the cases where we omit them.

Example 7.7. Suppose that $F\colon \mathsf{C} \longrightarrow \mathsf{D}$ is an equivalence with quasi-inverse F^{-1}. Then F^{-1} is both a left and a right adjoint for F as can be seen from the existence of the natural isomorphisms

$$\mathrm{id}_{\mathsf{C}} \xrightarrow{\sim} F^{-1}F \xrightarrow{\sim} \mathrm{id}_{\mathsf{C}}$$

which give one of the units and counits respectively.

Example 7.8. The functors

$$\begin{array}{c} \mathsf{Set} \\ F \Big\downarrow \;\; \Big\uparrow U \\ \mathsf{Grp} \end{array}$$

which we used as our motivating example are an adjoint pair, with F being left adjoint to U (as is alluded to by the placement of F and U in our pictorial representation).

This illustrates a general phenomenon and there are many examples of forgetful functors which admit left adjoints, generally called *free functors*, and right adjoints, generally called *cofree functors*. Let us give another concrete example.

Example 7.9. There is a forgetful functor $U\colon \mathsf{Top} \longrightarrow \mathsf{Set}$ that takes a space and forgets the topology. This functor has both a left adjoint and a right adjoint which we now describe.

First, consider the functor $D\colon \mathsf{Set} \longrightarrow \mathsf{Top}$ which takes a set and equips it with the discrete topology, i.e. sends a set X to the space DX in which every set is both open and closed. The functor D acts on functions of sets by doing nothing—every function out of a discrete space is continuous. We claim that D is left adjoint to U and prove this by exhibing the unit and counit. The unit $\eta\colon \mathrm{id}_{\mathsf{Set}} \longrightarrow UD$ is just the identity, and in fact $UD = \mathrm{id}_{\mathsf{Set}}$ on the nose. The counit $\varepsilon\colon DU \longrightarrow \mathrm{id}_{\mathsf{Top}}$ is also the identity on points, and is automatically continuous since DUY is discrete for any space Y. Note however, that ε_Y is a homeomorphism if and only if Y is discrete. Since neither η nor ε changes anything at the level of sets the triangle identities are easily checked.

We next consider the functor $I\colon \mathsf{Set} \longrightarrow \mathsf{Top}$ which takes a set and equips it with the indiscrete topology, i.e. for $X \in \mathsf{Set}$ the space IX has only $\varnothing$ and IX as open subsets. Any map into an indiscrete space is continuous, so again we can just let I do nothing on morphisms. We still have $UI = \mathrm{id}_{\mathsf{Set}}$ and so we can set ε to be the identity. The unit $\eta\colon \mathrm{id}_{\mathsf{Top}} \longrightarrow IU$ is, again, the identity on points and is continuous since any map to an indiscrete space is continuous. The triangle identities are again easily checked.

Thus we have $D \dashv U \dashv I$ and we can regard, given a set X, the space DX as being the free space on X and IX as the cofree space on X. Heuristically one can make sense of this via the observations, made above, that any map from DX to another space is continuous, so is determined at the level of underlying sets, and similarly for maps into IX.

Example 7.10. Consider $\mathsf{Mod}\,\mathbb{Z}$, the category of abelian groups. For a fixed abelian group M the functor $M \otimes_{\mathbb{Z}} -$, given by tensoring with M, has a right adjoint given by the functor $\mathrm{Hom}_{\mathbb{Z}}(M, -)$ which M corepresents. This is just an expression of the defining isomorphisms for the tensor product: given N and L in $\mathsf{Mod}\,\mathbb{Z}$ we have

$$\mathrm{Hom}_{\mathbb{Z}}(M \otimes_{\mathbb{Z}} N, L) \cong \mathrm{Hom}_{\mathbb{Z}}(M, \mathrm{Hom}_{\mathbb{Z}}(N, L)).$$

8. Limits and colimits

We now come to another central concept in category theory, namely (and unsurprisingly given the section title) limits and, their dual notion, colimits. We will start by giving a highbrow definition and then unwind what this means—then come the examples and many special cases of interest.

Our treatment is very abstract, and it might help to consult another source in conjunction to get a feeling for what we are trying to express.

8.1. **Diagrams and diagram categories.** A *diagram* is another term for a small category J. The intention is that we shift our focus and think of a functor $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ as being a 'J-shaped' collection of objects and morphisms in C. For instance, we could take J to be

$$* \qquad \text{or} \qquad j \xrightarrow{f} j' \qquad \text{or} \qquad j \underset{f_2}{\overset{f_1}{\rightrightarrows}} j'$$

and then giving a functor $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ would be picking an object, a morphism, or two parallel morphisms respectively. We denote the category $*$ with one object and only an identity morphism by $\mathbb{1}$, and the category $1 \longrightarrow 2$ (as in the second example above) with two objects and one non-identity morphism by $\mathbb{2}$.

