0 Preliminaries

I argue that set theory should not be based on membership, as in Zermelo-Frankel set theory, but rather on isomorphism-invariant structure.
—William Lawvere (Freitas, 2007)

Introduction. Traditionally, the first chapter of a textbook on mathematics begins by recalling basic notions from set theory. This chapter begins by introducing basic notions from category theory, the shift being from the internal anatomy of sets to their relationships with other sets. The idea of focusing on the relationships between mathematical objects, rather than on their internals, is fundamental to modern mathematics, and category theory is the framework for working from this perspective. Our goal for chapter 0 is to present what is perhaps familiar to you—functions, sets, topological spaces—from the contemporary perspective of category theory. Notably, category theory originated in topology in the 1940s with work of Samuel Eilenberg and Saunders MacLane (Eilenberg and MacLane, 1945).

This chapter's material is organized into three sections. Section 0.1 begins with a quick review of topological spaces, bases, and continuous functions. Motivated by a few key features of topological spaces and continuous functions, we'll proceed to section 0.2 and introduce three basic concepts of category theory: categories, functors, and natural transformations. The same section highlights one of the main philosophies of category theory, namely that studying a mathematical object is akin to studying its relationships to other objects. This golden thread starts in section 0.2 and weaves its way through the remaining pages of the book—we encourage you to keep an eye out for the occasional glimmer. Finally, equipped with the categorical mindset, we'll revisit some familiar ideas from basic set theory in section 0.3.

0.1 Basic Topology

Definition 0.1 A *topological space* $(X, \mathcal{T})$ consists of a set X and a collection $\mathcal{T}$ of subsets of X that satisfy the following properties:

(i) The empty set $\varnothing$ and X are in $\mathcal{T}$.
(ii) Any union of elements in $\mathcal{T}$ is also in $\mathcal{T}$.
(iii) Any finite intersection of elements in $\mathcal{T}$ is also in $\mathcal{T}$.

The collection $\mathcal{T}$ is called a *topology* on X, and we'll write X in place of $(X, \mathcal{T})$ if the topology is understood. Occasionally, we'll also refer to the topological space X as simply a *space*. Elements of the topology $\mathcal{T}$ are called *open sets*, and a set is called *closed* if and only if its complement is open.

Example 0.1 Suppose X is any set. The collection 2^X of all subsets of X forms a topology called the *discrete topology* on X, and the set $\{\varnothing, X\}$ forms a topology on X called the *indiscrete topology* or the *trivial topology*.

Sometimes, two topologies on the same set are comparable. When $\mathcal{T} \subseteq \mathcal{T}'$, the topology $\mathcal{T}$ can be called *coarser* than $\mathcal{T}'$, or the topology $\mathcal{T}'$ can be called *finer* than $\mathcal{T}$. Instead of coarser and finer, some people say "smaller and larger" or "weaker and stronger," but the terminology becomes clearer—as with most things in life—with coffee. A *coarse* grind yields a *small* number of chunky coffee pieces, whereas a *fine* grind results in a *large* number of tiny coffee pieces. Finely ground beans make stronger coffee; coarsely ground beans make weaker coffee.

In practice, it can be easier to work with a small collection of open subsets of X that generates the topology.

Definition 0.2 A collection $\mathcal{B}$ of subsets of a set X is a *basis* for a topology on X if and only if

(i) For each $x \in X$ there is a $B \in \mathcal{B}$ such that $x \in B$.

(ii) If $x \in A \cap B$ where $A, B \in \mathcal{B}$, then there is at least one $C \in \mathcal{B}$ such that $x \in C \subseteq A \cap B$.

The topology $\mathcal{T}$ *generated* by the basis $\mathcal{B}$ is defined to be the coarsest topology containing $\mathcal{B}$. Equivalently, a set $U \subseteq X$ is open in the topology generated by the basis $\mathcal{B}$ if and only if for every $x \in U$ there is a $B \in \mathcal{B}$ such that $x \in B \subseteq U$.

We call the collection of $B \in \mathcal{B}$ with $x \in B$ the *basic open neighborhoods* of x. More generally, in any topology $\mathcal{T}$, those $U \in \mathcal{T}$ containing x are called *open neighborhoods* of x and together are denoted $\mathcal{T}_x$.

Example 0.2 A *metric space* is a pair (X, d) where X is a set and $d\colon X \times X \to \mathbb{R}$ is a function satisfying

- $d(x, y) \geq 0$ for all $x, y \in X$,
- $d(x, y) = d(y, x)$ for all $x, y \in X$,
- $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$,
- $d(x, y) = 0$ if and only if $x = y$ for all $x, y \in X$.

The function d is called a *metric* or a *distance function*. If (X, d) is a metric space, $x \in X$, and $r > 0$, then the ball centered at x of radius r is defined to be

$$B(x, r) = \{y \in X \mid d(x, y) < r\}.$$

The balls $\{B(x,r)\}$ form a basis for a topology on X called *the metric topology*. So any set with a metric gives rise to a topological space. On the other hand, if Y is a space with topology $\mathcal{T}$ and if there is a metric d on Y such that the metric topology is the same as $\mathcal{T}$, then Y is said to be *metrizable*.

Any subset of a metric space is a metric space. In particular, subsets of $\mathbb{R}^n$ provide numerous examples of topological spaces since $\mathbb{R}^n$ with the usual Euclidean distance function is a metric space. For example,

- the real line $\mathbb{R}$,
- the unit interval $I := [0,1]$,
- the closed unit ball $D^n := \{(x_1,\dots,x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 \leq 1\}$, and
- the *n-sphere* $S^n := \{(x_1,\dots,x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}$

are all important topological spaces.

We'll see more examples of topological spaces in chapter 1 and will discuss some important features in chapter 2. One way we'll glean information is by studying how spaces relate to each other. These relationships are best understood as functions that interact nicely—in the sense of the following definition—with open subsets of the spaces.

Definition 0.3 A function $f\colon X \to Y$ between two topological spaces is *continuous* if and only if $f^{-1}U$ is open in X whenever U is open in Y.

