CHAPTER 1

Categories, Functors, Natural Transformations

Frequently in modern mathematics there occur phenomena of “naturality”.

Samuel Eilenberg and Saunders Mac Lane,
“Natural isomorphisms in group theory”
[EM42b]

A group extension of an abelian group \( H \) by an abelian group \( G \) consists of a group \( E \) together with an inclusion of \( G \hookrightarrow E \) as a normal subgroup and a surjective homomorphism \( E \twoheadrightarrow H \) that displays \( H \) as the quotient group \( E/G \). This data is typically displayed in a diagram of group homomorphisms:

\[
0 \to G \to E \to H \to 0. \tag{1.0.1}
\]

A pair of group extensions \( E \) and \( E' \) of \( G \) and \( H \) are considered to be equivalent whenever there is an isomorphism \( E \cong E' \) that commutes with the inclusions of \( G \) and quotient maps to \( H \), in a sense that is made precise in §1.6. The set of equivalence classes of abelian group extensions \( E \) of \( H \) by \( G \) defines an abelian group \( \text{Ext}(H, G) \).

In 1941, Saunders Mac Lane gave a lecture at the University of Michigan in which he computed for a prime \( p \) that \( \text{Ext}(\mathbb{Z}[\frac{1}{p}]/\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}_p \), the group of \( p \)-adic integers, where \( \mathbb{Z}[\frac{1}{p}]/\mathbb{Z} \) is the Prüfer \( p \)-group. When he explained this result to Samuel Eilenberg, who had missed the lecture, Eilenberg recognized the calculation as the homology of the 3-sphere complement of the \( p \)-adic solenoid, a space formed as the infinite intersection of a sequence of solid tori, each wound around \( p \) times inside the preceding torus. In teasing apart this connection, the pair of them discovered what is now known as the universal coefficient theorem in algebraic topology, which relates the homology \( H_\ast \) and cohomology groups \( H^\ast \) associated to a space \( X \) via a group extension [ML05]:

\[
0 \to \text{Ext}(H_{n-1}(X), G) \to H^n(X, G) \to \text{Hom}(H_n(X), G) \to 0.
\]

To obtain a more general form of the universal coefficient theorem, Eilenberg and Mac Lane needed to show that certain isomorphisms of abelian groups expressed by this group extension extend to spaces constructed via direct or inverse limits. And indeed this is the case, precisely because the homomorphisms in the diagram (1.0.1) are natural with respect to continuous maps between topological spaces.

The adjective “natural” had been used colloquially by mathematicians to mean “defined without arbitrary choices.” For instance, to define an isomorphism between a finite-dimensional vector space \( V \) and its dual, the vector space of linear maps from \( V \) to the

\[^{1}\text{The zeros appearing on the ends provide no additional data. Instead, the first zero implicitly asserts that the map } G \to E \text{ is an inclusion and the second that the map } E \to H \text{ is a surjection. More precisely, the displayed sequence of group homomorphisms is exact, meaning that the kernel of each homomorphism equals the image of the preceding homomorphism.}\]
ground field $k$, requires a choice of basis. However, there is an isomorphism between $V$ and its double dual that requires no choice of basis; the latter, but not the former, is natural.

To give a rigorous proof that their particular family of group isomorphisms extended to inverse and direct limits, Eilenberg and Mac Lane sought to give a mathematically precise definition of the informal concept of “naturality.” To that end, they introduced the notion of a natural transformation, a parallel collection of homomorphisms between abelian groups in this instance. To characterize the source and target of a natural transformation, they introduced the notion of a functor. And to define the source and target of a functor in the greatest generality, they introduced the concept of a category. This work, described in “The general theory of natural equivalences” [EM45], published in 1945, marked the birth of category theory.

While categories and functors were first conceived as auxiliary notions, needed to give a precise meaning to the concept of naturality, they have grown into interesting and important concepts in their own right. Categories suggest a particular perspective to be used in the study of mathematical objects that pays greater attention to the maps between them. Functors, which translate mathematical objects of one type into objects of another, have a more immediate utility. For instance, the Brouwer fixed point theorem translates a seemingly intractable problem in topology to a trivial one ($0 \not= 1$) in algebra. It is to these topics that we now turn.

Categories are introduced in §1.1 in two guises: firstly as universes categorizing mathematical objects and secondly as mathematical objects in their own right. The first perspective is used, for instance, to define a general notion of isomorphism that can be specialized to mathematical objects of every conceivable variety. The second perspective leads to the observation that the axioms defining a category are self-dual. Thus, as explored in §1.2, for any proof of a theorem about all categories from these axioms, there is a dual proof of the dual theorem obtained by a syntactic process that is interpreted as “turning around all the arrows.”

Functors and natural transformations are introduced in §1.3 and §1.4 with examples intended to shed light on the linguistic and practical utility of these concepts. The category-theoretic notions of isomorphism, monomorphism, and epimorphism are invariant under certain classes of functors, including in particular the equivalences of categories, introduced in §1.5. At a high level, an equivalence of categories provides a precise expression of the intuition that mathematical objects of one type are “the same as” objects of another variety: an equivalence between the category of matrices and the category of finite-dimensional vector spaces equates high school and college linear algebra.

In addition to providing a new language to describe emerging mathematical phenomena, category theory also introduced a new proof technique: that of the diagram chase. The introduction to the influential book [ES52] presents commutative diagrams as one of the “new techniques of proof” appropriate for their axiomatic treatment of homology theory. The technique of diagram chasing is introduced in §1.6 and applied in §1.7 to construct new natural transformations as horizontal or vertical composites of given ones.

---

2A brief account of functors and natural isomorphisms in group theory appeared in a 1942 paper [EM42b].
3As is the case for the duality in projective plane geometry, this duality can be formulated precisely as a feature of the first-order theories that axiomatize these structures.
1.1. Abstract and concrete categories

It frames a possible template for any mathematical theory: the theory should have nouns and verbs, i.e., objects, and morphisms, and there should be an explicit notion of composition related to the morphisms; the theory should, in brief, be packaged by a category.

Barry Mazur, “When is one thing equal to some other thing?” [Maz08]

Definition 1.1.1. A category consists of

- a collection of objects $X, Y, Z, \ldots$
- a collection of morphisms $f, g, h, \ldots$

so that:

- Each morphism has specified domain and codomain objects; the notation $f : X \rightarrow Y$ signifies that $f$ is a morphism with domain $X$ and codomain $Y$.
- Each object has a designated identity morphism $1_X : X \rightarrow X$.
- For any pair of morphisms $f, g$ with the codomain of $f$ equal to the domain of $g$, there exists a specified composite morphism $g \cdot f$ whose domain is equal to the domain of $f$ and whose codomain is equal to the codomain of $g$, i.e.:

\[
f : X \rightarrow Y, \quad g : Y \rightarrow Z \quad \Rightarrow \quad g \cdot f : X \rightarrow Z.
\]

This data is subject to the following two axioms:

- For any $f : X \rightarrow Y$, the composites $1_Y f$ and $f 1_X$ are both equal to $f$.
- For any composable triple of morphisms $f, g, h$, the composites $h(gf)$ and $(hg)f$ are equal and henceforth denoted by $hgf$.

\[
f : X \rightarrow Y, \quad g : Y \rightarrow Z, \quad h : Z \rightarrow W \quad \Rightarrow \quad hgf : X \rightarrow W.
\]

That is, the composition law is associative and unital with the identity morphisms serving as two-sided identities.

Remark 1.1.2. The objects of a category are in bijective correspondence with the identity morphisms, which are uniquely determined by the property that they serve as two-sided identities for composition. Thus, one can define a category to be a collection of morphisms with a partially-defined composition operation that has certain special morphisms, which are used to recognize composable pairs and which serve as two-sided identities; see [Ehr65, §1.1] or [FS90, §1.1]. But in practice it is not so hard to specify both the objects and the morphisms and this is what we shall do.

It is traditional to name a category after its objects; typically, the preferred choice of accompanying structure-preserving morphisms is clear. However, this practice is somewhat contrary to the basic philosophy of category theory: that mathematical objects should always be considered in tandem with the morphisms between them. By Remark 1.1.2, the algebra of morphisms determines the category, so of the two, the objects and morphisms, the morphisms take primacy.

Example 1.1.3. Many familiar varieties of mathematical objects assemble into a category.

\footnote{The composite may be written less concisely as $g \cdot f$ when this adds typographical clarity.}
1. CATEGORIES, FUNCTORS, NATURAL TRANSFORMATIONS

(i) \( \text{Set} \) has sets as its objects and functions, with specified domain and codomain,\(^5\) as its morphisms.
(ii) \( \text{Top} \) has topological spaces as its objects and continuous functions as its morphisms.
(iii) \( \text{Set} \) and \( \text{Top} \) have sets or spaces with a specified basepoint\(^6\) as objects and basepoint-preserving (continuous) functions as morphisms.
(iv) \( \text{Group} \) has groups as objects and group homomorphisms as morphisms. This example lent the general term “morphisms” to the data of an abstract category. The categories \( \text{Ring} \) of associative and unital rings and ring homomorphisms and \( \text{Field} \) of fields and field homomorphisms are defined similarly.
(v) For a fixed unital but not necessarily commutative ring \( R \), \( \text{Mod}_R \) is the category of left \( R \)-modules and \( R \)-module homomorphisms. This category is denoted by \( \text{Vec}_k \) when the ring happens to be a field \( k \) and abbreviated as \( \text{Ab} \) in the case of \( \text{Mod}_Z \), as a \( Z \)-module is precisely an abelian group.
(vi) \( \text{Graph} \) has graphs as objects and graph morphisms (functions carrying vertices to vertices and edges to edges, preserving incidence relations) as morphisms. In the variant \( \text{DirGraph} \), objects are directed graphs, whose edges are now depicted as arrows, and morphisms are directed graph morphisms, which must preserve sources and targets.
(vii) \( \text{Man} \) has smooth (i.e., infinitely differentiable) manifolds as objects and smooth maps as morphisms.
(viii) \( \text{Meas} \) has measurable spaces as objects and measurable functions as morphisms.
(ix) \( \text{Poset} \) has partially-ordered sets as objects and order-preserving functions as morphisms.
(x) \( \text{Ch}_R \) has chain complexes of \( R \)-modules as objects and chain homomorphisms as morphisms.\(^7\)
(xi) For any signature \( \sigma \), specifying constant, function, and relation symbols, and for any collection of well formed sentences \( T \) in the first-order language associated to \( \sigma \), there is a category \( \text{Model}_T \) whose objects are \( \sigma \)-structures that model \( T \), i.e., sets equipped with appropriate constants, relations, and functions satisfying the axioms \( T \). Morphisms are functions that preserve the specified constants, relations, and functions, in the usual sense.\(^8\) Special cases include (iv), (v), (vi), (ix), and (x).

The preceding are all examples of concrete categories, those whose objects have underlying sets and whose morphisms are functions between these underlying sets, typically the “structure-preserving” morphisms. A more precise definition of a concrete category is given in 1.6.17. However, “abstract” categories are also prevalent:

Example 1.1.4.

(i) For a unital ring \( R \), \( \text{Mat}_R \) is the category whose objects are positive integers and in which the set of morphisms from \( n \) to \( m \) is the set of \( m \times n \) matrices with values in

\(^5\)[EM45, p. 239] emphasizes that the data of a function should include specified sets of inputs and potential outputs, a perspective that was somewhat radical at the time.
\(^6\)A basepoint is simply a chosen distinguished point in the set or space.
\(^7\)A chain complex \( C \) is a collection \((C_n)_{n \in \mathbb{Z}}\) of \( R \)-modules equipped with \( R \)-module homomorphisms \( d : C_n \to C_{n-1} \), called boundary homomorphisms, with the property that \( d^2 = 0 \), i.e., the composite of any two boundary maps is the zero homomorphism. A map of chain complexes \( f : C \to C' \) is comprised of a collection of homomorphisms \( f_n : C_n \to C'_n \) so that \( df_n = f_{n-1}d \) for all \( n \in \mathbb{Z} \).
\(^8\)Model theory pays greater attention to other types of morphisms, for instance the elementary embeddings, which are (automatically injective) functions that preserve and reflect satisfaction of first-order formulae.
R. Composition is by matrix multiplication

\[ n \xrightarrow{A} m, \quad m \xrightarrow{B} k \xrightarrow{A} k \]

with identity matrices serving as the identity morphisms.

