A CONCRETE INTRODUCTION TO CATEGORIES

WILLIAM R. SCHMITT
DEPARTMENT OF MATHEMATICS
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, D.C. 20052

Contents

1. Categories

1.1. First Definition and Examples.

Definition 1.1. A *category* $\mathcal{C}$ consists of the following data:

(i) A set $\mathrm{Ob}(\mathcal{C})$ of *objects*.
(ii) For every pair of objects $a, b \in \mathrm{Ob}(\mathcal{C})$, a set $\mathcal{C}(a, b)$ of *arrows*, or *morphisms*, *from* a *to* b.
(iii) For all triples $a, b, c \in \mathrm{Ob}(\mathcal{C})$, a *composition map*

$$\begin{array}{ccc} \mathcal{C}(a,b) \times \mathcal{C}(b,c) & \to & \mathcal{C}(a,c) \\ (f,g) & \mapsto & gf = g \cdot f. \end{array}$$

(iv) For each object $a \in \mathrm{Ob}(\mathcal{C})$, an arrow $1_a \in \mathcal{C}(a, a)$, called the *identity* of a.

These data are subject to the following axioms:

Associativity: $h(gf) = (hg)f$, for all $(f, g, h) \in \mathcal{C}(a, b) \times \mathcal{C}(b, c) \times \mathcal{C}(c, d)$.

Identity: $f = f \cdot 1_a = 1_b \cdot f$, for all $f \in \mathcal{C}(a, b)$.

Disjointness: $\mathcal{C}(a, b) \cap \mathcal{C}(a', b') = \emptyset$, if $(a, b) \neq (a', b')$ in $\mathrm{Ob}(\mathcal{C}) \times \mathrm{Ob}(\mathcal{C})$.

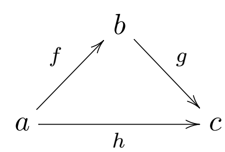
We usually write $f\colon a \to b$, or $a \xrightarrow{f} b$, to indicate that an arrow f belongs to $\mathcal{C}(a, b)$; in this case the object a is called the *domain*, or *source*, of f and written $\mathrm{dom\,f}$, and b is the *codomain*, or *target*, of f and written $\mathrm{cod\,f}$. We often write $a \in \mathcal{C}$ and $f \in \mathcal{C}$ to indicate that a is an object and f is an arrow belonging to $\mathcal{C}$; since we usually denote objects by the letters $a, b, c, \ldots$ and $z, y, x, \ldots$ and morphisms by the letters $f, g, h, \ldots$ this should never cause any unnecessary confusion.

The usual proof $(1_a = 1_a \cdot 1'_a = 1'_a)$ shows that the identity $1_a \in \mathcal{C}(a, a)$ is unique, justifying the use of the word "the" in (iv) above.

The prototypical example is the category **Set**, whose objects are all sets and arrows are all functions. For the sake of precision, let us recall the usual definition of a function from a set S to a set T; it is a subset f of the cartesian product $S \times T$ such that the composition $f \hookrightarrow S \times T \to S$ is a bijection (where the first map is the inclusion and the second is the projection onto the first factor). This definition is elegant but, for example, if S is a proper subset of T then it fails to distinguish between the inclusion map $S \hookrightarrow T$ and the identity map $1_S\colon S \to S$, since they consist of precisely the same set of ordered pairs. Thus we have $\mathbf{Set}(S, T) \cap \mathbf{Set}(S, S) \neq \emptyset$, in violation of the disjointness axiom for arrow sets. We will take the usual way around this problem, which is to define a function $f\colon S \to T$ as an ordered triple $(\bar{f}, S, T)$, where $\bar{f}$ is a function in the above sense. This redefinition affects nothing in practice, but ensures that the disjointness property holds.

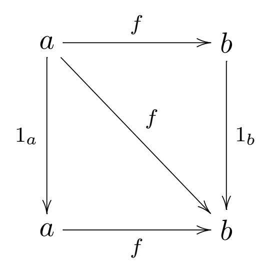
Definition 1.2. A *diagram* in a category $\mathcal{C}$ is a directed graph with vertices labelled by objects, and edges labelled by arrows of $\mathcal{C}$. A diagram is *commutative* if, for every pair of vertices a and b belonging to it, the labels along any two directed paths from a to b compose to give the same morphism in $\mathcal{C}$.