It is straightforward to check that for any topological space X, the identity $\mathrm{id}_X\colon X \to X$ is continuous, and for any topological spaces X, Y, Z and any continuous functions $f\colon X \to Y$ and $g\colon Y \to Z$, the composition $gf := g \circ f\colon X \to Z$ is continuous, and moreover that this composition is associative. On the surface, these observations may appear to be ho-hum, but quite the opposite is true. Collectively, the seemingly routine observations above amount to the statement that *topological spaces together with continuous functions form a category*.

0.2 Basic Category Theory

In this section, we'll give the formal definition of a *category* along with several examples.

0.2.1 Categories

Definition 0.4 A *category* C consists of the following data:

(i) a class of *objects*,

(ii) for every two objects X, Y, a set[1] $\mathsf{C}(X, Y)$ whose elements are called *morphisms* and denoted by arrows; for example $f\colon X \to Y$,

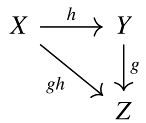
(iii) a *composition* rule defined for morphisms: if $f\colon X \to Y$ and $g\colon Y \to Z$, then there is a morphism $gf\colon X \to Z$.

These data must satisfy the following two conditions:

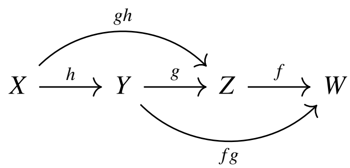
(i) Composition is associative. That is, if $h\colon X \to Y$, $g\colon Y \to Z$, $f\colon Z \to W$, then $f(gh) = (fg)h$.

(ii) There exist identity morphisms. That is, for every object X, there exists a morphism $\mathrm{id}_X\colon X \to X$ with the property that $f\,\mathrm{id}_X = f = \mathrm{id}_Y\, f$ whenever f is a morphism from X to Y. By the usual argument, identity morphisms are unique: if $\mathrm{id}'_X\colon X \to X$ is another identity morphism, then $\mathrm{id}'_X = \mathrm{id}'_X\,\mathrm{id}_X = \mathrm{id}_X$.

The associativity condition can also be expressed by way of a commutative diagram. A *diagram* can be thought of as a directed graph with morphisms as edges and with objects as vertices, though we'll give a more categorical definition in chapter 4. A diagram is said to *commute* (or is *commutative*) if all paths that share the same initial and final vertex are the same. For example, if $h\colon X \to Y$ and $g\colon Y \to Z$ are composable morphisms, then there is a commutative diagram:



And if $f\colon Z \to W$ is a third morphism, then the property, "composition is associative,"—that is, $f(gh) = (fg)h$—is equivalent to the statement that the following diagram commutes:



Simply put, a diagram is a visualization of morphism composition. A commutative diagram is a visualization of equalities between compositions. We'll see many more examples

[1] Some authors denote the set of morphisms from X to Y by $\hom_{\mathsf{C}}(X, Y)$ or simply $\hom(X, Y)$ if the category C is understood. Since keeping track of subscripts can be a chore, we'll usually promote the subscript to the forefront and write $\mathsf{C}(X, Y)$. While this notation may take some getting used to, we encourage the reader to do so. Authors also differ in their definitions of "category." We require a category to have *only a set's worth* of morphisms between any two objects—a property known as being *locally small*. Others sometimes allow categories to have more arrows.

in the pages to come. And speaking of examples, we already mentioned that topological spaces with continuous functions form a category. Here is another.

Example 0.3 For any given field $\mathbf{k}$, there is a category denoted $\mathsf{Vect}_{\mathbf{k}}$ whose objects $V, W, \ldots$ are vector spaces over $\mathbf{k}$ and whose morphisms are linear transformations. To verify the claim, suppose $T\colon V \to W$ and $S : W \to U$ are linear transformations. Then for any $v, v' \in V$ and any $k \in \mathbf{k}$,

$$ST(kv + v') = S(kTv + Tv') = kSTv + STv'$$

and so $ST\colon V \to U$ is indeed a linear transformation. Associativity of composition is automatic since linear transformations are functions and composition of functions is always associative. And for any vector space, the identity function is a linear transformation. More generally, modules over a fixed ring R together with R-module homomorphisms form a category, $R\mathsf{Mod}$.

Here are a few more examples. This time we'll leave the verifications as an exercise.

- Set: The objects are sets, the morphisms are functions, and composition is composition of functions.
- Set_*: The objects are sets S having a distinguished element. (Such sets are called *pointed sets.*) A morphism $f\colon S \to T$ is a function satisfying $fs_0 = t_0$ whenever s_0 is the distinguished element of S and t_0 is the distinguished element of T. (Such functions are said to "respect" the distinguished elements.) Composition is composition of functions.
- Top: The objects are topological spaces, the morphisms are continuous functions, and composition is composition of functions.
- Top_*: The objects are topological spaces with a distinguished point, often called a basepoint. (Such spaces are called *pointed* or *based* spaces.) The morphisms are continuous functions that map basepoint to basepoint, and composition is composition of functions.
- hTop: The objects are topological spaces, the morphisms are homotopy classes of continuous functions, and composition is composition of these homotopy classes. The notion of homotopy will be introduced in section 1.6.
- Grp: The objects are groups, the morphisms are group homomorphisms, and composition is composition of homomorphisms.
- Every group G can be viewed as a category with one object $\bullet$ and with a morphism $g\colon \bullet \to \bullet$ for each group element g. The composition of two morphisms f, g corresponds to the group element gf.

- A directed multigraph determines a category whose objects are the vertices and whose morphisms are the directed paths along finitely many arrows joined head-to-tail.[2] For example, the directed graph

$$\bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \cdots$$

  determines a category that will make an appearance in example 4.1. For simple graphs, one might want to display more information by drawing identities and composite morphisms. To avoid confusion, one can state whether identities and composites are displayed. Here, of course, they are not drawn. This example also illustrates another notational convenience: we'll frequently use symbols such as $\bullet$ and $\circ$ as anonymous placeholders. So unless otherwise indicated, each "$\bullet$" should be considered a distinct object.
- For any category C, there is an opposite category C^{op}. The objects are the same as the objects of C, but the morphisms are reversed. Composition in C^{op} is defined by composition in C, that is, $\mathsf{C}^{\mathrm{op}}(X, Y) = \mathsf{C}(Y, X)$. To check that composition makes sense, suppose $f \in \mathsf{C}^{\mathrm{op}}(X, Y)$ and $g \in \mathsf{C}^{\mathrm{op}}(Y, Z)$ so that $f \colon Y \to X$ and $g \colon Z \to Y$. Then $fg \colon Z \to X$ and therefore $fg \in \mathsf{C}^{\mathrm{op}}(X, Z)$ as needed.