(ii) A group \( G \) (or, more generally, a monoid\(^9\)) defines a category \( BG \) with a single object. The group elements are its morphisms, each group element representing a distinct endomorphism of the single object, with composition given by multiplication. The identity element \( e \in G \) acts as the identity morphism for the unique object in this category.

\[
\mathcal{B}S_3 = \begin{array}{c}
\circ \\
(12) \\
(13) \\
\uparrow \\
(132) \\
\downarrow \\
(23)
\end{array}
\]

(iii) A poset \((P, \leq)\) (or, more generally, a preorder\(^{10}\)) may be regarded as a category. The elements of \( P \) are the objects of the category and there exists a unique morphism \( x \to y \) if and only if \( x \leq y \). Transitivity of the relation \( \leq \) implies that the required composite morphisms exist. Reflexivity implies that identity morphisms exist.

(iv) In particular, any ordinal \( \alpha = \{ \beta \mid \beta < \alpha \} \) defines a category whose objects are the smaller ordinals. For example, \( \emptyset \) is the category with no objects and no morphisms. \( \{ \} \) is the category with a single object and only its identity morphism. \( 2 \) is the category with two objects and a single non-identity morphism, conventionally depicted as \( 0 \to 1 \). \( \omega \) is the category \textit{freely generated by the graph}

\[
0 \to 1 \to 2 \to 3 \to \cdots
\]

in the sense that every non-identity morphism can be uniquely factored as a composite of morphisms in the displayed graph; a precise definition of the notion of free generation is given in Example 4.1.13.

(v) A set may be regarded as a category in which the elements of the set define the objects and the only morphisms are the required identities. A category is \textit{discrete} if every morphism is an identity.

(vi) \( \text{Htpy} \), like \( \text{Top} \), has spaces as its objects but morphisms are homotopy classes of continuous maps. \( \text{Htpy} \) has based spaces as its objects and basepoint-preserving homotopy classes of based continuous maps as its morphisms.

(vii) \( \text{Measure} \) has measure spaces as objects. One reasonable choice for the morphisms is to take equivalence classes of measurable functions, where a parallel pair of functions are equivalent if their domain of difference is contained within a set of measure zero.

Thus, the philosophy of category theory is extended. The categories listed in Example 1.1.3 suggest that mathematical objects ought to be considered together with the appropriate notion of morphism between them. The categories listed in Example 1.1.4 illustrate that

---

\(^9\)A \textit{monoid} is a set \( M \) equipped with an associative binary multiplication operation \( M \times M \to M \) and an identity element \( e \in M \) serving as a two-sided identity. In other words, a monoid is precisely a one-object category.

\(^{10}\)A \textit{preorder} is a set with a binary relation \( \leq \) that is reflexive and transitive. In other words, a preorder is precisely a category in which there are no parallel pairs of distinct morphisms between any fixed pair of objects. A \textit{poset} is a preorder that is additionally antisymmetric: \( x \leq y \) and \( y \leq x \) implies that \( x = y \).
these morphisms are not always functions.\textsuperscript{11} The morphisms in a category are also called arrows or maps, particularly in the contexts of Examples 1.1.4 and 1.1.3, respectively.

**Remark 1.1.5.** Russell’s paradox implies that there is no set whose elements are “all sets.” This is the reason why we have used the vague word “collection” in Definition 1.1.1. Indeed, in each of the examples listed in 1.1.3, the collection of objects is not a set. Eilenberg and Mac Lane address this potential area of concern as follows:

... the whole concept of a category is essentially an auxiliary one; our basic concepts are essentially those of a functor and of a natural transformation ... . The idea of a category is required only by the precept that every function should have a definite class as domain and a definite class as range, for the categories are provided as the domains and ranges of functors. Thus one could drop the category concept altogether and adopt an even more intuitive standpoint, in which a functor such as “Hom” is not defined over the category of “all” groups, but for each particular pair of groups which may be given. [EM45]

The set-theoretical issues that confront us while defining the notion of a category will compound as we develop category theory further. For that reason, common practice among category theorists is to work in an extension of the usual Zermelo–Fraenkel axioms of set theory, with new axioms allowing one to distinguish between “small” and “large” sets, or between sets and classes. The search for the most useful set-theoretical foundations for category theory is a fascinating topic that unfortunately would require too long of a digression to explore.\textsuperscript{12} Instead, we sweep these foundational issues under the rug, not because these issues are not serious or interesting, but because they distract from the task at hand.\textsuperscript{13}

For the reasons just discussed, it is important to introduce adjectives that explicitly address the size of a category.

**Definition 1.1.6.** A category is small if it has only a set’s worth of arrows.

By Remark 1.1.2, a small category has only a set’s worth of objects. If $C$ is a small category, then there are functions $\text{mor} \ C \xrightarrow{\text{dom}} \text{dom} C \xleftarrow{\text{cod}} \text{cod} C \xrightarrow{\text{ob} C}$ that send a morphism to its domain and its codomain and an object to its identity.

\textsuperscript{11}Reid’s *Undergraduate algebraic geometry* emphasizes that the morphisms are not always functions, writing “Students who disapprove are recommended to give up at once and take a reading course in category theory instead” [Rei88, p. 4].

\textsuperscript{12}The preprint [Shu08] gives an excellent overview, though it is perhaps better read after Chapters 1–4.

\textsuperscript{13}If pressed, let us assume that there exists a countable sequence of inaccessible cardinals, meaning uncountable cardinals that are regular and strong limit. A cardinal $\kappa$ is regular if every union of fewer than $\kappa$ sets each of cardinality less than $\kappa$ has cardinality less than $\kappa$, and strong limit if $\lambda < \kappa$ implies that $2^\lambda < \kappa$. Inaccessibility means that sets of size less than $\kappa$ are closed under power sets and $\kappa$-small unions. If $\kappa$ is inaccessible, then the $\kappa$-stage of the von Neumann hierarchy, the set $V_\kappa$ of sets of rank less than $\kappa$, is a model of Zermelo–Fraenkel set theory with choice (ZFC); the set $V_\kappa$ is a Grothendieck universe. The assumption that there exists a countable sequence of inaccessible cardinals means that we can “do set theory” inside the universe $V_\kappa$, and then enlarge the universe if necessary as often as needed.

If ZFC is consistent, these axioms cannot prove the existence of an inaccessible cardinal or the consistency of the assumption that one exists (by Gödel’s second incompleteness theorem). Nonetheless, from the perspective of the hierarchy of large cardinal axioms, the existence of inaccessibles is a relatively mild hypothesis.
None of the categories in Example 1.1.3 are small—each has too many objects—but “locally” they resemble small categories in a sense made precise by the following notion:

Definition 1.1.7. A category is **locally small** if between any pair of objects there is only a set’s worth of morphisms.

It is traditional to write

\[(1.1.8) \quad \mathcal{C}(X, Y) \quad \text{or} \quad \text{Hom}(X, Y)\]

for the set of morphisms from \(X\) to \(Y\) in a locally small category \(\mathcal{C}\).\(^{14}\) The set of arrows between a pair of fixed objects in a locally small category is typically called a **hom-set**, whether or not it is a set of “homomorphisms” of any particular kind. Because the notation \((1.1.8)\) is so convenient, it is also adopted for the collection of morphisms between a fixed pair of objects in a category that is not necessarily locally small.

A category provides a context in which to answer the question “When is one thing the same as another thing?” Almost universally in mathematics, one regards two objects of the same category to be “the same” when they are isomorphic, in a precise categorical sense that we now introduce.

Definition 1.1.9. An **isomorphism** in a category is a morphism \(f : X \to Y\) for which there exists a morphism \(g : Y \to X\) so that \(gf = 1_X\) and \(fg = 1_Y\). The objects \(X\) and \(Y\) are **isomorphic** whenever there exists an isomorphism between \(X\) and \(Y\), in which case one writes \(X \cong Y\).

An **endomorphism**, i.e., a morphism whose domain equals its codomain, that is an isomorphism is called an **automorphism**.

Example 1.1.10.

(i) The isomorphisms in \(\text{Set}\) are precisely the **bijections**.

(ii) The isomorphisms in \(\text{Group}, \text{Ring}, \text{Field}, \text{or Mod}_R\) are the bijective homomorphisms.

(iii) The isomorphisms in the category \(\text{Top}\) are the **homeomorphisms**, i.e., the continuous functions with continuous inverse, which is a stronger property than merely being a bijective continuous function.

(iv) The isomorphisms in the category \(\text{Htpy}\) are the **homotopy equivalences**.

(v) In a poset \((P, \leq)\), the axiom of antisymmetry asserts that \(x \leq y\) and \(y \leq x\) imply that \(x = y\). That is, the only isomorphisms in the category \((P, \leq)\) are identities.

Examples 1.1.10(ii) and (iii) suggest the following general question: In a concrete category, when are the isomorphisms precisely those maps in the category that induce bijections between the underlying sets? We will see an answer in Lemma 5.6.1.

Definition 1.1.11. A **groupoid** is a category in which every morphism is an isomorphism.

Example 1.1.12.

(i) A **group** is a groupoid with one object.\(^ {15}\)

(ii) For any space \(X\), its **fundamental groupoid** \(\Pi_1(X)\) is a category whose objects are the points of \(X\) and whose morphisms are endpoint-preserving homotopy classes of paths.

---

\(^{14}\)Mac Lane credits Emmy Noether for emphasizing the importance of homomorphisms in abstract algebra, particularly the homomorphism onto a quotient group, which plays an integral role in the statement of her first isomorphism theorem. His recollection is that the arrow notation first appeared around 1940, perhaps due to Hurewicz [ML88]. The notation \(\text{Hom}(X, Y)\) was first used in [EM42a] for the set of homomorphisms between a pair of abelian groups.

\(^{15}\)This is not simply an example; it is a definition.
A subcategory \( D \) of a category \( C \) is defined by restricting to a subcollection of objects and subcollection of morphisms subject to the requirements that the subcategory \( D \) contains the domain and codomain of any morphism in \( D \), the identity morphism of any object in \( D \), and the composite of any composable pair of morphisms in \( D \). For example, there is a subcategory \( \text{CRing} \subset \text{Ring} \) of commutative unital rings. Both of these form subcategories of the category \( \text{Rng} \) of not-necessarily unital rings and homomorphisms that need not preserve the multiplicative unit.16

**Lemma 1.1.13.** Any category \( C \) contains a maximal groupoid, the subcategory containing all of the objects and only those morphisms that are isomorphisms.

**Proof.** Exercise 1.1.ii. \( \square \)

For instance, \( \text{Fin}_{\text{iso}} \), the category of finite sets and bijections, is the maximal sub-groupoid of the category \( \text{Fin} \) of finite sets and all functions. Example 1.4.9 will explain how this groupoid can be regarded as a categorification of the natural numbers, providing a vantage point from which to prove the laws of elementary arithmetic.

**Exercises.**

**Exercise 1.1.i.**

(i) Show that a morphism can have at most one inverse isomorphism.

(ii) Consider a morphism \( f : x \to y \). Show that if there exists a pair of morphisms \( g, h : y \to x \) so that \( gf = 1_x \) and \( fh = 1_y \), then \( g = h \) and \( f \) is an isomorphism.

**Exercise 1.1.ii.** Let \( C \) be a category. Show that the collection of isomorphisms in \( C \) defines a subcategory, the maximal groupoid inside \( C \).

**Exercise 1.1.iii.** For any category \( C \) and any object \( c \in C \), show that:

(i) There is a category \( c/C \) whose objects are morphisms \( f : c \to x \) with domain \( c \) and in which a morphism from \( f : c \to x \) to \( g : c \to y \) is a map \( h : x \to y \) between the codomains so that the triangle

\[
\begin{array}{ccc}
  f & c & g \\
  \downarrow & \downarrow & \downarrow \\
  x & h & y \\
\end{array}
\]

commutes, i.e., so that \( g = hf \).

(ii) There is a category \( C/c \) whose objects are morphisms \( f : x \to c \) with codomain \( c \) and in which a morphism from \( f : x \to c \) to \( g : y \to c \) is a map \( h : x \to y \) between the domains so that the triangle

\[
\begin{array}{ccc}
  x & h & y \\
  f & \downarrow & \downarrow \\
  c & g & y \\
\end{array}
\]

commutes, i.e., so that \( f = gh \).