For example, each composable pair of arrows $a \xrightarrow{f} b \xrightarrow{g} c$ in $\mathcal{C}$ corresponds to a unique commutative triangle

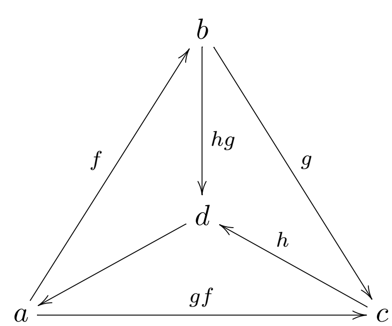


where $h = gf$ in $\mathcal{C}$.

The identity axiom for categories is equivalent to the commutativity of the diagram



for all arrows $f: a \to b$ in $\mathcal{C}$, and the associativity axiom is expressed by the commutativity of



for all composable triples (h, g, f). The unlabelled arrow here is the triple composition $h(gf) = (hg)f$, which may be denoted simply as hgf.

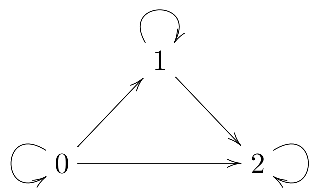
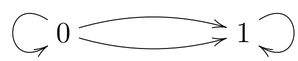
Here are some examples of categories:

Name	Objects	Arrows
Set	sets	functions
$\mathbf{Set}_*$	sets with distinguished base-point	base-point-preserving functions
Mon	monoids	monoid homomorphisms
Grp	groups	group homomorphisms
Ab	abelian groups	group homomorphisms
Rng	rings	ring homomorphisms
ComRng	commutative rings with 1	ring homomorphisms
${}_R\mathbf{Mod}$	left R-modules (where R is some fixed ring	R-linear maps
$\mathbf{Mod}_R$	right R-modules	R-linear maps
$\mathbf{Vect}_K$	K-vector spaces	K-linear maps
Top	topological spaces	continuous maps
Man	smooth manifolds	smooth (infinitely differentiable) maps

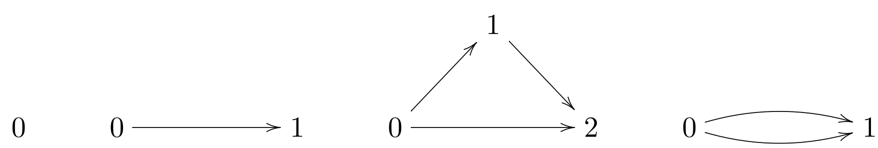
All of the categories in the above table are examples of *concrete* categories, that is, categories in which objects are sets with additional structure, arrows are structure-preserving functions, and the composition law is ordinary composition of functions. The *empty* category **0**, which is the unique category having no objects and no morphisms, is also a concrete category. The verification that a concrete category is indeed a category is usually quite simple; one merely needs to check that all identity arrows are present and that composition of two functions of the given type yields a function of the same type. Associativity, which is often the most difficult axiom to check, holds automatically since functional composition is always associative.

Example 1.3. A *discrete* category is one whose only arrows are the identity arrows. A discrete category is completely determined by its set of objects; thus discrete categories and sets are essentially the same notion. A discrete category is not necessarily concrete.

If a category is small enough, then it may be represented in its entirety by a diagram. Here are some examples of such categories that occur frequently enough so that they have standard names:

Name	Category
1	
2	
3	
$\upuparrows$	

We usually omit the identity arrows from such drawings, since they are redundant; hence the above categories may be depicted as



Example 1.4. A *monoid* is a category $\mathcal{M}$ having precisely one object. In such a category, the object, say $*$, is of no interest (at least from the point of view of category theory). All of the structure of $\mathcal{M}$ resides in its set of arrows $\mathcal{M}(*, *)$, which is a set together with an associative binary operation and identity element for this operation, that is, a monoid by the traditional definition.

Definition 1.5. An arrow $f \colon a \to b$ in a category $\mathcal{C}$ is an *isomorphism* if there exists $g \colon b \to a$ such that $fg = 1_b$ and $gf = 1_a$. Objects a and b are *isomorphic* if there exists an isomorphism between them, in which case we write $a \cong b$.