We can, as we have seen, consider the category

$$\mathsf{C}^{\mathsf{J}} := [\mathsf{J}, \mathsf{C}]$$

of J-shaped diagrams in C, i.e. we give a new name and symbol to the functor category.

Given a functor $G\colon \mathsf{J} \longrightarrow \mathsf{I}$ (i.e. a map of diagrams) we get an induced restriction functor

$$\mathsf{C}^{\mathsf{I}} \longrightarrow \mathsf{C}^{\mathsf{J}} \quad F \mapsto F \circ G$$

(aka $[G, \mathsf{C}]$ as in Example 5.9). For any category J there is a unique functor $\pi\colon \mathsf{J} \longrightarrow \mathbb{1}$ given by collapsing all the objects and morphisms. One checks (hint hint) that $\mathsf{C}^{\mathbb{1}}$ is equivalent to C and so π induces a functor

$$\Delta_{\mathsf{J}}\colon \mathsf{C} \longrightarrow \mathsf{C}^{\mathsf{J}}$$

called the *constant functor of shape* J; the name is due to the fact that $\Delta_{\mathsf{J}}(c)$ is the diagram of shape J with all objects equal to c and all maps the identity.

Definition 8.1. Let $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ be a diagram of shape J in C. The *limit* of F, if it exists, is a representing object (in the sense of Example 4.14) for the functor

$$\mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(-), F)\colon \mathsf{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}$$

i.e. it is an object $\lim F$ of C and a natural isomorphism of functors

$$\mathsf{C}(-, \lim F) \cong \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(-), F).$$

Dually the *colimit* of F, if it exists, is a corepresenting object for the functor

$$\mathsf{C}^{\mathsf{J}}(F, \Delta_{\mathsf{J}}(-))\colon \mathsf{C} \longrightarrow \mathsf{Set},$$

i.e. it is an object $\operatorname{colim} F$ of C and a natural isomorphism of functors

$$\mathsf{C}(\operatorname{colim} F, -) \cong \mathsf{C}^{\mathsf{J}}(F, \Delta_{\mathsf{J}}(-)).$$

8.2. **Back to earth.** Now let us unwind what this actually means. Fix a diagram $F\colon \mathsf{J} \longrightarrow \mathsf{C}$. Given $c \in \mathsf{C}$ a natural transformation $\phi\colon \Delta_{\mathsf{J}}(c) \longrightarrow F$ (often called a cone over F) is given by, for each object $j \in \mathsf{J}$ a map $\phi_j\colon c \longrightarrow F(j)$ such that for each map $f\colon j \longrightarrow j'$ in J the triangle

$$\begin{array}{ccc} c & & \\ {\scriptstyle \phi_j}\downarrow & \searrow^{\phi_{j'}} & \\ F(j) & \xrightarrow[F(f)]{} & F(j') \end{array}$$

commutes. We can think of this as saying that c sits 'above' $F(\mathsf{J})$ and maps to it in a way compatible with the arrows in the diagram. Let us explain that Definition 8.1 is really saying that the limit is the object that sits over $F(\mathsf{J})$ and approximates it best.

If $\lim F$ exists then we know, by plugging $\lim F$ into the defining formula that

$$\mathsf{C}(\lim F, \lim F) \cong \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(\lim F), F),$$

so corresponding to $1_{\lim F}$ there is a map $\delta\colon \Delta_{\mathsf{J}}(\lim F) \longrightarrow F$. This δ should be thought of as the best approximation of F by a constant functor.

That $\lim F$ is a representing object for $\mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(-), F)$ says giving a map $c \longrightarrow \lim F$ in C is the same as giving a map $\Delta_{\mathsf{J}}(c) \longrightarrow F$. In one direction this is clear: if we have $f\colon c \longrightarrow \lim F$ then we get

$$\Delta_{\mathsf{J}}(c) \xrightarrow{\Delta_{\mathsf{J}}(f)} \Delta_{\mathsf{J}}(\lim F) \xrightarrow{\delta} F$$

i.e., we just map c to $\lim F$ by f and follow this by the universal map δ to get the components of a natural transformation. The other direction is where the definition really lives—it asserts that given a map $\Delta_{\mathsf{J}}(c) \longrightarrow F$ there is a unique map $c \longrightarrow \lim F$ corresponding to this. In fact, this tells us that any map $\Delta_{\mathsf{J}}(c) \longrightarrow F$ must factor via δ courtesy of a map $c \longrightarrow \lim F$. Let us actually prove this second statement:

Proof. Let us give a name to the representability isomorphism

$$\phi\colon \mathsf{C}(-, \lim F) \xrightarrow{\sim} \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(-), F)$$

and note that having done so $\delta = \phi_{\lim F}(1_{\lim F})$. Suppose that we are given $f\colon c \longrightarrow \lim F$, with corresponding map $\phi_c\colon \Delta_{\mathsf{J}}(c) \longrightarrow F$. We know, by naturality, that the square

$$\begin{array}{ccc} \mathsf{C}(\lim F, \lim F) & \xrightarrow{\phi_{\lim F}} & \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(\lim F), F) \\ {\scriptstyle \mathsf{C}(f, \lim F)}\downarrow & & \downarrow{\scriptstyle \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(f), F)} \\ \mathsf{C}(c, \lim F) & \xrightarrow[\phi_c]{} & \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(c), F) \end{array}$$

commutes. Thus starting, with $1_{\lim F}$ in the top left we see that

$$\phi_{\lim F}(1_{\lim F}) \circ \Delta_{\mathsf{J}}(f) = \phi_c(f \circ 1_{\lim F}) = \phi_c(f)$$

by going both ways around the square. By definition $\phi_{\lim F}(1_{\lim F}) = \delta$ and so we see $\phi_c(f) = \delta \circ \Delta_{\mathsf{J}}(f)$. Since f was arbitrary this shows that

$$\phi_c = \delta \circ \Delta_{\mathsf{J}}$$

which proves that for any $g\colon \Delta_{\mathsf{J}}(c) \longrightarrow F$ there is a unique map, namely $\phi_c^{-1}(g)$, such that $g = \delta \circ \Delta_{\mathsf{J}}(\phi_c^{-1}(g))$ factors via $\delta\colon \Delta_{\mathsf{J}}(\lim F) \longrightarrow F$ as claimed. $\square$

We can rephrase Definition 8.1.

Definition 8.2. The limit of $F\colon \mathsf{J} \longrightarrow \mathsf{C}$, if it exists, is an object $\lim F \in \mathsf{C}$ and a map $\delta\colon \Delta_{\mathsf{J}}(\lim F) \longrightarrow F$, i.e. a collection of maps $\delta_j\colon \lim F \longrightarrow F(j)$ commuting with maps in $F(\mathsf{J})$ such that given any family of compatible maps $\phi_j\colon c \longrightarrow F(j)$ there is a unique $f\colon c \longrightarrow \lim F$ such that

$$\phi_j = \delta_j \circ f,$$

that is, any $\phi\colon \Delta_{\mathsf{J}}(c) \longrightarrow F$ can be uniquely factored as $\delta \circ \Delta_{\mathsf{J}}(f)$ for some (unique) $f\colon c \longrightarrow \lim F$.

Dually, $\operatorname{colim} F$ if it exists is an object $\operatorname{colim} F \in \mathsf{C}$ and a map $\gamma\colon F \longrightarrow \Delta_{\mathsf{J}}(\operatorname{colim} F)$ such that for any map $\psi\colon F \longrightarrow \Delta_{\mathsf{J}}(c)$ there is a unique $g\colon \operatorname{colim} F \longrightarrow c$ such that $\psi = \Delta_{\mathsf{J}}(g) \circ \gamma$.

Informally, $\lim F$ is the universal object sitting over F and $\operatorname{colim} F$ is the universal object sitting under it.

8.3. **Via adjunction.** The formulas given in Definition 8.1 are reminiscent of the definition of an adjunction. Indeed, consider the functor

$$\Delta_{\mathsf{J}}\colon \mathsf{C} \longrightarrow \mathsf{C}^{\mathsf{J}}$$

and suppose it had a left adjoint, which we suggestively name colim, and a right adjoint which we suggestively name lim. This would tell us there were, for $c \in \mathsf{C}$ and $F \in \mathsf{C}^{\mathsf{J}}$, natural bijections

$$\mathsf{C}(\operatorname{colim} F, c) \cong \mathsf{C}^{\mathsf{J}}(F, \Delta_{\mathsf{J}}(c)) \quad \text{and} \quad \mathsf{C}(c, \lim F) \cong \mathsf{C}^{\mathsf{J}}(\Delta_{\mathsf{J}}(c), F)$$

In other words a left adjoint for Δ_{J} sends a diagram F to its colimit and a right adjoint sends F to its limit (and in particular the existence of the adjoint includes that the respective limit or colimit exists). The universal maps discussed above arise from the unit and counit

$$\mathrm{id}_{\mathsf{C}^{\mathsf{J}}} \longrightarrow \Delta_{\mathsf{J}}(\operatorname{colim} -) \quad \text{and} \quad \Delta_{\mathsf{J}}(\lim -) \longrightarrow \mathrm{id}_{\mathsf{C}^{\mathsf{J}}}.$$

Definition 8.3. If the left adjoint $\operatorname{colim} \dashv \Delta_{\mathsf{J}}$ exists we say C *has colimits of shape* J, and similarly if lim exists. If for every diagram J the left adjoint to Δ_{J} exists we say C is *cocomplete* and if for every J the right adjoint lim exists we say C is *complete*. Thus C is (co)complete if it has all small (co)limits.