The categories \( c/C \) and \( C/c \) are called slice categories of \( C \) under and over \( c \), respectively.

---

16To justify our default notion of ring, see Poonen’s “Why all rings should have a 1” [Poo14]. The relationship between unital and non-unital rings is explored in greater depth in §4.6.
1.2. Duality

The dual of any axiom for a category is also an axiom. A simple metamathematical argument thus proves the dual principle. If any statement about a category is deducible from the axioms for a category, the dual statement is likely deducible.

Saunders Mac Lane, “Duality for groups” [ML50]

Upon first acquaintance, the primary role played by the notion of a category might appear to be taxonomic: vector spaces and linear maps define one category, manifolds and smooth functions define another. But a category, as defined in 1.1.1, is also a mathematical object in its own right, and as with any mathematical definition, this one is worthy of further consideration. Applying a mathematician’s gaze to the definition of a category, the following observation quickly materializes. If we visualize the morphisms in a category as arrows pointing from their domain object to their codomain object, we might imagine simultaneously reversing the directions of every arrow. This leads to the following notion.

**Definition 1.2.1.** Let $C$ be any category. The **opposite category** $C^{\text{op}}$ has

- the same objects as in $C$, and
- a morphism $f^{\text{op}}$ in $C^{\text{op}}$ for each a morphism $f$ in $C$ so that the domain of $f^{\text{op}}$ is defined to be the codomain of $f$ and the codomain of $f^{\text{op}}$ is defined to be the domain of $f$: i.e.,

$$f^{\text{op}} : X \to Y \in C^{\text{op}} \iff f : Y \to X \in C.$$ 

That is, $C^{\text{op}}$ has the same objects and morphisms as $C$, except that “each morphism is pointing in the opposite direction.” The remaining structure of the category $C^{\text{op}}$ is given as follows:

- For each object $X$, the arrow $1^{\text{op}}_X$ serves as its identity in $C^{\text{op}}$.
- To define composition, observe that a pair of morphisms $f^{\text{op}}, g^{\text{op}}$ in $C^{\text{op}}$ is composable precisely when the pair $g, f$ is composable in $C$, i.e., precisely when the codomain of $g$ equals the domain of $f$. We then define $g^{\text{op}} \cdot f^{\text{op}}$ to be $(f \cdot g)^{\text{op}}$: i.e.,

$$f^{\text{op}} : X \to Y, g^{\text{op}} : Y \to Z \in C^{\text{op}} \iff g^{\text{op}} f^{\text{op}} : X \to Z \in C^{\text{op}}$$

$$g : Z \to Y, f : Y \to X \in C \iff fg : Z \to X \in C$$

The data described in Definition 1.2.1 defines a category $C^{\text{op}}$—i.e., the composition law is associative and unital—if and only if $C$ defines a category. In summary, the process of “turning around the arrows” or “exchanging domains and codomains” exhibits a syntactical self-duality satisfied by the axioms for a category. Note that the category $C^{\text{op}}$ contains precisely the same information as the category $C$. Questions about the one can be answered by examining the other.

**Example 1.2.2.**

(i) $\text{Mat}_R^{\text{op}}$ is the category whose objects are non-zero natural numbers and in which a morphism from $m$ to $n$ is an $m \times n$ matrix with values in $R$. The upshot is that a reader who would have preferred the opposite handedness conventions when defining $\text{Mat}_R$ would have lost nothing by adopting them.

(ii) When a preorder $(\mathcal{P}, \leq)$ is regarded as a category, its opposite category is the category that has a morphism $x \to y$ if and only if $y \leq x$. For example, $\omega^{\text{op}}$ is the category...
freely generated by the graph
\[ \cdots \to 3 \to 2 \to 1 \to 0. \]

(iii) If \( G \) is a group, regarded as a one-object groupoid, the category \((BG)^{op} \cong B(G^{op})\) is again a one-object groupoid, and hence a group. The group \(G^{op}\) is called the **opposite group** and is used to define right actions as a special case of left actions; see Example 1.3.9.

This syntactical duality has a very important consequence for the development of category theory. Any theorem containing a universal quantification of the form “for all categories \( C \)” also necessarily applies to the opposites of these categories. Interpreting the result in the dual context leads to a **dual theorem**, proven by the dual of the original proof, in which the direction of each arrow appearing in the argument is reversed. The result is a two-for-one deal: any proof in category theory simultaneously proves two theorems, the original statement and its dual.\(^{17}\) For example, the reader may have found Exercise 1.1.iii redundant, precisely because the statements (i) and (ii) are dual; see Exercise 1.2.i.

To illustrate the principle of duality in category theory, let us consider the following result, which provides an important characterization of the isomorphisms in a category.

**Lemma 1.2.3.** The following are equivalent:

(i) \( f: x \to y \) is an isomorphism in \( C \).

(ii) For all objects \( c \in C \), post-composition with \( f \) defines a bijection

\[ f_*: C(c, x) \to C(c, y). \]

(iii) For all objects \( c \in C \), pre-composition with \( f \) defines a bijection

\[ f^*: C(y, c) \to C(x, c). \]

**Remark 1.2.4.** In language introduced in Chapter 2, Lemma 1.2.3 asserts that isomorphisms in a locally small category are defined **representably** in terms of isomorphisms in the category of sets. That is, a morphism \( f: x \to y \) in an arbitrary locally small category \( C \) is an isomorphism if and only if the post-composition function \( f_*: C(c, x) \to C(c, y) \) between hom-sets defines an isomorphism in \( \text{Set} \) for each object \( c \in C \).

In set theoretical foundations that permit the definition of functions between large sets, the proof given here applies also to non-locally small categories. In our exposition, the set theoretical hypotheses of smallness and local smallness will only appear when there are essential subtleties concerning the sizes of the categories in question. This is not one of those occasions.

**Proof of Lemma 1.2.3.** We will prove the equivalence (i) \( \iff \) (ii) and conclude the equivalence (i) \( \iff \) (iii) by duality.

Assuming (i), namely that \( f: x \to y \) is an isomorphism with inverse \( g: y \to x \), then, as an immediate application of the associativity and identity laws for composition in a category, post-composition with \( g \) defines an inverse function

\[ g_*: C(c, y) \to C(c, x) \]

to \( f_* \) in the sense that the composites

\[ g_* f_*: C(c, x) \to C(c, x) \quad \text{and} \quad f_* g_*: C(c, y) \to C(c, y) \]

\[^{17}\text{More generally, the proof of a statement of the form “for all categories } C_1, C_2, \ldots, C_n \text{” leads to } 2^n \text{ dual theorems. In practice, however, not all of the dual statements will differ meaningfully from the original; see e.g., §4.3.}\]
are both the identity function: for any \( h: c \to x \) and \( k: c \to y \), \( g_*f_*(h) = gfh = h \) and \( f_*g_*(k) = fgk = k \).

Conversely, assuming (ii), there must be an element \( g \in C(y, x) \) whose image under \( f_*: C(y, x) \to C(y, y) \) is \( 1_y \). By construction, \( 1_y = fg \). But now, by associativity of composition, the elements \( gf, 1_x \in C(x, x) \) have the common image \( f \) under the function \( f_*: C(x, x) \to C(x, y) \), whence \( gf = 1_x \). Thus, \( f \) and \( g \) are inverse isomorphisms.

We have just proven the equivalence (i) \( \iff \) (ii) for all categories and in particular for the category \( C^{op} \): i.e., a morphism \( f^{op}: y \to x \) in \( C^{op} \) is an isomorphism if and only if

\[ f^{op}_*: C^{op}(c, y) \to C^{op}(c, x) \]

is an isomorphism for all \( c \in C \).

Interpreting the data of \( C^{op} \) in its opposite category \( C \), the statement (1.2.5) expresses the same mathematical content as

\[ f^*: C(y, c) \to C(x, c) \]

is an isomorphism for all \( c \in C \).

That is: \( C^{op}(c, x) = C(x, c) \), post-composition with \( f^{op} \) in \( C^{op} \) translates to pre-composition with \( f \) in the opposite category. The notion of isomorphism, as defined in 1.1.9, is self-dual: \( f^{op}: y \to x \) is an isomorphism in \( C^{op} \) if and only if \( f: x \to y \) is an isomorphism in \( C \). So the equivalence (i) \( \iff \) (ii) in \( C^{op} \) expresses the equivalence (i) \( \iff \) (iii) in \( C \).\( \square \)

Concise expositions of the duality principle in category theory may be found in [Awo10, §3.1] and [HIS97, §II.3]. As we become more comfortable with arguing by duality, dual proofs and eventually also dual statements will seldom be described in this much detail.

Categorical definitions also have duals; for instance:

**Definition 1.2.7.** A morphism \( f: x \to y \) in a category is

(i) a **monomorphism** if for any parallel morphisms \( h, k: w \cong x \), \( fh = fk \) implies that \( h = k \); or

(ii) an **epimorphism** if for any parallel morphisms \( h, k: y \cong z \), \( hf = kf \) implies that \( h = k \).

Note that a monomorphism or epimorphism in \( C \) is, respectively, an epimorphism or monomorphism in \( C^{op} \). In adjectival form, a monomorphism is **monic** and an epimorphism is **epic**. In common shorthand, a monomorphism is a **mono** and an epimorphism is an **epi**. For graphical emphasis, monos are often decorated with a tail “\( \rightarrow \)” while epis may be decorated at their head “\( \twoheadrightarrow \)”.

The following dual statements re-express Definition 1.2.7:

(i) \( f: x \to y \) is a monomorphism in \( C \) if and only if for all objects \( c \in C \), post-composition with \( f \) defines an injection \( f_*: C(c, x) \to C(c, y) \).

(ii) \( f: x \to y \) is an epimorphism in \( C \) if and only if for all objects \( c \in C \), pre-composition with \( f \) defines an injection \( f^*: C(y, c) \to C(x, c) \).

**Example 1.2.8.** Suppose \( f: X \to Y \) is a monomorphism in the category of sets. Then, in particular, given any two maps \( x, x': 1 \cong X \), whose domain is the singleton set, if \( fx = fx' \) then \( x = x' \). Thus, monomorphisms are injective functions. Conversely, any injective function can easily be seen to be a monomorphism.

Similarly, a function \( f: X \to Y \) is an epimorphism in the category of sets if and only if it is surjective. Given functions \( h, k: Y \cong Z \), the equation \( hf = kf \) says exactly that \( h \) is equal to \( k \) on the image of \( f \). This only implies that \( h = k \) in the case where the image is all of \( Y \).

\[ \text{A similar translation, as just demonstrated between the statements (1.2.5) and (1.2.6), transforms the proof of (i) \( \iff \) (ii) into a proof of (i) \( \iff \) (iii).} \]
Thus, monomorphisms and epimorphisms should be regarded as categorical analogs of the notions of injective and surjective functions. In practice, if \( C \) is a category in which objects have “underlying sets,” then any morphism that induces an injective or surjective function between these defines a monomorphism or epimorphism; see Exercise 1.6.iii for a precise discussion. However, even in such categories, the notions of monomorphism and epimorphism can be more general, as demonstrated in Exercise 1.6.v.

**Example 1.2.9.** Suppose that \( x \xrightarrow{s} y \xrightarrow{r} x \) are morphisms so that \( rs = 1_x \). The map \( s \) is a **section** or **right inverse** to \( r \), while the map \( r \) defines a **retraction** or **left inverse** to \( s \). The maps \( s \) and \( r \) express the object \( x \) as a **retract** of the object \( y \).

In this case, \( s \) is always a monomorphism and, dually, \( r \) is always an epimorphism. To acknowledge the presence of these one-sided inverses, \( s \) is said to be a **split monomorphism** and \( r \) is said to be a **split epimorphism**.\(^{19}\)

**Example 1.2.10.** By the previous example, an isomorphism is necessarily both monic and epic, but the converse need not hold in general. For example, the inclusion \( \mathbb{Z} \hookrightarrow \mathbb{Q} \) is both monic and epic in the category \( \text{Ring} \), but this map is not an isomorphism: there are no ring homomorphisms from \( \mathbb{Q} \) to \( \mathbb{Z} \).

Since the notions of monomorphism and epimorphism are dual, their abstract categorical properties are also dual, such as exhibited by the following lemma.