Category theory is an appropriate setting in which to discuss an age-old question: "When are two objects really the same?" The concept of "sameness" being a special kind of relationship means that objects are the same if there is a particular morphism—an *isomorphism*—between them. How we talk about sameness is at the heart of category theory. We take Wittgenstein's (1922) critique,

> Roughly speaking, to say of two things that they are identical is nonsense, and to say of one thing that it is identical with itself is to say nothing at all

as an invitation to ask for less, namely isomorphic (or uniquely isomorphic) objects, rather than identical objects.

Definition 0.5 Let X and Y be objects in any category, and suppose $f \colon X \to Y$.

(i) f is *left invertible* if and only if there exists a morphism $g \colon Y \to X$ so that $gf = \mathrm{id}_X$. The morphism g is called a *left inverse* of f.

(ii) f is *right invertible* if and only if there exists a morphism $h \colon Y \to X$ so that $fh = \mathrm{id}_Y$. The morphism h is called a *right inverse* of f.

In the case when f has both a left inverse g and a right inverse h, then

$$g = g\,\mathrm{id}_Y = gfh = \mathrm{id}_X\,h = h$$

and the single morphism $g = h$ is called the *inverse* of f. (We encourage you to verify our use of "the." If f has an inverse, it is unique.) Therefore

[2] Including—for each object—a unique path of length zero starting and ending at that object.

(iii) f is *invertible* and is said to be an *isomorphism* if it is both left and right invertible. Two objects X and Y are *isomorphic*, denoted $X \cong Y$, if there exists an isomorphism $f : X \to Y$.

The notion of *being isomorphic* is a form of equivalence, meaning it is reflexive, symmetric, and transitive—properties you should check. Isomorphic objects, therefore, form equivalence classes. Some categories have their own special terminology for both *isomorphism* and these *isomorphism classes*. For instance,

- Isomorphisms in Set are called *bijections*, and two isomorphic sets are said to have the same *cardinality*. A *cardinal* is an isomorphism class of sets.
- Isomorphisms in Top are called *homeomorphisms*, and two isomorphic spaces are said to be *homeomorphic*.
- Isomorphisms in $h\mathsf{Top}$ are called *homotopy equivalences*, and isomorphic spaces in $h\mathsf{Top}$ are said to be *homotopic*. (See section 1.6 for a discussion of homotopy.)

Mathematics is particularly concerned with properties that are preserved under isomorphisms in a given category. For instance, topology is essentially the study of properties that are preserved under homeomorphisms. Such properties are referred to as *topological properties* and distinguish spaces: if X and Y are homeomorphic and X has (or doesn't have) a certain property, then Y must have (or cannot have) that property, too.

Example 0.4 The cardinality of a topological space is a topological property since any homeomorphism $f : X \to Y$ is necessarily a bijective function, and so X and Y, when viewed as sets, must have the same cardinality. *Metrizability* is also a topological property. *Connectedness* (section 2.1), *compactness* (section 2.3), *Hausdorff* (section 2.2), *first countability* (section 3.2), . . . are examples of other topological properties that will be discussed in the indicated sections.

However, not every familiar property is a topological one.

Example 0.5 A metric space is called *complete* if every Cauchy sequence converges. Being a complete metric space is *not* a topological property. For instance, the map $(-1, 1) \to \mathbb{R}$ by $x \mapsto \frac{x}{(1-x^2)}$ is a homeomorphism, yet $\mathbb{R}$ is a complete metric space while $(-1, 1)$ is not. This example also shows that being bounded is also not a topological property: a metric space is said to be *bounded* if the metric is a bounded function. Clearly, $(-1, 1)$ is bounded while $\mathbb{R}$ is not.

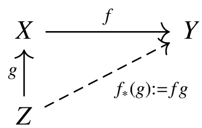
Comparing an object X with another object can help us to understand X better. Taking this idea further, we may also compare X with *all* objects at once. In other words, we can learn a great deal of information about X by looking at all morphisms both *out of* and *in to* X. This is the content of the next theorem. In essence, it states that the isomorphism class of an object is completely determined by morphisms to and from it. This is one of the main maxims of category theory:

an object is completely determined by its relationships with other objects.

In fact, it's a corollary of a major result to be discussed in section 0.2.3. But before we state the theorem formally, here is some useful terminology.

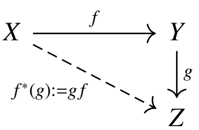
Definition 0.6 For each morphism $f\colon X \to Y$ and object Z in a category, there is a map of sets $f_*\colon \mathsf{C}(Z,X) \to \mathsf{C}(Z,Y)$ called the *pushforward* of f defined by postcomposition $f_*\colon g \mapsto fg$.

$$\mathsf{C}(Z,X) \xrightarrow{f_*} \mathsf{C}(Z,Y)$$



There is also a map of sets $f^*\colon \mathsf{C}(Y,Z) \to \mathsf{C}(X,Z)$ called the *pullback* defined by precomposition $f^*\colon g \mapsto gf$.

$$\mathsf{C}(X,Z) \xleftarrow{f^*} \mathsf{C}(Y,Z)$$



The reason for the visual names *pushforward* and *pullback* is hopefully clear from the diagrams above. With this terminology in hand, here is the theorem whose summary we provided above.

Theorem 0.1 The following are equivalent.

- $f\colon X \to Y$ is an isomorphism.
- For every object Z, the pushforward $f_*\colon \mathsf{C}(Z,X) \to \mathsf{C}(Z,Y)$ is an isomorphism of sets.
- For every object Z, the pullback $f^*\colon \mathsf{C}(Y,Z) \to \mathsf{C}(X,Z)$ is an isomorphism of sets.