8.4. **Important special cases and examples.** We now go through a number of choices of diagram J, indicate the various special names for the corresponding limits and colimits, describe what precisely they are in down to earth terms, and give examples.

8.4.1. *Initial and terminal objects.* We first take for J the empty category $\varnothing$ with no objects. In this case $\mathsf{C}^{\mathsf{J}} \cong \mathbb{1}$ is the category with a unique object, the empty functor $\varnothing\colon \varnothing \longrightarrow \mathsf{C}$, which just includes the empty set into $\operatorname{Ob} \mathsf{C}$ and $\Delta_\varnothing$ is the functor $\pi\colon \mathsf{C} \longrightarrow \mathbb{1}$ collapsing C. In this case $\operatorname{colim} \varnothing$ is an object i such that

$$\mathsf{C}(i, c) \cong \mathsf{C}^{\varnothing}(\varnothing, \pi(c)) = \mathsf{C}^{\varnothing}(\varnothing, \varnothing) = \{*\}.$$

So, we want an object $i \in \mathsf{C}$ such that $\mathsf{C}(i, c) \cong \{*\}$ for every $c \in \mathsf{C}$, i.e. such that there is a unique map from i to c for every object $c \in \mathsf{C}$. Such an object is called an *initial object* of C. Dually, the limit of the empty diagram is an object t receiving a unique map from every object of C, i.e. such that $\mathsf{C}(c, t) \cong \{*\}$ for every $c \in \mathsf{C}$. Such an object is called a *terminal object* of C.

Before giving examples let us show that a terminal object, if one exists, is unique up to unique isomorphism. This is the case for any limit or colimit (and we hope this argument provides the germ of an idea of how one might generalize the proof).

Lemma 8.4. *If t and t' are terminal in C then there is a unique isomorphism $f\colon t \xrightarrow{\sim} t'$.*

Proof. By definition there is a unique map $f\colon t \longrightarrow t'$ and a unique map $f'\colon t' \longrightarrow t$. The composite $f'f$ is an endomorphism of t. By definition, $\mathsf{C}(t,t)$ has a single element, which much be 1_t and so $f'f = 1_t$. Similarly we see $ff' = 1'_t$ and so $t \cong t'$ by the unique maps f and f'. □

Example 8.5. The category Set has an initial object $\varnothing$, the empty set, and any set with a single element, for instance $\{*\}$ is a terminal object. Similarly, in Top the initial object is the empty space and any space with a single point is terminal.

Example 8.6. In Grp the trivial group 1 is both initial and terminal.

Definition 8.7. We call an object which is both initial and terminal a *zero object*.

The terminology is motivated by the next example.

Example 8.8. If R is a ring then 0, the zero module, is both initial and terminal in $\mathsf{Mod}\, R$, the category of right R-modules. Given R-modules M and N the zero map $0\colon M \longrightarrow N$ is the unique map factoring as $M \longrightarrow 0 \longrightarrow N$.

Example 8.9. If G is a non-trivial group then BG has neither an initial nor a terminal object.

Example 8.10. In the category of pointed sets Set_* the object $(*,*)$, i.e. a singleton set based at its unique element, is both initial and terminal. Indeed, it is initial since any map $(*,*) \longrightarrow (X, x_0)$ must, by definition, send $*$ to x_0 and there is always such a map (being pointed means the empty set is not an object of Set_*). On the other hand, any map $(X, x_0) \longrightarrow (*,*)$ must just collapse everything so there is again precisely one.

Thus $(*,*)$, or just $*$ for short, is a zero object for Set_*. Thus there is always a zero map $(X, x_0) \longrightarrow (Y, y_0)$ which is given by the obvious unique composite via $*$, and acts by collapsing everything to y_0.