**Lemma 1.2.11.**

(i) If \( f : x \to y \) and \( g : y \to z \) are monomorphisms, then so is \( gf : x \to z \).

(ii) If \( f : x \to y \) and \( g : y \to z \) are morphisms so that \( gf \) is monic, then \( f \) is monic.

Dually:

(i') If \( f : x \to y \) and \( g : y \to z \) are epimorphisms, then so is \( gf : x \to z \).

(ii') If \( f : x \to y \) and \( g : y \to z \) are morphisms so that \( gf \) is epic, then \( g \) is epic.

**Proof.** Exercise 1.2.iii. \( \square \)

**Exercises.**

**Exercise 1.2.i.** Show that \( C/c \cong (c/\text{C}^{\text{op}})^{\text{op}} \). Defining \( C/c \) to be \((c/\text{C}^{\text{op}})^{\text{op}}\), deduce Exercise 1.1.iii(ii) from Exercise 1.1.iii(i).

**Exercise 1.2.ii.**

(i) Show that a morphism \( f : x \to y \) is a split epimorphism in a category \( C \) if and only if for all \( c \in C \), post-composition \( f_c : C(c, x) \to C(c, y) \) defines a surjective function.

(ii) Argue by duality that \( f \) is a split monomorphism if and only if for all \( c \in C \), pre-composition \( f^*_c : C(y, c) \to C(x, c) \) is a surjective function.

**Exercise 1.2.iii.** Prove Lemma 1.2.11 by proving either (i) or (i') and either (ii) or (ii'), then arguing by duality. Conclude that the monomorphisms in any category define a subcategory of that category and dually that the epimorphisms also define a subcategory.

**Exercise 1.2.iv.** What are the monomorphisms in the category of fields?

**Exercise 1.2.v.** Show that the inclusion \( \mathbb{Z} \hookrightarrow \mathbb{Q} \) is both a monomorphism and an epimorphism in the category \( \text{Ring} \) of rings. Conclude that a map that is both monic and epic need not be an isomorphism.

**Exercise 1.2.vi.** Prove that a morphism that is both a monomorphism and a split epimorphism is necessarily an isomorphism. Argue by duality that a split monomorphism that is an epimorphism is also an isomorphism.

\(^{19}\)The axiom of choice asserts that every epimorphism in the category of sets is a split epimorphism.
1.3. Functoriality

... every sufficiently good analogy is yearning to become a functor.

John Baez. “Quantum Quandaries: A Category-Theoretic Perspective” [Bae06]

A key tenet in category theory, motivating the very definition of a category, is that any mathematical object should be considered together with its accompanying notion of structure-preserving morphism. In “General theory of natural equivalences” [EM45], Eilenberg and Mac Lane argue further:

... whenever new abstract objects are constructed in a specified way out of given ones, it is advisable to regard the construction of the corresponding induced mappings on these new objects as an integral part of their definition.

Categories are themselves mathematical objects, if of a somewhat unfamiliar sort, which leads to a question: What is a morphism between categories?

Definition 1.3.1. A functor $F : \mathcal{C} \to \mathcal{D}$, between categories $\mathcal{C}$ and $\mathcal{D}$, consists of the following data:

- An object $Fc \in \mathcal{D}$, for each object $c \in \mathcal{C}$.
- A morphism $Ff : Fc \toFc' \in \mathcal{D}$, for each morphism $f : c \to c' \in \mathcal{C}$, so that the domain and codomain of $Ff$ are, respectively, equal to $F$ applied to the domain or codomain of $f$.

The assignments are required to satisfy the following two functoriality axioms:

- For any composable pair $f, g$ in $\mathcal{C}$, $Fg \cdot Ff = F(g \cdot f)$.
- For each object $c$ in $\mathcal{C}$, $F(1_c) = 1_{Fc}$.

Put concisely, a functor consists of a mapping on objects and a mapping on morphisms that preserves all of the structure of a category, namely domains and codomains, composition, and identities.\(^{20}\)

Example 1.3.2.

(i) There is an endofunctor\(^{21}\) $P : \text{Set} \to \text{Set}$ that sends a set $A$ to its power set $PA = \{A' \subset A\}$ and a function $f : A \to B$ to the direct-image function $f_* : PA \to PB$ that sends $A' \subset A$ to $f(A') \subset B$.

(ii) Each of the categories listed in Example 1.1.3 has a forgetful functor, a general term that is used for any functor that forgets structure, whose codomain is the category of sets. For example, $U : \text{Group} \to \text{Set}$ sends a group to its underlying set and a group homomorphism to its underlying function. The functor $U : \text{Top} \to \text{Set}$ sends a space to its set of points. There are two natural forgetful functors $V, E : \text{Graph} \to \text{Set}$ that send a graph to its vertex or edge sets, respectively; if desired, these can be combined

\(^{20}\)While a functor should be regarded as a mapping from the data of one category to the data of another, parentheses are used as seldom as possible unless demanded for notational clarity.

\(^{21}\)An endofunctor is a functor whose domain is equal to its codomain.
to define a single functor \( V \sqcup E : \text{Graph} \to \text{Set} \) that carries a graph to the disjoint union of its vertex and edge sets. These mappings are functorial because in each instance a morphism in the domain category has an underlying function.

(iii) There are intermediate forgetful functors \( \text{Mod}_R \to \text{Ab} \) and \( \text{Ring} \to \text{Ab} \) that forget some but not all of the algebraic structure. The inclusion functors \( \text{Ab} \leftarrow \text{Group} \) and \( \text{Field} \leftarrow \text{Ring} \) may also be regarded as “forgetful.” Note that the latter two, but neither of the former, are injective on objects: a group is either abelian or not, but an abelian group might admit the structure of a ring in multiple ways.

(iv) Similarly, there are forgetful functors \( \text{Group} \to \text{Set} , \) and \( \text{Ring} \to \text{Set} , \) that take the basepoint to be the identity and zero elements, respectively. These assignments are functorial because group and ring homomorphisms necessarily preserve these elements.

(v) There are functors \( \text{Top} \to \text{Htpy} \) and \( \text{Top}_* \to \text{Htpy}_* \) that act as the identity on objects and send a (based) continuous function to its homotopy class.

(vi) The fundamental group defines a functor \( \pi_1 : \text{Top}_* \to \text{Group} ; \) a continuous function \( f : (X, x) \to (Y, y) \) of based spaces induces a group homomorphism \( f_* : \pi_1(X, x) \to \pi_1(Y, y) \) and this assignment is functorial, satisfying the two functoriality axioms described above. A precise expression of the statement that “the fundamental group is a homotopy invariant” is that this functor factors through the functor \( \text{Top}_* \to \text{Htpy}_* \) to define a functor \( \pi_1 : \text{Htpy}_* \to \text{Group} . \)

(vii) A related functor \( \Pi_1 : \text{Top} \to \text{Groupoid} \) assigns an unbased topological space its fundamental groupoid, the category defined in Example 1.1.12(ii). A continuous function \( f : X \to Y \) induces a functor \( f_* : \Pi_1(X) \to \Pi_1(Y) \) that carries a point \( x \in X \) to the point \( f(x) \in Y \). This mapping extends to morphisms in \( \Pi_1(X) \) because continuous functions preserve untangled and path homotopy classes.

(viii) For each \( n \in \mathbb{Z} \), there are functors \( Z_n, B_n, H_n : \text{Ch}_R \to \text{Mod}_R . \) The functor \( Z_n \) computes the \( n \)-cycles defined by \( Z_n \text{C}_* = \ker(d : C_n \to C_{n-1}) \). The functor \( B_n \) computes the \( n \)-boundary defined by \( B_n \text{C}_* = \text{im}(d : C_{n+1} \to C_n) \). The functor \( H_n \) computes the \( n \)th homology \( H_n \text{C}_* := Z_n \text{C}_*/B_n \text{C}_* \). We leave it to the reader to verify that each of these three constructions is functorial. Considering all degrees simultaneously, the cycle, boundary, and homology functors assemble into functors \( Z_*, B_*, H_* : \text{Ch}_R \to \text{GrMod}_R \) from the category of chain complexes to the category of graded \( R \)-modules. The singular homology of a topological space is defined by precomposing \( H_* \) with a suitable functor \( \text{Top} \to \text{Ch}_R \).

(ix) There is a functor \( F : \text{Set} \to \text{Group} \) that sends a set \( X \) to the free group on \( X \). This is the group whose elements are finite “words” whose letters are elements \( x \in X \) or their formal inverses \( x^{-1} \), modulo an equivalence relation that equates the words \( xx^{-1} \) and \( x^{-1}x \) with the empty word. Multiplication is by concatenation, with the empty word serving as the identity. This is one instance of a large family of “free” functors studied in Chapter 4.

(x) The chain rule expresses the functoriality of the derivative. Let \( \text{Euclid} \), denote the category whose objects are pointed finite-dimensional Euclidean spaces \((\mathbb{R}^n, a)\)—or, better, open subsets thereof—and whose morphisms are pointed differentiable functions. The total derivative of \( f : \mathbb{R}^n \to \mathbb{R}^m \), evaluated at the designated basepoint \( a \in \mathbb{R}^n \), gives rise to a matrix called the Jacobian matrix defining the directional derivatives of \( f \) at the point \( a \). If \( f \) is given by component functions \( f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \), the \((i, j)\)-entry of this matrix is \( \frac{\partial}{\partial x^j} f_i(a) \). This defines the action on morphisms of a functor \( D : \text{Euclid}_* \to \text{Mat}_{\mathbb{R}} \); on objects, \( D \) assigns a pointed Euclidean space...
its dimension. Given \( g: \mathbb{R}^m \to \mathbb{R}^k \) carrying the designated basepoint \( f(a) \in \mathbb{R}^m \) to \( gf(a) \in \mathbb{R}^k \), functoriality of \( D \) asserts that the product of the Jacobian of \( f \) at \( a \) with the Jacobian of \( g \) at \( f(a) \) equals the Jacobian of \( gf \) at \( a \). This is the chain rule from multivariable calculus.\(^{22}\)

(xi) Any commutative monoid \( M \) can be used to define a functor \( M^{-}: \text{Fin}_e \to \text{Set} \). Writing \( n_e \in \text{Fin}_e \) for the set with \( n \) non-basepoint elements, define \( M^{n_e} \) to be \( M^n \), the \( n \)-fold cartesian product of the set \( M \) with itself. By convention, \( M^0_e \) is a singleton set. For any based map \( f: m_+ \to n_+ \), define the \( i \)th component of the corresponding function \( M^f: M^m \to M^n \) by projecting from \( M^m \) to the coordinates indexed by elements in the fiber \( f^{-1}(i) \) and then multiplying these using the commutative monoid structure; if the fiber is empty, the function \( M^f \) inserts the unit element in the \( i \)th coordinate. Note each of the sets \( M^n \) itself has a basepoint, the \( n \)-tuple of unit elements, and each of the maps in the image of the functor are based. It follows that the functor \( M^- \) lifts along the forgetful functor \( U: \text{Set} \to \text{Set} \).

There is a special property satisfied by this construction that allows one to extract the commutative monoid \( M \) from the functor \( \text{Fin}_e \to \text{Set} \). This observation was used by Segal to introduce a suitable notion of “commutative monoid” into algebraic topology [Seg74].

More examples of functors will appear shortly, but first we illustrate the utility of knowing that the assignment of a mathematical object of one type to mathematical objects of another type is \emph{functorial}. Applying the functoriality of the fundamental group construction \( \pi_1: \text{Top}_e \to \text{Group} \), one can prove:

\textbf{Theorem 1.3.3 (Brouwer Fixed Point Theorem).} \textit{Any continuous endomorphism of a 2-dimensional disk} \( D^2 \) \textit{has a fixed point.}

\textbf{Proof.} Assuming \( f: D^2 \to D^2 \) is such that \( f(x) \neq x \) for all \( x \in D^2 \), there is a continuous function \( r: D^2 \to S^1 \) that carries a point \( x \in D^2 \) to the intersection of the ray from \( f(x) \) to \( x \) with the boundary circle \( S^1 \). Note that the function \( r \) fixes the points on the boundary circle \( S^1 \subset D^2 \). Thus, \( r \) defines a retraction of the inclusion \( i: S^1 \hookrightarrow D^2 \), which is to say, the composite \( i \circ r \) is the identity.