Proof. We'll prove that a morphism $f\colon X \to Y$ is an isomorphism if and only if for every object Z, the pushforward $f_*\colon \mathsf{C}(Z,X) \to \mathsf{C}(Z,Y)$ is an isomorphism of sets; then, we'll leave the other statements as exercises.

Suppose $f\colon X \to Y$ is an isomorphism. Let $g\colon Y \to X$ be the inverse of f. Then for any Z, the map $g_*\colon \mathsf{C}(Z,Y) \to \mathsf{C}(Z,X)$ is the inverse of f_*.

Conversely, suppose that for any Z, the map $f_*\colon \mathsf{C}(Z,X) \to \mathsf{C}(Z,Y)$ is an isomorphism of sets. Choosing $Z = Y$, we have $f_*\colon \mathsf{C}(Y,X) \xrightarrow{\cong} \mathsf{C}(Y,Y)$, so in particular f_* is surjective. Therefore there exists a morphism $g\colon Y \to X$ so that $f_*g = \mathrm{id}_Y$, which by definition implies $fg = \mathrm{id}_Y$. To see that $gf = \mathrm{id}_X$, consider the case $Z = X$. We know $f_*\colon \mathsf{C}(X,X) \xrightarrow{\cong} \mathsf{C}(X,Y)$, so in particular f_* is injective. And since $f_*(\mathrm{id}_X) = f$ and $f_*(gf) = fgf = f$, injectivity of f_* implies $\mathrm{id}_X = gf$ as needed. □

In summary, if you understand all the morphisms $X \to Z$, then you know X up to isomorphism. Or if you understand all the morphisms $Z \to X$, then you know X up to isomorphism.

Now it turns out that categories themselves are objects worthy of study. And as per our maxim above, studying a category should amount to studying its relationships to other categories. But what is a relationship between categories? It is called a *functor*. That is the subject of the next section.

You might wonder if we are about to construct a *category of categories* where the objects are categories and the morphisms are functors. This can be done, but there are a few important considerations. One consideration is size. When we defined categories, we assumed that for any two objects X and Y, there is a *set* of morphisms $\mathsf{C}(X, Y)$. This is convenient so that, as in the case of the pushforward $f_*\colon \mathsf{C}(Z, X) \to \mathsf{C}(Z, Y)$ of a morphism $f\colon X \to Y$ in the previous theorem, we can consider functions between sets of morphisms. However, if C and D are categories, then there may be *more* than a set's worth of functors $\mathsf{C} \to \mathsf{D}$. There's also another, more subtle consideration to think about: The sharpest identification of an object generally available when doing category is *up to unique isomorphism*. This should apply to the morphisms between categories as well. It can be dizzying, but the takeaway is that invertible functors provide a notion of equivalence that's too rigid to be useful. So instead of an "isomorphism of categories," we talk about an "equivalence of categories," which is defined in a more subtle way. A third consideration is that one might want to remember that categories themselves have objects and morphisms, and the category of categories has more structure than a category—it's probably better to think of it as a "higher" category, a concept we won't stop to discuss here.

0.2.2 Functors

Definition 0.7 A *functor* F from a category C to a category D consists of the following data:

(i) An object FX of the category D for each object X in the category C,
(ii) A morphism $Ff\colon FX \to FY$ for every morphism $f\colon X \to Y$.

These data must be compatible with composition and identity morphisms in the following sense:

(iii) $(Fg)(Ff) = F(gf)$ for any morphisms $f\colon X \to Y$ and $g\colon Y \to Z$,
(iv) $F\,\mathrm{id}_X = \mathrm{id}_{FX}$ for any object X.

A functor as defined above is sometimes described as a *covariant* functor to distinguish it from its *contravariant* counterpart. A functor $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$ whose domain is the opposite category is called a *contravariant* functor from C to D. Contravariant functors "reverse arrows." That is, for every morphism $f\colon X \to Y$, the functor $F\colon \mathsf{C}^{\text{op}} \to \mathsf{D}$ assigns a morphism $Ff\colon FY \to FX$. We sometimes abuse terminology and refer to a contravariant functor simply as a functor. This usually doesn't cause confusion since the direction of the arrows can often be worked out easily.

Example 0.6 Here are some examples of functors.

- For an object X in a category C, there is a functor $\mathsf{C}(X,-)$ from C to Set that assigns to each object Z the set $\mathsf{C}(X,Z)$ and to each morphism $f\colon Y \to Z$ the pushforward f_* of f as in definition 0.6.

$$\begin{array}{ccc} Y & & \mathsf{C}(X,Y) \\ {\scriptstyle f}\big\downarrow & \longmapsto & \big\downarrow{\scriptstyle f_*} \\ Z & & \mathsf{C}(X,Z) \end{array}$$

- For an object X in a category C, there is a functor $\mathsf{C}(-,X)$ from C^{op} to Set that assigns to each object Z the set $\mathsf{C}(Z,X)$ and to each morphism $f\colon Y \to Z$ the pullback f^* of f as in definition 0.6.

$$\begin{array}{ccc} Y & & \mathsf{C}(Y,X) \\ {\scriptstyle f}\big\downarrow & \longmapsto & \big\uparrow{\scriptstyle f^*} \\ Z & & \mathsf{C}(Z,X) \end{array}$$