8.4.2. *Coproducts and products.* We next consider the case that $\mathsf{J} = \{1,2\}$ is a category with two objects and only identity morphisms. In this situation a functor $\mathsf{J} \longrightarrow \mathsf{C}$ picks out two objects $F(1)$ and $F(2)$. Since J has only identity maps a natural transformation $F \longrightarrow F'$ is just given by a pair of maps $F(1) \longrightarrow F'(1)$ and $F(2) \longrightarrow F'(2)$. Thus

$$\mathsf{C}^{\mathsf{J}} \cong \mathsf{C} \times \mathsf{C}$$

where $\mathsf{C} \times \mathsf{C}$ is the category with objects $\operatorname{Ob}\mathsf{C} \times \operatorname{Ob}\mathsf{C}$ and maps

$$\mathsf{C} \times \mathsf{C}((c_1, c_2), (c'_1, c'_2)) = \mathsf{C}(c_1, c'_1) \times \mathsf{C}(c_2, c'_2).$$

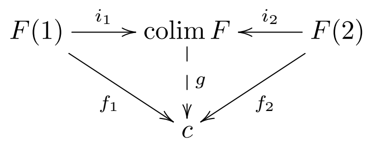
The constant diagram functor $\Delta_{\mathsf{J}}\colon \mathsf{C} \longrightarrow \mathsf{C} \times \mathsf{C}$ is just the inclusion given by sending c to (c,c) and similarly on morphisms.

Thus, given a diagram $F = (F(1), F(2))$ the colimit is an object $\operatorname{colim} F$ such that

$$\mathsf{C}(\operatorname{colim} F, c) \cong \mathsf{C}(F(1), c) \times \mathsf{C}(F(2), c),$$

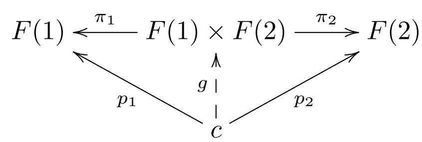
i.e., a map out of the colimit of F is the same thing as a map from $F(1)$ and a map from $F(2)$. This can be reinterpreted as saying it is an object $\operatorname{colim} F$ with maps $i_1\colon F(1) \longrightarrow \operatorname{colim} F$ and $i_2\colon F(2) \longrightarrow \operatorname{colim} F$ such that given any pair of maps $f_i\colon F(i) \longrightarrow c$ for

$i = 1, 2$ there is a unique map $\operatorname{colim} F \xrightarrow{g} c$ making the following diagram commute



In this case the object $\operatorname{colim} F$ is called the *coproduct* of $F(1)$ and $F(2)$ and is usually written $F(1) \coprod F(2)$.

Dually, the limit of F, usually written $F(1) \times F(2)$ and called the *product* of $F(1)$ and $F(2)$ satisfies the universal property indicated by the following diagram with respect to mapping to $F(1)$ and $F(2)$



Remark 8.11. This can be generalised to the case when J instead has n objects to give the (co)product of the n objects $\mathsf{J} \longrightarrow \mathsf{C}$ picks out. In fact, one can take J to be any set viewed as a category (with only identity morphisms) and consider the corresponding (co)limit to define the (co)product of a family of objects indexed by $\operatorname{Ob} \mathsf{J}$.

Remark 8.12. One can also consider the case where $\mathsf{J} = \mathbb{1}$ the category with a single object. In this case both lim and colim agree and are the identity. One can also think of an initial object as being an empty coproduct and a terminal object as being an empty product.

Example 8.13. The coproduct of $X, Y \in \mathsf{Set}$ is their disjoint union $X \coprod Y$. Their product $X \times Y$ is the usual cartesian product. This is the inspiration for the general notation. The same constructions, namely disjoint union and cartesian product, give the coproduct and product in Top as well.

Example 8.14. If R is a ring then the product and coproduct of M and N in $\mathsf{Mod}\, R$ coincide and are equal to $M \oplus N$, the direct sum of M and N (aka the cartesian product). In this case, when the coproduct and product coincide (for finite indexing sets), the resulting object is often called a *biproduct*.

Example 8.15. Consider the category Set_* of pointed sets. The coproduct of (X, x_0) and (Y, y_0) is $(X \vee Y, *)$ (really the notation should denote the base points, as the result may well depend upon them, but it is standard to gloss over this), the *wedge sum* of (X, x_0) and (Y, y_0) which is defined by be

$$(X, x_0) \vee (Y, y_0) = (X \coprod Y)/(x_0 \sim y_0)$$

i.e. we take the disjoint union of X and Y and identify the base points.

The coproduct in pointed spaces Top_* is given similarly and also called the wedge sum.