![Diagram](https://via.placeholder.com/150)

Pick any basepoint on the boundary circle \( S^1 \) and apply the functor \( \pi_1 \) to obtain a composable pair of group homomorphisms:

\[
\pi_1(S^1) \xrightarrow{\pi_1(i)} \pi_1(D^2) \xrightarrow{\pi_1(r)} \pi_1(S^1).
\]

By the functoriality axioms, we must have

\[
\pi_1(r) \cdot \pi_1(i) = \pi_1(r i) = \pi_1(1_{S^1}) = 1_{\pi_1(S^1)}.
\]

However, a computation involving covering spaces reveals that \( \pi_1(S^1) = \mathbb{Z} \), while \( \pi_1(D^2) = 0 \), the trivial group. The composite endomorphism \( \pi_1(r) \cdot \pi_1(i) \) of \( \mathbb{Z} \) must be zero, since it factors through the trivial group. Thus, it cannot equal the identity homomorphism, which

\(^{22}\)Taking a more sophisticated perspective, the derivative defines the action on morphisms of a functor from the category \( \text{Man}_e \), to the category of real vector spaces that sends a pointed manifold to its tangent space.
carries the generator $1 \in \mathbb{Z}$ to itself ($0 \neq 1$). This contradiction proves that the retraction $r$ cannot exist, and so $f$ must have a fixed point. \hfill $\square$

Functoriality also plays a key role in the emerging area of topological data analysis.

**Example 1.3.4** (in search of a clustering functor). A *clustering algorithm* is a function that converts a finite metric space into a partition of its points into sets of “clusters.” An impossibility theorem of Kleinberg proves that there are no clustering algorithms that satisfy three reasonable axioms [Kle03]. A key insight of Carlsson and Mémoli is that these axioms can be encoded as morphisms in a category of finite metric spaces in such a way that what is desired is not a clustering function but a clustering functor into a suitable category [CM13]. Ghrist’s *Elementary Applied Topology* [Ghr14, p. 216] describes this move as follows:

> What is the good of this? Category theory is criticized as an esoteric language: formal and fruitless for conversation. *This is not so.*

The virtue of reformulating (the negative) Theorem [of Kleinberg] functorially is a clearer path to a positive statement. If the goal is to have a theory of clustering; if clustering is, properly, a nontrivial functor; if no nontrivial functors between the proposed categories exist; then, naturally, the solution is to alter the domain or codomain categories and classify the ensuing functors. One such modification is to consider a category of persistent clusters.

One pair of categories considered in [CM13] are the categories FinMetric, of finite metric spaces and distance non-increasing functions, and Cluster, of clusters and refinements. An object in Cluster is a partitioned set. Given a function $f : X \to Y$, the preimages of a partition of the set $Y$ define a partition of $X$. A morphism in Cluster is a function $f : X \to Y$ of underlying sets so that the given partition on $X$ refines the partition on $X$ defined by the preimages of the given partition on $Y$.

Carlsson and Mémoli observe that the only scale-invariant functors FinMetric $\to$ Cluster either assign each metric space the discrete partition (into singletons) or the indiscrete partition (into a single cluster); both cases fail to satisfy Kleinberg’s surjectivity condition. This suggests that clusters should be replaced by a notion of “persistent” clusters. A *persistent cluster* on $X$ is a functor from the poset $((0, \infty), \leq)$ to the poset of clusters on $X$, where $\phi \leq \psi$ if and only if the partition $\phi$ refines the partition $\psi$. The idea is that when the parameter $r \in [0, \infty)$ is small, the partition on $X$ might be very fine, but the clusters are allowed to coalesce as one “zooms out,” i.e., as $r$ increases.

There is a category PCluster whose objects are persistent clusters and whose morphisms are functions of underlying sets $f : X \to Y$ that define morphisms in Cluster for each $r \in [0, \infty)$. Carlsson and Mémoli prove that there is a unique functor FinMet $\to$ PCluster, which takes the metric space with two points of distance $r$ to the persistent cluster with one cluster for $t \geq r$ and two clusters for $0 \leq t < r$ and satisfies two other reasonable conditions; see [CM13] for the details.

---

23The same argument, with the $n$th homotopy group functor $\pi_n : \text{Top} \to \text{Group}$ in place of $\pi_1$, proves that any continuous endomorphism of an $n$-dimensional disk has a fixed point.

24Namely, there are no clustering algorithms that are invariant under rigid scaling, consistent under alterations to the distance function that “sharpen” the point clusters, and have the property that some distance function realizes each possible partition.

25Emphasis his.
The functors defined in 1.3.1 are called **covariant** so as to distinguish them from another variety of functor that we now introduce.

**Definition 1.3.5.** A **contravariant functor** $F$ from $C$ to $D$ is a functor $F : C^{\text{op}} \to D$.\(^{26}\)

Explicitly, this consists of the following data:

- An object $Fc \in D$, for each object $c \in C$.
- A morphism $Ff : Fc' \to Fc \in D$, for each morphism $f : c \to c' \in C$, so that the domain and codomain of $Ff$ are, respectively, equal to $F$ applied to the codomain or domain of $f$.

The assignments are required to satisfy the following two **functoriality axioms**:

- For any composable pair $f, g$ in $C$, $Ff \cdot Fg = F(g \cdot f)$.
- For each object $c \in C$, $F(1_c) = 1_{Fc}$.

**Notation 1.3.6.** To avoid unnatural arrow-theoretic representations, a morphism in the domain of a functor $F : C^{\text{op}} \to D$ will always be depicted as an arrow $f : c \to c'$ in $C$, pointing from its domain in $C$ to its codomain in $C$. Similarly, its image will always be depicted as an arrow $Ff : Fc' \to Fc$ in $D$, pointing from its domain to its codomain. Note that these conventions require that the domain and codomain objects switch their relative places, from left to right, but in examples, for instance in the case where $C$ and $D$ are concrete categories, these positions are the familiar ones. Graphically, the mapping on morphisms given by a contravariant functor is depicted as follows:

$$
\begin{array}{cccc}
C^{\text{op}} & \xrightarrow{F} & D \\
\uparrow c & \mapsto & \uparrow Fc \\
\downarrow f & \mapsto & \uparrow Ff \\
\uparrow c' & \mapsto & \uparrow Fc'
\end{array}
$$

In accordance with this convention, if $f : c \to c'$ and $g : c' \to c''$ are morphisms in $C$, their composite will always be written as $gf : c \to c''$. The image of this morphism under the contravariant functor $F : C^{\text{op}} \to D$ is $F(gf) : Fc'' \to Fc$, the composite $Ff \cdot Fg$ of $Fg : Fc'' \to Fc'$ and $Ff : Fc' \to Fc$.

In summary, even in the presence of opposite categories, we always make an effort to draw arrows pointing in the “correct way” and depict composition in the usual order.\(^{27}\)

**Example 1.3.7.**

(i) The contravariant power set functor $P : \text{Set}^{\text{op}} \to \text{Set}$ sends a set $A$ to its power set $PA$ and a function $f : A \to B$ to the inverse-image function $f^{-1} : PB \to PA$ that sends $B' \subseteq B$ to $f^{-1}(B') \subseteq A$.

(ii) There is a functor $(-)^{\ast} : \text{Vect}^{\text{op}}_k \to \text{Vect}_k$ that carries a vector space to its **dual vector space** $V^{\ast} = \text{Hom}(V, k)$. A vector in $V^{\ast}$ is a **linear functional** on $V$, i.e., a linear map

\(^{26}\)In this text, a contravariant functor $F$ from $C$ to $D$ will always be written as $F : C^{\text{op}} \to D$. Some mathematicians omit the “op” and let the context or surrounding verbiage convey the variance. We think this is bad practice, as the co- or contravariance is an essential part of the data of a functor, which is not necessarily determined by its assignation on objects. More to the point, we find that this notational convention helps mitigate the consequences of temporary distraction. Seeing $F : C^{\text{op}} \to D$ written on a chalkboard immediately conveys that $F$ is a contravariant functor from $C$ to $D$, even to the most spaced-out observer. A similar principle will motivate other notational conventions introduced in Definition 3.1.15 and Notation 4.1.5.

\(^{27}\)Of course, technically there is no meaning to the phrase “opposite category”: every category is the opposite of some other category (its opposite category). But in practice, there is no question which of $\text{Set}$ and $\text{Set}^{\text{op}}$ is the “opposite category,” and sufficiently many of the other cases can be deduced from this one.
This functor is contravariant, with a linear map \( \phi : V \rightarrow W \) sent to the linear map \( \phi^* : V^* \rightarrow W^* \) that pre-composes a linear functional \( W \rightarrow \mathbb{k} \) with \( \phi \) to obtain a linear functional \( V \rightarrow W \rightarrow \mathbb{k} \).

(iii) The functor \( O : \text{Top}^{\text{op}} \rightarrow \text{Poset} \) that carries a space \( X \) to its poset \( O(X) \) of open subsets is contravariant on the category of spaces: a continuous map \( f : X \rightarrow Y \) gives rise to a function \( f^{-1} : O(Y) \rightarrow O(X) \) that carries an open subset \( U \subset Y \) to its preimage \( f^{-1}(U) \), which is open in \( X \); this is the definition of continuity. A similar functor \( C : \text{Top}^{\text{op}} \rightarrow \text{Poset} \) carries a space to its poset of closed subsets.

(iv) There is a contravariant functor \( \text{Spec} : \text{CRing}^{\text{op}} \rightarrow \text{Top} \) that sends a commutative ring \( R \) to its set \( \text{Spec}(R) \) of prime ideals given the Zariski topology. The closed subsets in the Zariski topology are those subsets \( V_I \subset \text{Spec}(R) \) of prime ideals containing a fixed ideal \( I \subset R \). This construction is contravariantly functorial: for any ring homomorphism \( \phi : R \rightarrow S \) and prime ideal \( p \subset S \), the inverse image \( \phi^{-1}(p) \subset R \) defines a prime ideal of \( R \), and the inverse image function \( \phi^{-1} : \text{Spec}(S) \rightarrow \text{Spec}(R) \) is continuous with respect to the Zariski topology.

(v) For a generic small category \( C \), a functor \( C^{\text{op}} \rightarrow \text{Set} \) is called a (set-valued) presheaf on \( C \). A typical example is the functor \( O(X)^{\text{op}} \rightarrow \text{Set} \) whose domain is the poset \( O(X) \) of open subsets of a topological space \( X \) and whose value at \( U \subset X \) is the set of continuous real-valued functions on \( U \). The action on morphisms is by restriction. This presheaf is a sheaf, if it satisfies an axiom that is introduced in Definition 3.3.4.

(vi) Presheaves on the category \( \Delta \), of finite non-empty ordinals and order-preserving maps, are called simplicial sets. \( \Delta \) is also called the simplex category. The ordinal \( n + 1 = [0, 1, \ldots, n] \) may be thought of as a direct version of the topological \( n \)-simplex and, with this interpretation in mind, is typically denoted by \( \{ [n] \} \) by algebraic topologists.

The following result, which appears immediately after functors are first defined in [EM42b], is arguably the first lemma in category theory.

**Lemma 1.3.8.** Functors preserve isomorphisms.

**Proof.** Consider a functor \( F : C \rightarrow D \) and an isomorphism \( \phi : x \rightarrow y \) in \( C \) with inverse \( g : y \rightarrow x \). Applying the two functoriality axioms:

\[
F(g)F(f) = F(gf) = F(1_x) = 1_{F(x)}.
\]

Thus, \( Fg : Fy \rightarrow Fx \) is a left inverse to \( Ff : Fx \rightarrow Fy \). Exchanging the roles of \( f \) and \( g \) (or arguing by duality) shows that \( Fg \) is also a right inverse. \( \square \)

**Example 1.3.9.** Let \( G \) be a group, regarded as a one-object category \( BG \). A functor \( X : BG \rightarrow C \) specifies an object \( X \in C \) (the unique object in its image) together with an endomorphism \( g_* : X \rightarrow X \) for each \( g \in G \). This assignment must satisfy two conditions:

(i) \( h_*g_* = (hg)_* \), for all \( g, h \in G \).

(ii) \( e_* = 1_X \), where \( e \in G \) is the identity element.