- Fix a set X. There is a functor $X\times -$ from Set to Set defined on objects by $Y \mapsto X\times Y$ and on morphisms f by $\text{id}\times f$. As we'll see in example 5.1, the functors $X\times -$ and $\mathsf{C}(X,-)$ form a special pair called the *product-hom adjunction.*
- Fix a vector space V over a field $\mathbf{k}$. There is a functor $V\otimes -$ from $\mathsf{Vect}_{\mathbf{k}}$ to $\mathsf{Vect}_{\mathbf{k}}$ defined on objects by $W \mapsto V\otimes W$ and on morphisms by $f \mapsto \text{id}\otimes f$.
- There is a *forgetful functor*, usually denoted U for "underlying," from Grp to Set that forgets the group operation. Concretely, it sends a group G to its underlying set UG and a group homomorphism to its underlying function.
- There is a *free functor* F from Set to Grp that assigns the free group FS to the set S. The free and forgetful functors form a special pair called the *free-forgetful adjunction*. We'll expand on this in section 5.2.
- There are other "forgetful" functors besides the one that forgets the group operation. Any functor that forgets structure, such as the $U:\mathsf{Top}\to\mathsf{Set}$ that forgets the topology, may be referred to as a forgetful functor.
- The *fundamental group*, to be discussed in detail in chapter 6, defines a functor π_1 from Top_* to Grp.
- The construction of the Grothendieck group of a commutative monoid is functorial. That is, there is a "Grothendieck group" functor from the category of commutative

monoids to the category of commutative groups that constructs a group from a commutative monoid by attaching inverses.

Here are a few different kinds of functors.

Definition 0.8 Let F be a functor from a category C to a category D. If for all objects X and Y in C, the map

$$\mathsf{C}(X, Y) \to \mathsf{D}(FX, FY) \qquad \text{given by} \qquad f \mapsto Ff$$

(i) is injective, then F is called *faithful*;
(ii) is surjective, then F is called *full*;
(iii) is bijective, then F is called *fully faithful*.

Fully faithful functors preserve all relationships among objects of the domain category: F is fully faithful if and only if each $FX \to FY$ is the image of *exactly one* morphism $X \to Y$. Think of it as an embedding, so to speak, of one category into another. But note, a fully faithful functor need not be injective on objects, so we use the term "embedding" loosely here. A fully faithful functor that *is* injective on objects is called a *full embedding*.

We'll define one of category theory's most famous fully faithful functors in the following section, but first we emphasize an important utility of functors. A functor $F\colon \mathsf{C} \to \mathsf{D}$ encodes invariants of isomorphism classes of objects within C. This is a fundamental idea, arising from the fact that functors take compositions to compositions and identities to identities. Consequentially,

functors take isomorphisms to isomorphisms.

So if two objects X and Y are "the same" in C, then FX and FY must be "the same" in D, the contrapositive being just as useful. For instance, the value that any functor $\mathsf{Top} \to \mathsf{C}$ assigns to a topological space is automatically a topological property (or we might say *a topological invariant*). So the question arises, "What is a useful choice of category C?" If we choose C to be an algebraic category, then we enter into the realm of *algebraic topology.* A *homology theory*, for example, is a functor $H\colon \mathsf{Top} \to R\mathsf{Mod}$ and is an excellent means of distinguishing topological spaces: if HX and HY are not isomorphic R-modules, then X and Y are not isomorphic spaces.

Having adopted the perspective that functors are invariants, it's natural to wonder: "When are two invariants the same?" To answer this, one needs a way to compare functors.

0.2.3 Natural Transformations and the Yoneda Lemma

Definition 0.9 Let F and G be functors $\mathsf{C} \to \mathsf{D}$. A *natural transformation* η from F to G consists of a morphism $\eta_X\colon FX \to GX$ for each object X in C. Moreover, these morphisms in D must satisfy the property that $\eta_Y Ff = Gf\eta_X$ for every morphism $f\colon X \to Y$ in C. In other words, the following diagram must commute:

$$\begin{array}{ccc} FX & \xrightarrow{Ff} & FY \\ {\scriptstyle \eta_X}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ GX & \xrightarrow{Gf} & GY \end{array}$$

For any two functors $F, G\colon \mathsf{C} \to \mathsf{D}$, let $\mathsf{Nat}(F, G)$ denote the natural transformations from F to G. If $\eta_X\colon FX \xrightarrow{\cong} GX$ is an isomorphism for each X, then η is called a *natural isomorphism* or a *natural equivalence*, and we say F and G are *naturally isomorphic*, denoted $F \cong G$.

Keep in mind that a natural transformation η is the *totality* of all the morphisms η_X, and each η_X can be thought of as a component, so to speak, of η. Simply put, a natural transformation is a collection of maps from one diagram to another, and these maps are special in that they *commute* with all the arrows in the diagrams.[3]

The language of natural transformations not only provides us an avenue in which comparing invariants (functors) becomes possible, but it also prompts us to revisit an idea mentioned in section 0.2.1. There we introduced the categorical philosophy that studying a mathematical object is more of a global, as opposed to a local, endeavor. That is, we can paint a better—rather, *a complete*—picture of an object once we investigate its interactions with all other objects. This theme finds its origins in an—if not *the most*—important result in category theory.

Yoneda Lemma For every object X in C and for every functor $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$, the set of natural transformations from $\mathsf{C}(-, X)$ to F is isomorphic to FX,

$$\mathsf{Nat}(\mathsf{C}(-, X), F) \cong FX$$

In other words, elements of the set FX are in bijection with natural transformations from $\mathsf{C}(-, X)$ to F. We omit the proof but take note of the special case when $F = \mathsf{C}(-, Y)$:

$$\mathsf{Nat}(\mathsf{C}(-, X), \mathsf{C}(-, Y)) \cong \mathsf{C}(X, Y) \tag{0.1}$$

It's closely related to theorem 0.1 in the following way.

First observe that given two categories C and D, we can form a new category D^{C} whose objects are functors $\mathsf{C} \to \mathsf{D}$ and whose morphisms are natural transformations. To ensure that D^{C} is locally small, we may require C to be small and D to be locally small. When considering contravariant functors and taking $D = \mathsf{Set}$, we obtain the category $\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$. An object here is called a *presheaf*. This is a very nice category—it has all finite limits and

[3] So a natural transformation may be viewed simultaneously as a single arrow between two functors or as a collection of arrows between two diagrams. Therefore you might hope that diagrams can be viewed as functors. As we will see in section 4.1, this is indeed the case.

colimits (to be discussed in chapter 4), it is Cartesian closed, and it forms what's called a topos. We won't dwell on these properties, though. Instead we turn our attention to a special functor $y\colon \mathsf{C} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$. It's defined by sending an object X to the presheaf $\mathsf{C}(-, X)$ and a morphism $f\colon X \to Y$ to the natural transformation f_*,

$$\begin{array}{ccc} X & & \mathsf{C}(-,X) \\ {\scriptstyle f}\downarrow & \longmapsto & \downarrow{\scriptstyle f_*} \\ Y & & \mathsf{C}(-,Y) \end{array}$$

This is a slight abuse of notation, though. By $f_*\colon \mathsf{C}(-, X) \to \mathsf{C}(-, Y)$ we mean the natural transformation whose component morphisms are the pushforward of f.