Example 8.16. Let us return to the example of modules over a ring R. Given a set $\{M_i \mid i \in I\}$ of modules, with I infinite, their product and coproduct still exist but no longer coincide. They are denoted

$$\prod_{i \in I} M_i \quad \text{and} \quad \bigoplus_{i \in I} M_i$$

respectively. The former is the infinite cartesian product, in the sense that it is defined to be sequences of elements, the ith in M_i, and addition and the action of R are given componentwise. The coproduct, often called the direct sum, is the submodule of the product consisting of sequences with all but finitely many entries zero.

Coproducts and products may fail to exist.

Example 8.17. The category of groups Grp has all coproducts and products. Products are given by the usual cartesian products. The coproduct of two groups G and H is the *free product* $G * H$, which is defined as follows: suppose we have presentations for G and H, say

$$G = \langle X_G \mid R_G \rangle \quad \text{and} \quad H = \langle X_H \mid R_H \rangle$$

where the X's are sets of generators and the R's are sets of relations. The free product is then given by

$$G * H = \langle X_G \coprod X_H \mid R_G \coprod R_H \rangle.$$

This is independent of the presentations chosen. One can construct it without reference to presentations by taking (reduced) words in the elements of G and H.

Now consider $\mathsf{Grp}^{\mathsf{f}}$, the category of finite groups. This category does *not* have coproducts. We don't give a proof, but we leave it as an exercise to check that the free product of two finite non-trivial groups is never finite (this is not enough, one needs to check the coproduct couldn't be something else, but it's something).

8.4.3. *Coequalizers and equalizers.* For our next act, we take

$$\mathsf{J} = \; 1 \underset{g}{\overset{f}{\rightrightarrows}} 2$$

so a functor $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ is given by a pair of parallel maps $F(f), F(g)\colon F(1) \longrightarrow F(2)$ in C. A natural transformation from $F \longrightarrow G$ in C^{J} is given by a pair of maps $\alpha_i\colon F(i) \longrightarrow G(i)$ such that both the square for f and the square for g of

$$\begin{array}{ccc} F(1) & \underset{F(g)}{\overset{F(f)}{\rightrightarrows}} & F(2) \\ {\scriptstyle \alpha_1}\downarrow & & \downarrow{\scriptstyle \alpha_2} \\ G(1) & \underset{G(g)}{\overset{G(f)}{\rightrightarrows}} & G(2) \end{array}$$

commute. In particular, a map $\Delta_{\mathsf{J}}(c) \longrightarrow F$ is given by α_1, α_2 such that

$$\begin{array}{ccc} c & = & c \\ {\scriptstyle \alpha_1}\downarrow & & \downarrow{\scriptstyle \alpha_2} \\ F(1) & \underset{F(g)}{\overset{F(f)}{\rightrightarrows}} & F(2) \end{array}$$

commutes. But then $\alpha_2 = F(f)\alpha_1 = F(g)\alpha_1$, so such a map is really just determined by a map $\alpha\colon c \longrightarrow F(1)$ such that both composites to $F(2)$ are equal.

Given $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ the limit, if it exists, is defined by

$$\mathsf{C}(c, \lim F) \cong \mathsf{C}(\Delta_{\mathsf{J}}(c), F) = \{f \in \mathsf{C}(c, F(1)) \mid F(f)\alpha = F(g)\alpha\}.$$

Thinking of Definition 8.2 we see that $\lim F$ if it exists, is given by a map $\delta\colon \lim F \longrightarrow F(1)$ satisfying $F(f)\delta = F(g)\delta$ and such that any map $\alpha\colon c \longrightarrow F(1)$ such that $F(f)\alpha = F(g)\alpha$ factors uniquely via δ. In this case $\lim F$ is called the *equalizer of* $F(f)$ *and* $F(g)$.

Dually, $\operatorname{colim} F$ is given by a map $\delta\colon F(2) \longrightarrow \operatorname{colim} F$ such that any map $\beta\colon F(2) \longrightarrow c$ with $\beta F(f) = \beta F(g)$ factors uniquely via δ and is called the *coequalizer of $F(f)$ and $F(g)$*.

Example 8.18. We begin with Set, and suppose we are given $f, g\colon X \longrightarrow Y$ a pair of functions from X to Y. The equalizer of f and g is

$$\{x \in X \mid f(x) = g(x)\} \subseteq X,$$

the inclusion of the set of elements on which f and g coincide. This evidently satisfies the universal property—if $h\colon W \longrightarrow X$ satisfies $fh = gh$, then $h(W)$ must land in the collection of elements on which f and g agree.

The coequalizer is the quotient of Y by the equivalence relation generated by $f(x) \sim g(x)$, i.e. it is the set of equivalence classes with respect to the equivalence relation we get by identifying the images of f and g.