In summary, the functor \( BG \rightarrow C \) defines an action of the group \( G \) on the object \( X \in C \). When \( C = \text{Set} \), the object \( X \) endowed with such an action is called a \( G \)-set. When \( C = \text{Vect}_k \), the object \( X \) is called a \( G \)-representation. When \( C = \text{Top} \), the object \( X \) is called a \( G \)-space. Note the utility of this categorical language for defining several analogous concepts simultaneously.

The action specified by a functor \( BG \rightarrow C \) is sometimes called a left action. A right action is a functor \( BG^{\text{op}} \rightarrow C \). As before, each \( g \in G \) determines an endomorphism.
g^* : X → X in C and the identity element must act trivially. But now, for a pair of elements g, h ∈ G these actions must satisfy the composition rule (hg)^* = g^*h^*.

Because the elements g ∈ G are isomorphisms when regarded as morphisms in the 1-object category BG that represents the group, their images under any such functor must also be isomorphisms in the target category. In particular, in the case of a G-representation V: BG → Vect_k, the linear map g_* : V → V must be an automorphism of the vector space V. The point is that the functoriality axioms (i) and (ii) imply automatically that each g_* is an automorphism and that (g^{-1})_* = (g_*)^{-1}; the proof is a special case of Lemma 1.3.8.

In summary:

**Corollary 1.3.10.** When a group G acts functorially on an object X in a category C, its elements g must act by automorphisms g_* : X → X and, moreover, (g_*)^{-1} = (g_*)^{-1}.

A functor may or may not preserve monomorphisms or epimorphisms, but an argument similar to the proof of Lemma 1.3.8 shows that a functor necessarily preserves split monomorphisms and split epimorphisms. The retraction or section defines an “equational witness” for the mono or the epi.

**Definition 1.3.11.** If C is locally small, then for any object c ∈ C we may define a pair of covariant and contravariant functors represented by c:

\[
\begin{align*}
C & \rightarrow \text{Set} \\
C^{\text{op}} & \rightarrow \text{Set}
\end{align*}
\]

The notation suggests the action on objects: the functor C(c, -) carries x ∈ C to the set C(c, x) of arrows from c to x in C. Dually, the functor C(-, c) carries x ∈ C to the set C(x, c).

The functor C(c, -) carries a morphism f : x → y to the post-composition function f_* : C(c, x) → C(c, y) introduced in Lemma 1.2.3(ii). Dually, the functor C(-, c) carries f to the pre-composition function f^* : C(y, c) → C(x, c) introduced in 1.2.3(iii). Note that post-composition defines a covariant action on hom-sets, while pre-composition defines a contravariant action. There are no choices involved here; post-composition is always a covariant operation, while pre-composition is always a contravariant one. This is just the natural order of things.

We leave it to the reader to verify that the assignments just described satisfy the two functoriality axioms. Note that Lemma 1.3.8 specializes in the case of represented functors to give a proof of the implications (i) ⇒ (ii) and (i) ⇒ (iii) of Lemma 1.2.3. These functors will play a starring role in Chapter 2, where a number of examples in disguise are discussed.

The data of the covariant and contravariant functors introduced in Definition 1.3.11 may be encoded in a single bifunctor, which is the name for a functor of two variables. Its domain is given by the product of a pair of categories.

**Definition 1.3.12.** For any categories C and D, there is a category C × D, their product, whose

- objects are ordered pairs (c, d), where c is an object of C and d is an object of D,
- morphisms are ordered pairs (f, g): (c, d) → (c', d'), where f : c → c' ∈ C and g : d → d' ∈ D, and
in which composition and identities are defined componentwise.

Definition 1.3.13. If \( C \) is locally small, then there is a **two-sided represented functor**

\[
C(-, -) : C^{\text{op}} \times C \to \text{Set}
\]
defined in the evident manner. A pair of objects \((x, y)\) is mapped to the hom-set \(C(x, y)\). A pair of morphisms \(f : w \to x\) and \(h : y \to z\) is sent to the function

\[
C(x, y) \xrightarrow{(f', h')} C(w, z)
\]

\[
g \mapsto hgf
\]

that takes an arrow \(g : x \to y\) and then pre-composes with \(f\) and post-composes with \(h\) to obtain \(hgf : w \to z\).

At the beginning of this section, it was suggested that functors define morphisms between categories. Indeed, categories and functors assemble into a category. Here the size issues are even more significant than we have encountered thus far. To put a lid on things, define \(\text{Cat}\) to be the category whose objects are small categories and whose morphisms are functors between them. This category is locally small but not small: it contains \(\text{Set}\), \(\text{Poset}\), \(\text{Monoid}\), \(\text{Group}\), and \(\text{Groupoid}\) as proper subcategories (see Exercises 1.3.i and 1.3.ii). However, none of these categories are objects of \(\text{Cat}\).

The non-small categories of Example 1.1.3 are objects of \(\text{CAT}\), some category of “large” categories and functors between them. Russell’s paradox suggests that \(\text{CAT}\) should not be so large as to contain itself, so we require the objects in \(\text{CAT}\) to be locally small categories; the category \(\text{CAT}\) defined in this way is not locally small, and so is thus excluded. There is an inclusion functor \(\text{Cat} \hookrightarrow \text{CAT}\) but no obvious functor pointing in the other direction.

The category of categories gives rise to a notion of an **isomorphism of categories**, defined by interpreting Definition 1.1.9 in \(\text{Cat}\) or in \(\text{CAT}\). Namely, an isomorphism of categories is given by a pair of inverse functors \(F : C \to D\) and \(G : D \to C\) so that the composites \(GF\) and \(FG\), respectively, equal the identity functors on \(C\) and on \(D\). An isomorphism induces a bijection between the objects of \(C\) and objects of \(D\) and likewise for the morphisms.

Example 1.3.14. For instance:

(i) The functor \((-)^{\text{op}} : \text{CAT} \to \text{CAT}\) defines a non-trivial automorphism of the category of categories. Note that a functor \(F : C \to D\) also defines a functor \(F : C^{\text{op}} \to D^{\text{op}}\).

(ii) For any group \(G\), the categories \(BG\) and \(B^{\text{op}}\) are isomorphic via the functor \((-)^{-1}\) that sends each morphism \(g \in G\) to its inverse. Any right action can be converted into a left action by precomposing with this isomorphism, which has the effect of “inserting inverses in the formula” defining the endomorphism associated to a particular group element.

(iii) Similarly, any groupoid is isomorphic to its opposite category via the functor that acts as the identity on objects and sends a morphism to its unique inverse morphism.

(iv) Any ring \(R\) has an opposite ring \(R^{\text{op}}\) with the same underlying abelian group but with the product of elements \(r\) and \(s\) in \(R^{\text{op}}\) defined to be the product \(s \cdot r\) of the elements \(s\) and \(r\) in \(R\). A left \(R\)-module is the same thing as a right \(R^{\text{op}}\)-module, which is to say there is a covariant isomorphism of categories \(\text{Mod}_R \cong R^{\text{op}}\text{-Mod}\) between the category of left \(R\)-modules and the category of right \(R^{\text{op}}\)-modules.

(v) For any space \(X\), there is a contravariant isomorphism of poset categories \(O(X) \cong C(X)^{\text{op}}\) that associates an open subset of \(X\) to its closed complement.
1.3. FUNCTORIALITY

Contrary to the impression created by Examples 1.3.14 (ii), (iii), and (vi), a category is not typically isomorphic to its opposite category.

Example 1.3.15. Let $E/F$ be a finite Galois extension: this means that $F$ is a finite-index subfield of $E$ and that the size of the group $\text{Aut}(E/F)$ of automorphisms of $E$ fixing every element of $F$ is at least (in fact, equal to) the index $[E:F]$. In this case, $G := \text{Aut}(E/F)$ is called the Galois group of the Galois extension $E/F$.

Consider the orbit category $O_G$ associated to the group $G$. Its objects are subgroups $H \subset G$, which we identify with the left $G$-set $G/H$ of left cosets of $H$. Morphisms $G/H \rightarrow G/K$ are $G$-equivariant maps, i.e., functions that commute with the left $G$-action. By an elementary exercise left to the reader, every morphism $G/H \rightarrow G/K$ has the form $gH \mapsto gyK$, where $\gamma \in G$ is an element so that $\gamma^{-1}Hy \subset K$.

Let $\text{Field}_F^E$ denote the subcategory of $F/\text{Field}$ whose objects are intermediate fields $F \subset K \subset E$. A morphism $K \rightarrow L$ is a field homomorphism that fixes the elements of $F$ pointwise. Note that the group of automorphisms of the object $E \in \text{Field}_F^E$ is the Galois group $G = \text{Aut}(E/F)$.

We define a functor $\Phi : O_G^{op} \rightarrow \text{Field}_F^E$ that sends $H \subset G$ to the subfield of $E$ of elements that are fixed by $H$ under the action of the Galois group. If $G/H \rightarrow G/K$ is induced by $\gamma$, then the field homomorphism $x \mapsto \gamma x$ sends an element $x \in E$ that is fixed by $K$ to an element $\gamma x \in E$ that is fixed by $H$. This defines the action of the functor $\Phi$ on morphisms. The fundamental theorem of Galois theory asserts that $\Phi$ defines a bijection on objects but in fact more is true: $\Phi$ defines an isomorphism of categories $O_G^{op} \cong \text{Field}_F^E$.

These examples aside, the notion of isomorphism of categories is somewhat unnatural. To illustrate, consider the category $\text{Set}^0$ of sets and partially-defined functions. A partial function $f : X \rightarrow Y$ is a function from a (possibly-empty) subset $X' \subset X$ to $Y$: the subset $X'$ is the domain of definition of the partial function $f$. The composite of two partial functions $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ is the partial function whose domain of definition is the intersection of the domain of definition of $f$ with the preimage of the domain of definition of $g$.

There is a functor $(-)_+ : \text{Set}^0 \rightarrow \text{Set}_*$, whose codomain is the category of pointed sets, that sends a set $X$ to the pointed set $X_+$, which is defined to be the disjoint union of $X$ with a freely-added basepoint. By the axiom of regularity, we might define $X_* := X \cup \{x\}$.28 A partial function $f : X \rightarrow Y$ gives rise to a pointed function $f_* : X_+ \rightarrow Y_+$ that sends every point outside of the domain of definition of $f$ to the formally added basepoint of $Y_+$. The inverse functor $U : \text{Set}_* \rightarrow \text{Set}^0$ discards the basepoint and sends a based function $f : (X,x) \rightarrow (Y,y)$ to the partial function $X\setminus\{x\} \rightarrow Y\setminus\{y\}$ with the maximal possible domain of definition.

By construction, we see that the composite $U(-)_+$ is the identity endofunctor of the category $\text{Set}^0$. By contrast, the other composite $U(-)_+ : \text{Set}_* \rightarrow \text{Set}$, sends a pointed set $(X,x)$ to $(X\setminus\{x\} \cup \{x\}, X\setminus\{x\})$. These sets are isomorphic but they are not identical. Nor is another set-theoretical construction of the “freely added basepoint” likely to define a genuine inverse to the functor $U : \text{Set}_* \rightarrow \text{Set}^0$. It is too restrictive to ask for the categories $\text{Set}^0$ and $\text{Set}_*$ to be isomorphic.

---

28In the axioms of Zermelo-Fraenkel set theory, elements of sets (like everything else in its mathematical universe) are themselves sets. The axiom of regularity prohibits a set from being an element of itself. As $X \not\in X$, we are free to add the element $X$ as a disjoint basepoint.
Indeed, there is a better way to decide whether two categories may safely be regarded as “the same.” To define it, we must relax the identities $GF = 1_C$ and $FG = 1_D$ between functors $F : C \to D$ and $G : D \to C$ that define an isomorphism of categories. This is possible because the collections $\text{Hom}(C, C)$ and $\text{Hom}(D, D)$ are not mere (possibly large) sets: they have higher-dimensional structure. For any pair of categories $C$ and $D$, the collection $\text{Hom}(C, D)$ of functors is itself a category. To explain this, we introduce what in French is called a morphisme de foncteurs, the notion that launched the entire subject of category theory: a natural transformation.

**Exercises.**

**Exercise 1.3.i.** What is a functor between groups, regarded as one-object categories?

**Exercise 1.3.ii.** What is a functor between preorders, regarded as categories?

**Exercise 1.3.iii.** Find an example to show that the objects and morphisms in the image of a functor $F : C \to D$ do not necessarily define a subcategory of $D$.