The isomorphism in (0.1) shows that the functor y is fully faithful and therefore *embeds* C into the functor category $\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$. For this reason, y is called the *Yoneda embedding*. The punchline is that each object X in C can be viewed as the contravariant functor $\mathsf{C}(-, X)$. Practically speaking, this means information about X can be obtained by studying the set of all morphisms in to it. But what about morphisms *out of* it?

There is a version of the Yoneda lemma for covariant functors (called *co-presheaves*) in the category $\mathsf{Set}^{\mathsf{C}}$ and, accordingly, a contravariant Yoneda embedding $\mathsf{C}^{\mathrm{op}} \to \mathsf{Set}^{\mathsf{C}}$. The corresponding result is that X can also be viewed as a functor $\mathsf{C}(X, -)$. So the moral of the story is that if you understand maps in and out of X, then you understand X. Thus we've come full circle to the theme introduced in section 0.2.1:

objects are completely determined by their relationships with other objects.

The adjective "completely" is justified by the following important corollary of the Yoneda lemma:

$$X \cong Y \qquad \text{if and only if} \qquad \mathsf{C}(-, X) \cong \mathsf{C}(-, Y).$$

One direction follows from the fact that the Yoneda embedding is a *functor*. The other direction follows from the fact that it is fully faithful. It's also true that $X \cong Y$ if and only if $\mathsf{C}(X, -) \cong \mathsf{C}(Y, -)$, as can be verified by considering the contravariant Yoneda embedding. And with that, we've just provided a restatement of theorem 0.1.

As a final remark, at times we'll adapt this philosophy and use it to consider just a few (and not necessarily all) morphisms to or from an object—this also yields fruitful information.

Example 0.7 Two sets X and Y are isomorphic if and only if $\mathsf{Set}(Z, X) \cong \mathsf{Set}(Z, Y)$ for all sets Z. That is, X and Y are the same if and only if they relate to all other sets in the same way. But this is overkill! Two sets are isomorphic if and only if they have the same cardinality, so to distinguish X and Y we need only look at the case when Z is the one-point set $*$. Indeed, a morphism $* \to X$ is a choice of element $x \in X$, and X and Y will have the same cardinality if and only if $\mathsf{Set}(*, X) \cong \mathsf{Set}(*, Y)$.

Perhaps it's not surprising that the full arsenal of the Yoneda lemma is not needed to distinguish sets. After all, they have no internal structure. The strength of lemma is more clearly seen when we look at objects such as groups or topological spaces that have more interesting aspects. But we include the example to draw attention to some notation you'll likely have noticed: we choose to omit the parentheses around the arguments of morphisms, for example, fx versus $f(x)$. This is consistent with category theory's frequent emphasis on morphisms, for as we've just seen, every element x in a set X can be viewed as a morphism $x\colon * \to X$. So given a function $f\colon X \to Y$, the image of x in Y is ultimately the composite morphism fx.

Keeping in tune with the previous paragraph, we now proceed to a discussion on basic set theory by recasting familiar material in a more categorical light.

0.3 Basic Set Theory

We'll begin with a brief review of functions.

0.3.1 Functions

A function is said to be *injective* if and only if it is *left cancellative*. That is, $f\colon X \to Y$ is injective if and only if for all functions $g_1, g_2\colon Z \to X$ with $fg_1 = fg_2$, it follows that $g_1 = g_2$. Equivalently, f is injective if and only if $f_*\colon \mathsf{Set}(Z, X) \to \mathsf{Set}(Z, Y)$ is injective for all Z. Yet another equivalent definition is that f is injective if and only if it has a left inverse, that is, if and only if there exists $g\colon Y \to X$ so that $gf = \mathrm{id}_Y$. Note that the composition of injective functions is injective; also, for any $f\colon X \to Y$ and $g\colon Y \to Z$, if gf is injective, then f is injective. We will denote injective functions by hooked arrows, as in $f\colon X \hookrightarrow Y$.

More generally, left-cancellative morphisms in any category are called *monomorphisms* or said to be *monic* and are denoted with arrows with tails as in $X \rightarrowtail Y$. In this more general case, left invertible implies left cancellative, but not conversely. For example, the map $n \mapsto 2n$ defines a left-cancellative group homomorphism $f\colon \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z}$. However, there is no group homomorphism $g\colon \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ so that $gf = \mathrm{id}_{\mathbb{Z}/2\mathbb{Z}}$.

A function is said to be *surjective* if and only if it is *right cancellative*. That is, $f\colon X \to Y$ is *surjective* if and only if for all functions $g_1, g_2\colon Y \to Z$ with $g_1 f = g_2 f$, it follows that $g_1 = g_2$. Equivalently, f is surjective if and only if $f^*\colon \mathsf{Set}(Y, Z) \to \mathsf{Set}(X, Z)$ is injective for all Z or, equivalent still, if and only if it has a right inverse. That is, $f\colon X \to Y$ is surjective if and only if there exists $g\colon Y \to X$ so that $fg = \mathrm{id}_X$. The composition of surjective functions is surjective, and for any $f\colon X \to Y$ and $g\colon Y \to Z$, if gf is surjective, then g is surjective. Surjective functions will be denoted with two headed arrows as in $f\colon X \twoheadrightarrow Y$.

In general, right-cancellative morphisms in any category are called *epimorphisms* or said to be *epic* and are also denoted with two-headed arrows. Right invertible implies right cancellative in any category, but not conversely. (You are asked to provide an example in exercise 0.3 at the end of the chapter.)