Example 8.19. Let R be a ring and consider $f\colon M \longrightarrow N$ in $\mathsf{Mod}\, R$. The equalizer of f and 0 is called the *kernel* of f and is given by

$$\ker f = \{m \in M \mid f(m) = 0\}$$

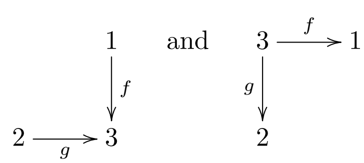
More generally, given another map $g\colon M \longrightarrow N$ the equalizer of f and g is, as in the last example, given by

$$\{m \in M \mid f(m) = g(m)\} \subseteq M$$

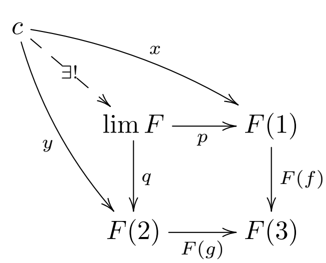
and is a submodule of M. It can also be written as $\ker(f - g)$, i.e. kernels and equalizers are equivalent concepts.

Dually, the coequalizer of f and 0 is the *cokernel* of f written $\operatorname{coker} f$ and can be defined as $M/\operatorname{im}(f)$ where $\operatorname{im}(f)$ is the image of f. The coequalizer of f and g can then be defined by $\operatorname{coker}(f - g)$, or constructed by killing the submodule of N generated by elements $f(m) - g(m)$ which amounts to the same thing (as it must by uniqueness).

8.4.4. *Pushouts and pullbacks.* The final particular diagrams we will consider are



which we denote by J and J^{op} (because they are formally dual). In this case the limit of a diagram $F\colon \mathsf{J} \longrightarrow \mathsf{C}$ is given as in the diagram



i.e. it is an object $\lim F$ together with maps p and q making the lower square commute, such that given any maps x, y out of c that make the outer diagram commute there is a unique map $c \longrightarrow \lim F$ making everything commute. In this case $\lim F$ is called the *pullback* of the diagram F and denoted by $F(1) \times_{F(3)} F(2)$.

Dually, given $F\colon \mathsf{J}^{\mathrm{op}} \longrightarrow \mathsf{C}$, the colimit $\operatorname{colim} F$ is called a *pushout* (and we leave it as an exercise to formulate the universal property—or to look it up on Wikipedia) and denoted $F(1) \coprod_{F(3)} F(2)$ or sometimes $F(1) \cup_{F(3)} F(2)$.

We first note that under favourable circumstances pullbacks and pushouts can be viewed as generalizations of products and coproducts.

Example 8.20. Suppose that C has a terminal object $*$ and consider for $c, c' \in \mathsf{C}$ the diagram

$$\begin{array}{ccc} & & c \\ & & \downarrow \\ c' & \longrightarrow & * \end{array}$$

The pullback $c \times_* c'$, if it exists, is the product $c \times c'$. Indeed, since every object has a unique map to the terminal object any pair of maps $b \longrightarrow c$ and $b \longrightarrow c'$ will make the diagram commute and so the universal properties of the pullback and product coincide in this case.

Analogously, if C has an initial object $\varnothing$ then the pushout of

$$\begin{array}{ccc} \varnothing & \longrightarrow & c \\ \downarrow & & \\ c' & & \end{array}$$

is the coproduct $c \coprod c'$.

Example 8.21. In fact there is a further connection between pullbacks and products. As the previous example may suggest one can think loosely of a pullback as 'compatible elements in the product'. This is made precise in the following sense: given a diagram

$$\begin{array}{ccc} & & c \\ & & \downarrow{\scriptstyle f} \\ c' & \xrightarrow[g]{} & d \end{array}$$

if $c \times c'$ exists then it comes with maps

$$c \xleftarrow{\pi} c \times c' \xrightarrow{\pi'} c'$$

and so composing with f and g respectively we get a diagram

$$c \times c' \overset{f\pi}{\underset{g\pi'}{\rightrightarrows}} d$$

We claim that the equalizer $e \xrightarrow{h} c \times c'$ is the pullback $c \times_d c'$ via the maps πh and $\pi' h$. We leave the verification of this fact as an exercise.

Dually, one can express a pushout $c \coprod_d c'$ as a coequalizer of a pair of maps to $c \coprod c'$ and, in this way, view it as c and c' glued together by identifying the 'images' of d in each (note we are being vague here, the image doesn't always make sense).