**Exercise 1.3.iv.** Verify that the constructions introduced in Definition 1.3.11 are functorial.

**Exercise 1.3.v.** What is the difference between a functor $C^{\text{op}} \to D$ and a functor $C \to D^{\text{op}}$? What is the difference between a functor $C \to D$ and a functor $C^{\text{op}} \to D^{\text{op}}$?

**Exercise 1.3.vi.** Given functors $F : D \to C$ and $G : E \to C$, show that there is a category, called the **comma category** $F \downarrow G$, which has

- as objects, triples $(d \in D, e \in E, f : Fd \to Ge \in C)$, and
- as morphisms $(d, e, f) \to (d', e', f')$, a pair of morphisms $(h : d \to d', k : e \to e')$ so that the square

$$
\begin{array}{ccc}
Fd & \overset{f}{\longrightarrow} & Ge \\
Fh \downarrow & & \downarrow Gk \\
Fd' & \overset{f'}{\longrightarrow} & Ge'
\end{array}
$$

commutes in $C$, i.e., so that $f' \cdot Fh = Gk \cdot f$.

Define a pair of projection functors $\text{dom} : F \downarrow G \to D$ and $\text{cod} : F \downarrow G \to E$.

**Exercise 1.3.vii.** Define functors to construct the slice categories $c/C$ and $C/c$ of Exercise 1.1.iii as special cases of comma categories constructed in Exercise 1.3.vi. What are the projection functors?

**Exercise 1.3.viii.** Lemma 1.3.8 shows that functors preserve isomorphisms. Find an example to demonstrate that functors need not reflect isomorphisms: that is, find a functor $F : C \to D$ and a morphism $f$ in $C$ so that $Ff$ is an isomorphism in $D$ but $f$ is not an isomorphism in $C$.

**Exercise 1.3.ix.** For any group $G$, we may define other groups:

- the **center** $Z(G) = \{ h \in G \mid hg = gh \forall g \in G \}$, a subgroup of $G$,
- the **commutator subgroup** $C(G)$, the subgroup of $G$ generated by elements $ghg^{-1}h^{-1}$ for any $g, h \in G$, and
- the **automorphism group** $\text{Aut}(G)$, the group of isomorphisms $\phi : G \to G$ in $\text{Group}$.

Trivially, all three constructions define a functor from the discrete category of groups (with only identity morphisms) to $\text{Group}$. Are these constructions functorial in

- the isomorphisms of groups? That is, do they extend to functors $\text{Group}_{\text{iso}} \to \text{Group}$?
• the epimorphisms of groups? That is, do they extend to functors $\text{Group}_{\text{epi}} \rightarrow \text{Group}$?
• all homomorphisms of groups? That is, do they extend to functors $\text{Group} \rightarrow \text{Group}$?

Exercise 1.3.x. Show that the construction of the set of conjugacy classes of elements of a group is functorial, defining a functor $\text{Conj}: \text{Group} \rightarrow \text{Set}$. Conclude that any pair of groups whose sets of conjugacy classes of elements have differing cardinalities cannot be isomorphic.

### 1.4. Naturality

It is not too misleading, at least historically, to say that categories are what one must define in order to define functors, and that functors are what one must define in order to define natural transformations. Peter Freyd, *Abelian categories* [Fre03]

Any finite-dimensional $k$-vector space $V$ is isomorphic to its **linear dual**, the vector space $V^* = \text{Hom}(V, k)$ of linear maps $V \rightarrow k$, because these vector spaces have the same dimension. This can be proven through the construction of an explicit **dual basis**: choose a basis $e_1, \ldots, e_n$ for $V$ and then define $e_1^*, \ldots, e_n^* \in V^*$ by

$$e_i^*(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

The collection $e_1^*, \ldots, e_n^*$ defines a basis for $V^*$ and the map $e_i \mapsto e_i^*$ extends by linearity to define an isomorphism $V \cong V^*$.

Now consider a related construction of the **double dual** $V^{**} = \text{Hom}(\text{Hom}(V, k), k)$ of $V$. If $V$ is finite dimensional, then the isomorphism $V \cong V^*$ is carried by the dual vector space functor $(-)^*: \text{Vect}_{\text{op}} \rightarrow \text{Vect}_k$ to an isomorphism $V^* \cong V^{**}$. The composite isomorphism $V \cong V^{**}$ sends the basis $e_1, \ldots, e_n$ to the dual dual basis $e_1^{**}, \ldots, e_n^{**}$.

As it turns out, this isomorphism has a simpler description. For any $v \in V$, the “evaluation function”

$$f \mapsto f(v): V^* \xrightarrow{\text{ev}_v} k$$

defines a linear functional on $V^*$. It turns out the assignment $v \mapsto \text{ev}_v$ defines a linear isomorphism $V \cong V^{**}$, this time requiring no “unnatural” choice of basis.\(^{30}\)

What distinguishes the isomorphism between a finite-dimensional vector space and its double dual from the isomorphism between a finite-dimensional vector space and its single dual is that the former assembles into the components of a natural transformation in the sense that we now introduce.

**Definition** 1.4.1. Given categories $C$ and $D$ and functors $F, G: C \Rightarrow D$, a **natural transformation** $\alpha: F \Rightarrow G$ consists of:

• an arrow $\alpha_c: Fc \rightarrow Gc$ in $D$ for each object $c \in C$, the collection of which define the **components** of the natural transformation,

\(^{29}\) A non-trivial theorem demonstrates that a homomorphism $\phi: G \rightarrow H$ is an epimorphism in $\text{Group}$ if and only if its underlying function is surjective; see [Lin70].

\(^{30}\) In fact, $e_i^{**}(e_j^*) = e_j^*(e_i) = \text{ev}_v(e_i^*)$, and so the two isomorphisms $V \cong V^{**}$ are the same—it is only our description that has improved.
so that, for any morphism \( f: c \to c' \) in \( C \), the following square of morphisms in \( D \)

\[
\begin{array}{ccc}
Fc & \xrightarrow{\alpha_c} & Gc \\
\downarrow_{\alpha'_c} & & \downarrow_{\alpha'_c} \\
Fc' & \xrightarrow{\alpha'_c} & Gc'
\end{array}
\]

(1.4.2)

commutes, i.e., has a common composite \( Fc \to Gc' \) in \( D \).

A \textbf{natural isomorphism} is a natural transformation \( \alpha: F \Rightarrow G \) in which every component \( \alpha_c \) is an isomorphism. In this case, the natural isomorphism may be depicted as \( \alpha: F \cong G \).

In practice, it is usually most elegant to define a natural transformation by saying that “the arrows \( X \) are natural,” which means that the collection of arrows defines the components of a natural transformation, leaving implicit the correct choices of domain and codomain functors, and source and target categories. Here \( X \) should be a collection of morphisms in a clearly identifiable (target) category, whose domains and codomains are defined using a common “variable” (an object of the source category). If this variable is \( c \), one might say “the arrows \( X \) are natural in \( c \)” to emphasize the domain object whose component is being described. However, the totality of the data of the source and target categories, the parallel pair of functors, and the components should always be considered part of the natural transformation. The naturality condition (1.4.2) cannot be stated precisely with any less: it refers to every object and every morphism in the domain category and is described using the images in the codomain category under the action of both functors. The “boundary data” needed to define a natural transformation \( \alpha \) is often displayed in a globular diagram:

\[
\begin{array}{ccc}
C & \xrightarrow{\alpha} & D \\
\downarrow & & \downarrow \\
G & & F
\end{array}
\]

The globular depiction of a natural transformation makes the notions of composable natural transformations that are introduced in §1.7 particularly intuitive.

\textbf{Example 1.4.3.}

(i) For vector spaces of any dimension, the map \( \text{ev}: V \to V^{**} \) that sends \( v \in V \) to the linear function \( \text{ev}_v: V^* \to \mathbb{k} \) defines the components of a natural transformation from the identity endofunctor on \( \text{Vect}_k \) to the double dual functor. To check that the naturality square

\[
\begin{array}{ccc}
V & \xrightarrow{\text{ev}} & V^{**} \\
\downarrow_{\phi} & & \downarrow_{\phi^{**}} \\
W & \xrightarrow{\text{ev}} & W^{**}
\end{array}
\]

commutes for any linear map \( \phi: V \to W \), it suffices to consider the image of a generic vector \( v \in V \). By definition, \( \text{ev}_v: V^* \to \mathbb{k} \) defines a functional \( f: W \to \mathbb{k} \) to \( f(\phi v) \). Recalling the definition of the action of the dual functor of Example 1.3.7(ii) on morphisms, we see that \( \phi^{**}(\text{ev}_v): W^* \to \mathbb{k} \) carries a functional \( f: W \to \mathbb{k} \) to \( f(\phi v) \), which amounts to the same thing.

(ii) By contrast, the identity functor and the single dual functor on finite-dimensional vector spaces are not naturally isomorphic. One technical obstruction is somewhat
beside the point: the identity functor is covariant while the dual functor is contravariant.\footnote{A more flexible notion of extranatural transformation can accommodate functors with conflicting variance \cite[IX.4]{ML98}, see Exercise 1.4.vi.} More significant is the essential failure of naturality. The isomorphisms $V \cong V^*$ that can be defined whenever $V$ is finite dimensional require a choice of basis, which is preserved by essentially no linear maps, indeed by no non-identity linear endomorphism.\footnote{A proof that there exists no extranatural isomorphism between the identity and dual functors on the categories of finite-dimensional vector spaces is given in \cite[p. 234]{EM45}.}

(iii) There is a natural transformation $\eta: 1_{\text{Set}} \Rightarrow P$ from the identity to the covariant power set functor whose components $\eta_A: A \to PA$ are the functions that carry $a \in A$ to the singleton subset $\{a\} \in PA$.

(iv) For $G$ a group, Example 1.3.9 shows that a functor $X: BG \to C$ corresponds to an object $X \in C$ equipped with a left action of $G$, which suggests a question: What is a natural transformation between a pair $X, Y: BG \Rightarrow C$ of such functors? Because the category $BG$ has only one object, the data of $\alpha: X \Rightarrow Y$ consists of a single morphism $\alpha: X \to Y$ in $C$ that is $G$-equivariant, meaning that for each $g \in G$, the diagram

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

commutes.

(v) The open and closed subset functors described in Example 1.3.7(iii) are naturally isomorphic when they are regarded as functors $O, C: \text{Top} \Rightarrow \text{Set}$ valued in the category of sets. The components $O(X) \cong C(X)$ of the natural isomorphism are defined by taking an open subset of $X$ to its complement, which is closed. Naturality asserts that the process of forming complements commutes with the operation of taking preimages.

(vi) The construction of the opposite group described in Example 1.2.2(iii) defines a (covariant!) endofunctor $(-)^{\text{op}}: \text{Group} \to \text{Group}$ of the category of groups; a homomorphism $\phi: G \to H$ induces a homomorphism $\phi^{\text{op}}: G^{\text{op}} \to H^{\text{op}}$ defined by $\phi^{\text{op}}(g) = \phi(g)$. This functor is naturally isomorphic to the identity. Define $\eta_G: G \to G^{\text{op}}$ to be the homomorphism that sends $g \in G$ to its inverse $g^{-1} \in G^{\text{op}}$; this mapping does not define an automorphism of $G$, because it fails to commute with the group multiplication, but it does define a homomorphism $G \to G^{\text{op}}$. Now given any homomorphism $\phi: G \to H$, the diagram

$$
\begin{array}{ccc}
G & \xrightarrow{\eta_G} & G^{\text{op}} \\
\downarrow \phi & & \downarrow \phi^{\text{op}} \\
H & \xrightarrow{\eta_H} & H^{\text{op}}
\end{array}
$$

commutes because $\phi^{\text{op}}(g^{-1}) = \phi(g)^{-1}$.