Finally, note that in Set a function that is both injective and surjective is an isomorphism. This is because left invertible and right invertible imply invertible. (You should check that having a left inverse and a right inverse both imply there's a single two-sided inverse). But left cancellative and right cancellative together do not imply invertible: there are categories—and Top is one of them—that have morphisms that are both monic and epic, which nonetheless fail to be isomorphisms.

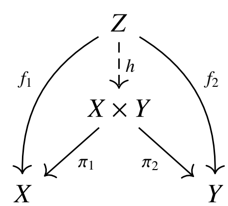
0.3.2 The Empty Set and One-Point Set

The empty set $\varnothing$ is *initial* in Set. That is, for any set X there is a unique function $\varnothing \to X$. On the other hand, the one-point set $*$ is *terminal*. That is, for any set X, there is a unique function $X \to *$. You might take issue with the definite article "the" in "the one-point set," but it is standard to use the definite article in circumstances that are *unique up to unique isomorphism*. That is the case here: if $*$ and $*'$ are both one-point sets, then there is a unique isomorphism $* \xrightarrow{\cong} *'$. Note also that this terminology is not unique to Set: we can make sense of initial and terminal objects in any category, though such objects may not always exist. An object C in a category C is called terminal if for every object X in C there is a unique morphism $X \to C$. Dually, an object D is called initial if for every object X in C there is a unique morphism $D \to X$.

0.3.3 Products and Coproducts in Set

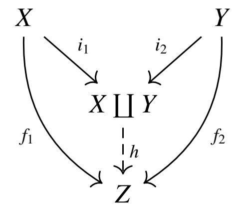
The *Cartesian product* of two sets X and Y is defined to be the set $X \times Y$ of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. While this tells us what the set $X \times Y$ *is*, it doesn't say much about the properties that it possesses or how it relates to other sets in the category. This prompts us to look for a more categorical description of the Cartesian product.

The Cartesian product of two sets X and Y is a set $X \times Y$ that comes with maps $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$. The product is characterized by the property that for any set Z and any functions $f_1 : Z \to X$ and $f_2 : Z \to Y$, there is a unique map $h : Z \to X \times Y$ with $\pi_1 h = f_1$ and $\pi_2 h = f_2$. By "characterized by," we mean that the Cartesian product is the *unique* (up to isomorphism) object in Set with this property.



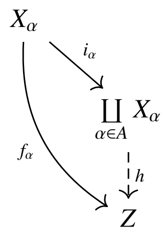
As an example, note the isomorphism of finite sets $\{1, \ldots, n\} \times \{1, \ldots, m\} \cong \{1, \ldots, nm\}$.

The *disjoint union* of two sets X and Y also has a categorical description. It is a set $X \coprod Y$ that comes with maps $i_1 \colon X \to X \coprod Y$ and $i_2 \colon Y \to X \coprod Y$ and is characterized by the property that for any set Z and any functions $f_1 \colon X \to Z$ and $f_2 \colon Y \to Z$, there is a unique map $h \colon X \coprod Y \to Z$ with $hi_1 = f_1$ and $hi_2 = f_2$.

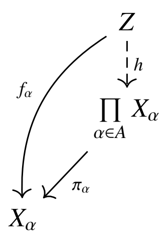


Sometimes, disjoint union is called the sum and is denoted $X + Y$ or $X \oplus Y$ instead of $X \coprod Y$. As an example, note that $\{1, \ldots, n\} + \{1, \ldots, m\} \cong \{1, \ldots, n + m\}$. The property characterizing the disjoint union is dual to the one characterizing the product, and the disjoint union is sometimes called the *coproduct* of sets.

We can also take products and disjoint unions of arbitrary collections of sets. The disjoint union of a collection of sets $\{X_\alpha\}_{\alpha \in A}$ is a set $\coprod_{\alpha \in A} X_\alpha$ together with maps $i_\alpha \colon X_\alpha \to \coprod X_\alpha$ satisfying the property that for any set Z and any collection of functions $\{f_\alpha \colon X_\alpha \to Z\}$, there is a unique map $h \colon \coprod X_\alpha \to Z$ with $hi_\alpha = f_\alpha$ for all $\alpha \in A$.



The product of a collection of sets $\{X_\alpha\}_{\alpha \in A}$ is sometimes described as the subset of functions $f \colon A \to \coprod X_\alpha$ satisfying $f(\alpha) \in X_\alpha$. But, as we mentioned earlier, what's more important than the construction of the product is to understand its universal property. The product of a collection of sets $\{X_\alpha\}_{\alpha \in A}$ is a set $\prod_{\alpha \in A} X_\alpha$ together with maps $\pi_\alpha \colon \prod X_\alpha \to X_\alpha$ characterized by the property that for any set Z and any collection of functions $\{f_\alpha \colon Z \to X_\alpha\}$, there is a unique map $h \colon Z \to \prod X_\alpha$ with $\pi_\alpha h = f_\alpha$ for all $\alpha \in A$.



0.3.4 Products and Coproducts in Any Category

The universal properties outlined above provide a template for defining products and coproducts—and more generally limits and colimits—in any category. A more complete discussion is found in chapter 4, but for now, we'd like to point out that in an arbitrary category, products and coproducts may not exist, and when they do they might not look like disjoint unions or Cartesian products. For example, the category Fld of fields doesn't have products: if there were a field $\mathbf{k}$ that were the product of $\mathbb{F}_2$ and $\mathbb{F}_3$, then there would be homomorphisms $\mathbf{k} \to \mathbb{F}_2$ and $\mathbf{k} \to \mathbb{F}_3$. But this is impossible since the characteristic of $\mathbf{k}$ would be equal to both 2 and 3. On the other hand the category $\mathsf{Vect}_\mathbf{k}$ and more generally $R\mathsf{Mod}$ has both products and coproducts. Products are Cartesian products, but coproducts are direct sums. In Grp coproducts are free products, while in the category of abelian groups, coproducts are direct sums. Even in the category Set, there is something to say about the existence of products and coproducts. The *axiom of choice* is precisely the statement that for any nonempty collection of sets $\{X_\alpha\}_{\alpha\in A}$, the product $\prod X_\alpha$ exists and is nonempty. (For more on the axiom of choice, see section 3.4.)