In particular, we (really you if you did the verification) have proved:

Lemma 8.22. *If C has finite products and equalizers then C has pullbacks. Dually, if C has finite coproducts and coequalizers it has pushouts.*

Example 8.23. Consider the category Top of topological spaces. Given any diagram

$$\begin{array}{ccc} & & X \\ & & \downarrow f \\ Y & \xrightarrow[g]{} & Z \end{array}$$

of spaces and continuous maps the pullback exists. It can be given, for instance, by the construction we saw in the previous example

$$X \times_Z Y = \{(x, y) \in X \times Y \mid f(x) = g(y)\}.$$

Dually, given any diagram

$$\begin{array}{ccc} Z & \xrightarrow{f} & X \\ g \downarrow & & \\ Y & & \end{array}$$

the pushout exists and can be constructed as the quotient space

$$X \coprod_Z Y = X \coprod Y / \sim$$

where $\sim$ is the equivalence relation generated by $f(z) = g(z)$ for $z \in Z$ (strictly speaking for f and g composed with the inclusions into the disjoint union).

8.4.5. *General limits and colimits.* We saw in the last section that pullbacks can be constructed from products and equalizers. In fact, all limits can be constructed in this way. Dually all colimits can be built from coproducts and coequalizers. We sketch a proof of this fact here (we sketch a proof for the case of limits, the case of colimits being formally dual).

Really, we give more of a construction than a proof, the details being left as an exercise. Suppose that $F \colon \mathsf{J} \longrightarrow \mathsf{C}$ is a diagram of shape J in C, and suppose that C admits all products and equalizers (one can get away with only assuming certain products and equalizers exist, and which one needs for a given shape will become clear as we proceed). Consider the two objects

$$\prod_{j \in \mathrm{Ob}\,\mathsf{J}} F(j) \quad \text{and} \quad \prod_{j \xrightarrow{f} j'} F(j')$$

where the second product is taken over all maps in J. We define two maps from the first object to the second. In order to specify these maps we note that by the universal property of the product

$$\mathsf{C}\Big(\prod_{j \in \mathrm{Ob}\,\mathsf{J}} F(j), \prod_{j \xrightarrow{f} j'} F(j')\Big) \cong \prod_{j \xrightarrow{f} j'} \mathsf{C}\Big(\prod_{j \in \mathrm{Ob}\,\mathsf{J}} F(j), F(j')\Big)$$

so it's enough to specify a map from the first product to $F(j')$ for each $f \colon j \longrightarrow j'$ in J.

Let p be the map corresponding to picking, for $g \colon j' \longrightarrow j''$, the map

$$\prod_{j \in \mathrm{Ob}\,\mathsf{J}} F(j) \xrightarrow{\pi_{F(j'')}} F(j'')$$

given by projection onto F of the target, and q be the map determined by the components

$$\prod_{j\in \mathrm{Ob}\,\mathsf{J}} F(j) \xrightarrow{\pi_{F(j')}} F(j') \xrightarrow{F(g)} F(j'')$$

and consider the equaliser

$$L \xrightarrow{\phi} \prod_{j\in \mathrm{Ob}\,\mathsf{J}} F(j) \underset{q}{\overset{p}{\rightrightarrows}} \prod_{j\xrightarrow{f} j'} F(j').$$

We claim that L together with ϕ (really with ϕ composed with the projections $\pi_{F(j)}$) is the limit of F. First note that, by definition of the equaliser, ϕ gives a natural transformation $\Delta_{\mathsf{J}}(L) \longrightarrow F$ (which we also call ϕ); this requires some unwrapping but one sees this to be true by looking at the components defining p and q and noting that equalising them says precisely that the square

$$\begin{array}{ccc} L & \xrightarrow{\pi_{F(j)}\phi} & F(j) \\ \| & & \downarrow{\scriptstyle F(f)} \\ L & \xrightarrow[\pi_{F(j')}\phi]{} & F(j') \end{array}$$

commutes for every $f\colon j \longrightarrow j'$ in J. Thus ϕ must factor through a unique map $\alpha\colon L \longrightarrow \lim F$. On the other hand, the map $\delta\colon \Delta_{\mathsf{J}}(\lim F) \longrightarrow F$, that comes from being the limit of F, is the data of maps $\delta_j\colon \lim F \longrightarrow F(j)$ and so defines a map $\lim F \longrightarrow \prod_{j\in \mathrm{Ob}\,\mathsf{J}} F(j)$. Naturality of δ is the statement that this map equalises p and q and so there is a unique map $\alpha^{-1}\colon \lim F \longrightarrow L$ such that δ factors via ϕ. Since both of these maps are unique and compatible with δ and ϕ it is a formal consequence that they must be inverse to one another. This proves that L is the limit of F.

References

[ML98] Saunders Mac Lane, *Categories for the working mathematician*, Second, Graduate Texts in Mathematics, vol. 5, Springer-Verlag, New York, 1998. MR1712872 (2001j:18001)