(vii) Define an endofunctor of $\text{Vect}_k$ by $V \mapsto V \otimes V$. There is a natural transformation from the identity functor to this endofunctor whose components are the zero maps, but this is the only such natural transformation: there is no basis-independent way to define a linear map $V \to V \otimes V$. The same result is true for the category of Hilbert
spaces and linear operators between them, in which context it is related to the “no cloning theorem” in quantum physics.\footnote{The states in a quantum mechanical system are modeled by vectors in a Hilbert space and the observables are operators on that space. See \cite{Bae06} for more.}

Another familiar isomorphism that is not natural arises in the classification of finitely generated abelian groups, objects of a category $\text{Ab}_{fg}$. Let $TA$ denote the torsion subgroup of an abelian group $A$, the subgroup of elements with finite order. In classifying finitely generated groups, one proves that every finitely generated abelian group $A$ is isomorphic to the direct sum $TA \oplus (A/TA)$, the summand $A/TA$ being the torsion-free part of $A$. However, these isomorphisms are not natural, as we now demonstrate.

**Proposition 1.4.4.** The isomorphisms $A \cong TA \oplus (A/TA)$ are not natural\footnote{Any finitely generated abelian group $A$ has a short exact sequence $0 \to TA \to A \to A/TA \to 0$. Proposition 1.4.4 asserts that there is no natural splitting.} in $A \in \text{Ab}_{fg}$.

**Proof.** Suppose the isomorphisms $A \cong TA \oplus (A/TA)$ were natural in $A$. Then the composite

\[(1.4.5) \quad A \to A/TA \to TA \oplus (A/TA) \cong A\]

of the canonical quotient map, the inclusion into the direct sum, and the hypothesized natural isomorphism would define a natural endomorphism of the identity functor on $\text{Ab}_{fg}$. We shall see that this is impossible.

To derive the contradiction, we first show that every natural endomorphism of the identity functor on $\text{Ab}_{fg}$ is multiplication by some $n \in \mathbb{Z}$. Clearly the component of $\alpha : 1_{\text{Ab}_{fg}} \Rightarrow 1_{\text{Ab}_{fg}}$ at $\mathbb{Z}$ has this description for some $n$. Now observe that homomorphisms $\mathbb{Z} \to A$ correspond bijectively to elements $a \in A$, choosing $a$ to be the image of $1 \in \mathbb{Z}$. Thus, commutativity of

\[
\begin{array}{ccc}
\mathbb{Z} & \xrightarrow{\alpha = n\cdot} & \mathbb{Z} \\
\downarrow{a} & & \downarrow{a} \\
A & \xrightarrow{\alpha_A} & A
\end{array}
\]

forces us to define $\alpha_A(a) = n \cdot a$.

In the case where $\alpha$ is the natural transformation defined by (1.4.5), by examining the component at $A = \mathbb{Z}$, we can see that $n \neq 0$. Finally, consider $A = \mathbb{Z}/2n\mathbb{Z}$. This group is torsion, so any map, such as $\alpha_{\mathbb{Z}/2n\mathbb{Z}}$, which factors through the quotient by its torsion subgroup is zero. But $n \neq 0 \in \mathbb{Z}/2n\mathbb{Z}$, a contradiction. \hfill $\square$

**Example 1.4.6.** The Riesz representation theorem can be expressed as a natural isomorphism of functors from the category $\text{cHaus}$ of compact Hausdorff spaces and continuous maps to the category $\text{Ban}$ of real Banach spaces and continuous linear maps. Let $\Sigma : \text{cHaus} \to \text{Ban}$ be the functor that carries a compact Hausdorff space $X$ to the Banach space $\Sigma(X)$ of signed Baire measures on $X$ and sends a continuous map $f : X \to Y$ to the map $\mu \mapsto \mu \circ f^{-1} : \Sigma(X) \to \Sigma(Y)$. Let $C^* : \text{cHaus} \to \text{Ban}$ be the functor that carries $X$ to the linear dual $C(X)^*$ of the Banach space $C(X)$ of continuous real-valued functions on $X$.

Now for each $\mu \in \Sigma(X)$, there is a linear functional $\phi_{\mu} : C(X) \to \mathbb{R}$ defined by

\[\phi_{\mu}(g) := \int_X g \, d\mu, \quad g \in C(X).\]
For each $\mu \in \Sigma(X)$, $f : X \to Y$, and $h \in C(Y)$ the equation
\[
\int_X hf \, d\mu = \int_Y h \, d(\mu \circ f^{-1})
\]
says that the assignment $\mu \mapsto \phi_{\mu}$ defines the components of a natural transformation $\eta : \Sigma \Rightarrow C^*$. The Riesz representation theorem asserts that this natural transformation is a natural isomorphism; see [Har83].

**Example 1.4.7.** Consider morphisms $f : w \to x$ and $h : y \to z$ in a locally small category $C$. Post-composition by $h$ and pre-composition by $f$ define functions between hom-sets
\[
C(x, y) \xrightarrow{h \cdot -} C(x, z) \quad - \cdot f \quad C(w, y) \xrightarrow{h \cdot -} C(w, z)
\]
(1.4.8)

In Definition 1.3.13 and elsewhere, $h \cdot -$ was denoted by $h_*$ and $- \cdot f$ was denoted by $f^*$, but we find this less-concise notation to be more evocative here. Associativity of composition implies that this diagram commutes: for any $g : x \to y$, the common image is $h g f : w \to z$.

Interpreting the vertical arrows as the images of $f$ under the actions of the functors $C(\cdot, y)$ and $C(\cdot, z)$, the square (1.4.8) demonstrates that there is a natural transformation $h_* : C(\cdot, y) \Rightarrow C(\cdot, z)$

whose components are defined by post-composition with $h : y \to z$. Flipping perspectives and interpreting the horizontal arrows as the images of $h$ under the actions of the functors $C(x, \cdot)$ and $C(w, \cdot)$, the square (1.4.8) demonstrates that there is a natural transformation $f^* : C(x, \cdot) \Rightarrow C(w, \cdot)$

whose components are defined by pre-composition with $f : w \to x$.

A final example describes the natural isomorphisms that supply proofs of the fundamental laws of elementary arithmetic.

**Example 1.4.9 (a categorification of the natural numbers).** For sets $A$ and $B$, let $A \times B$ denote their cartesian product, let $A + B$ denote their disjoint union, and let $AB$ denote the set of functions from $B$ to $A$. These constructions are related by natural isomorphisms

\[
\begin{align*}
A \times (B + C) & \cong (A \times B) + (A \times C) \\
A^{B+C} & \cong A^B \times A^C \\
(A \times B)^C & \cong A^C \times B^C \\
(A^B)^C & \cong A^{B \times C}
\end{align*}
\]
(1.4.10)

In the first instance, the isomorphism defines the components of a natural transformation between a pair of functors $\text{Set} \times \text{Set} \times \text{Set} \to \text{Set}$. For the others, the variance in the variables appearing as “exponents” is contravariant. This is because the assignment $(B, A) \mapsto A^B$ defines a functor $\text{Set}^{op} \times \text{Set} \to \text{Set}$, namely the two-sided represented functor introduced in Definition 1.3.13.

The displayed natural isomorphisms restrict to the category $\text{Fin}_{\text{iso}}$ of finite sets and bijections, a category which serves as the domain for the cardinality functor $\lvert - \rvert : \text{Fin}_{\text{iso}} \to \mathbb{N}$, whose codomain is the discrete category of natural numbers.\[^{35}\]

\[^{35}\text{Mathematical invariants often take the form of a functor from a groupoid to a discrete category.}\]
1. CATEGORIES, FUNCTORS, NATURAL TRANSFORMATIONS

and $c = |C|$, the cardinality functor carries these natural isomorphisms to the equations

$$a \times (b + c) = (a \times b) + (a \times c)$$

$$a^{b+c} = a^b \times a^c$$

$$(a \times b)^c = a^c \times b^c$$

$$(a^b)^c = a^{b \times c}$$

through a process called **decategorification**. Reversing directions, $\text{Fin}_{\text{iso}}$ is a categorification of the natural numbers, which reveals that the familiar laws of arithmetic follow from more fundamental natural isomorphisms between various constructions on sets. A slick proof of each of the displayed natural isomorphisms (1.4.10) appears in Corollary 4.5.6.

**Exercises.**

**Exercise 1.4.i.** Suppose $\alpha: F \Rightarrow G$ is a natural isomorphism. Show that the inverses of the component morphisms define the components of a natural isomorphism $\alpha^{-1}: G \Rightarrow F$.

**Exercise 1.4.ii.** What is a natural transformation between a parallel pair of functors between groups, regarded as one-object categories?

**Exercise 1.4.iii.** What is a natural transformation between a parallel pair of functors between preorders, regarded as categories?

**Exercise 1.4.iv.** In the notation of Example 1.4.7, prove that distinct parallel morphisms $f, g: c \Rightarrow d$ define distinct natural transformations

$$f_*, g_*: C(-, c) \Rightarrow C(-, d) \quad \text{and} \quad f^*, g^*: C(d, -) \Rightarrow C(c, -)$$

by post- and pre-composition.

**Exercise 1.4.v.** Recall the construction of the comma category for any pair of functors $F: D \rightarrow C$ and $G: E \rightarrow C$ described in Exercise 1.3.vi. From this data, construct a canonical natural transformation $\alpha: F \text{dom} \Rightarrow G \text{cod}$ between the functors that form the boundary of the square

$$
\begin{array}{ccc}
D & \xrightarrow{F} & E \\
\downarrow & & \downarrow \\
\text{dom} & \xrightarrow{\alpha} & \text{cod} \\
\downarrow & & \downarrow \\
F & \xrightarrow{G} & C
\end{array}
$$

**Exercise 1.4.vi.** Given a pair of functors $F: A \times B \times B^{\text{op}} \rightarrow D$ and $G: A \times C \times C^{\text{op}} \rightarrow D$, a family of morphisms

$$a_{a,b,c}: F(a, b, b) \rightarrow G(a, c, c)$$

in $D$ defines the components of an **extranatural transformation** $\alpha: F \Rightarrow G$ if for any $f: a \rightarrow a'$, $g: b \rightarrow b'$, and $h: c \rightarrow c'$ the following diagrams commute in $D$:

$$
\begin{array}{ccc}
F(a, b, b) & \xrightarrow{a_{a,b,c}} & G(a, c, c) \\
\downarrow & & \downarrow \\
F(f, 1, b) & \xrightarrow{F(f, 1, b)} & F(a, b, b) \\
\downarrow & & \downarrow \\
g(1, 1, b) & \xrightarrow{g(1, 1, b)} & G(a, c, c) \\
\downarrow & & \downarrow \\
a_{a,b,c} & \xrightarrow{a_{a,b,c}} & a_{a,b,c}
\end{array}
$$

$$
\begin{array}{ccc}
F(a', b', b') & \xrightarrow{a_{a',b',c}} & G(a', c, c) \\
\downarrow & & \downarrow \\
F(f, 1, b) & \xrightarrow{F(f, 1, b)} & F(a, b, b) \\
\downarrow & & \downarrow \\
g(1, 1, b) & \xrightarrow{g(1, 1, b)} & G(a, c, c) \\
\downarrow & & \downarrow \\
a_{a,b,c} & \xrightarrow{a_{a,b,c}} & a_{a,b,c}
\end{array}
$$

The left-hand square asserts that the components $\alpha_{a,-,c}: F(\cdot, b, b) \Rightarrow G(\cdot, c, c)$ define a natural transformation in $a$ for each $b \in B$ and $c \in C$. The remaining squares assert that the components $\alpha_{a,-,c}: F(a, \cdot, \cdot) \Rightarrow G(a, c, c)$ and $\alpha_{a,b,-}: F(a, b, \cdot) \Rightarrow G(\cdot, \cdot, \cdot)$ define transformations that are respectively extranatural in $b$ and in $c$. Explain why the functors $F$ and $G$ must have a common target category for this definition to make sense.
Category theory has provided the foundations for many of the twentieth century’s greatest advances in pure mathematics. This concise, original text for a one-semester introduction to the subject is derived from courses that author Emily Riehl taught at Harvard and Johns Hopkins Universities. The treatment introduces the essential concepts of category theory: categories, functors, natural transformations, the Yoneda lemma, limits and colimits, adjunctions, monads, Kan extensions, and other topics.

Suitable for advanced undergraduates and graduate students in mathematics, the text provides tools for understanding and attacking difficult problems in algebra, number theory, algebraic geometry, and algebraic topology. Drawing upon a broad range of mathematical examples from the categorical perspective, the author illustrates how the concepts and constructions of category theory arise from and illuminate more basic mathematical ideas. While the reader will be rewarded for familiarity with these background mathematical contexts, essential prerequisites are limited to basic set theory and logic.