Even though products and coproducts in an arbitrary category might look different than they do in Set, the constructions are closely related to the constructions in Set. That's because the universal properties of products and coproducts in an arbitrary category C yield bijections of sets

$$\mathsf{C}\left(\textstyle\coprod_\alpha X_\alpha, Z\right) \cong \textstyle\prod_\alpha \mathsf{C}\left(X_\alpha, Z\right) \qquad \Big| \qquad \mathsf{C}\left(Z, \textstyle\prod_\alpha X_\alpha\right) \cong \textstyle\prod_\alpha \mathsf{C}\left(Z, X_\alpha\right)$$

In other words, the product is characterized by the fact that maps *into* it are in bijection with maps into each of the factors. Dually, the coproduct is characterized by the fact that maps *out of* it are in bijection with maps out of each of the components. (In a sense that can be made precise, products and coproducts in a category are [co]representations of products and coproducts of sets.) To phrase it another way, coproducts come out of the first entry of hom as products, and products come out of the second entry of hom as products. An example where this comes up often is in $\mathsf{Vect}_\mathbf{k}$ and $R\mathsf{Mod}$ where the coproduct is direct sum and the product is Cartesian product. For an R module X, let $X^* := R\mathsf{Mod}(X, R)$ denote the dual space. Then setting $Z = R$ in the first isomorphism above yields

$$\left(\bigoplus X_\alpha\right)^* \quad\cong\quad \prod (X_\alpha)^*$$

So the fact that the "*the dual of the sum is the product of the duals*" is a consequence of the existence of coproducts in $R\mathsf{Mod}$.

0.3.5 Exponentiation in Set

In the category of sets, the set of morphisms $\mathsf{Set}(X, Y)$ is also denoted Y^X. Moreover there is a natural *evaluation* map eval $X\times\colon Y^X \to Y$ defined by $\mathrm{eval}(x, f) = fx$. The exponential notation is convenient for expressing various isomorphisms, such as

$$(X \times Y)^Z \quad \cong \quad X^Z \times Y^Z$$

which is a concise way to express the universal property of the Cartesian product: maps from a set Z into a product correspond to maps from Z into the factors. There is also the isomorphism

$$Y^{X\times Z} \quad \cong \quad (Y^X)^Z$$

which is an expression of the *product-hom adjunction*. We made brief mention of this adjunction in example 0.6 and will discuss it in greater detail in example 5.1. But here's a sneak preview. Fix a set X. Let L be the functor $X \times -$ and let R be the functor $\mathsf{Set}(X, -)$. In this notation, the isomorphism $Y^{X\times Z} \cong (Y^X)^Z$ becomes $\mathsf{Set}(LZ, Y) \cong \mathsf{Set}(Z, RY)$, which evokes the defining property of *adjoint* linear maps. (Hence the term "adjunction.")

0.3.6 Partially Ordered Sets

A partially ordered set or *poset* is a set $\mathcal{P}$ together with a relation $\leq$ on $\mathcal{P}$ that is reflexive, transitive, and antisymmetric. Reflexive means that for all $a \in \mathcal{P}$, $a \leq a$; transitive means that for all $a, b, c \in \mathcal{P}$, if $a \leq b$ and $b \leq c$, then $a \leq c$; antisymmetric means that for all $a, b \in P$, if $a \leq b$ and $b \leq a$, then $a = b$.

One can view a poset as a category whose objects are the elements of $\mathcal{P}$ by declaring there to be a morphism $a \to b$ if and only if $a \leq b$. Transitivity says not only can composition be defined, but it may only be defined in *one* way since there's at most one morphism between objects. Alternatively, one can define a poset to be a (small) category with the property that there's at most one morphism between objects.

We expect that you have encountered different definitions for some of the concepts in this section than what we have provided. In your previous work, an injective function, for example, might have been defined in terms of what it does to the elements of the domain. Here, instead, we've defined an injective function in terms of how it interacts with other functions. For another example, instead of defining what the disjoint $X \coprod Y$ *is* (which ultimately involves the Zermelo-Fraenkel axioms of union and extension), we've characterized it up to isomorphism by explaining how it interacts with other sets, bringing back to mind the Lawvere quote that opened the chapter.

Exercises

1. Suppose $\mathcal{S}$ is a collection of subsets of X whose union equals X. Prove there is a coarsest topology $\mathcal{T}$ containing $\mathcal{S}$ and that the collection of all finite intersections of sets in $\mathcal{S}$ is a basis for $\mathcal{T}$. In this situation, the collection $\mathcal{S}$ is called *a subbasis* for the topology $\mathcal{T}$.

2. Prove that a function $f\colon X \to Y$ between topological spaces is continuous if and only if $f^{-1}B$ is open for every B in a basis for the topology on Y.

3. Here are some examples and short exercises about morphisms.

a) Prove that left-invertible morphisms are monic and right-invertible morphisms are epic.

b) Give an example of an epimorphism which is not right invertible.

c) Prove that if a morphism is left invertible and right invertible, then it is invertible.

d) Give an example of a morphism in Top that is epic and monic but not an isomorphism.

e) In some category, give an example of two objects X and Y that are not isomorphic but which nonetheless have monomorphisms: $X \rightleftarrows Y$.

4. Discuss the initial object, the terminal object, products, and coproducts in the categories Grp and $\mathsf{Vect}_{\mathbf{k}}$.

5. Prove the other part of theorem 0.1. That is, prove that $f\colon X \to Y$ is an isomorphism in a category C if and only if $f_*\colon \mathsf{C}(Z, X) \to \mathsf{C}(Z, Y)$ is an isomorphism for every object Z.

6. Prove the Yoneda lemma in section 0.2.3. The key is to observe that $\mathsf{C}(X, X)$ has a special element, namely id_X. So, for any natural transformation $\eta\colon \mathsf{C}(-, X) \to F$, one obtains a special element $\eta\,\mathrm{id}_X \in FX$, which completely determines η.