We also sometimes refer to isomorphisms as *invertible* arrows. Note that identity arrows are isomorphisms. It is immediate that isomorphism is an equivalence relation on the object set of a category. In each of the concrete categories above, this notion of isomorphism corresponds to the usual one for that class of objects; for example, isomorphic objects in **Top** are homeomorphic topological spaces, isomorphic objects in **Grp** are isomorphic groups, and isomorphic objects in **Set** are sets having the same cardinality.

From the point of view of category theory, isomorphic objects are indistinguishable and may thus be regarded as 'the same' for most purposes. Rather than go ahead and identify isomorphic objects, however, we usually take a different approach and dispense with the equality relation on objects, using isomorphism in its place. This point of view is nicely summed up by the saying "doing category theory means never having to say 'equal' ". Indeed, throughout the subject, we repeatedly will encounter constructions that are defined only up to some (canonical) isomorphism.

Example 1.6. A *group* is a category having one object, in which every arrow is an isomorphism. As in the previous example, the set of arrows of such a category corresponds to the traditional definition.

It may seem unnecessarily abstract to define a monoid or group as above, but in fact this definition is in a sense *less* abstract than the usual definition. These structures were originally defined as they arose 'in nature', as sets of transformations of a given object (e.g., functions from a set to itself, or linear maps from a vector space to itself) that were closed under composition. The categorical definitions of monoid and group retain these underlying objects, and are thus much closer to the original, concrete, definitions.

Example 1.7. A *groupoid* is a category in which every arrow is an isomorphism. Hence a groupoid consists of a set of groups, some pairs of which are connected by isomorphisms. One example of a groupoid is the category **Bij** of sets and bijections. The groupoid **FinBij** of finite sets and bijections is of central importance in combinatorics. A different sort of example is the *fundamental* groupoid $\pi(T)$ of a topological space T, which has points of T as objects, with certain equivalence classes, called *homotopy classes*, of oriented paths between points as arrows. If we choose a single *base-point* x in T, then the group in $\pi(T)$ corresponding to the object x is called the *fundamental group* of the pointed space (T, x).

So a groupoid is to a group what an arbitrary category is to a monoid; thus it might make sense to call categories 'monoidoids', which would be a pretty unfortunate choice of terminology.

Example 1.8. A *preorder*, or *quasiorder* is a category $\mathcal{P}$ having the property that $|\mathcal{P}(a, b)| \leq 1$, for all $a, b \in \mathcal{P}$. If $\mathcal{P}$ is a preorder, then the relation $\leq$ defined on set of objects $\mathrm{Ob}(\mathcal{P})$ by setting $a \leq b$ if and only if $\mathcal{P}(a, b) \neq \emptyset$ is easily seen to be reflexive and transitive, and hence $\mathrm{Ob}(\mathcal{P})$ is a preorder in the usual sense.

Example 1.9. A category $\mathcal{P}$ is a *partial order*, or *poset*, if $|\mathcal{P}(a, b) \cup \mathcal{P}(b, a)| \leq 1$, for all $a, b \in \mathcal{P}$. If $\mathcal{P}$ is a partial order, then the object set $\mathrm{Ob}(\mathcal{P})$, equipped with the order relation $\leq$ as in the previous example, is a poset in the usual sense.

The categories **0**, **1**, **2**, and **3** are all partial orders (linear, or total, orders, in fact); they are the first four *ordinal numbers*, the nth of which may be regarded as the set $\{0, 1, \ldots, n-1\}$ equipped with the usual ordering.

1.2. **An Alternative Definition: The Arrows-Only Perspective.**

Definition 1.10. A (directed) *graph* $\mathcal{C}$ consists of a set $A = \mathrm{Ar}(\mathcal{C})$ of *arrows*, a set $O = \mathrm{Ob}(\mathcal{C})$ of *objects* and two functions

$$\mathrm{dom} : A \to O \qquad \text{and} \qquad \mathrm{cod} : A \to O,$$

called *domain* and *codomain*, respectively. The set of *composable* pairs of arrows in $\mathcal{C}$ is the subset

$$A \times_O A = \{(f, g) \colon \mathrm{dom}\, f = \mathrm{cod}\, g\}$$

of $A \times A$.

Definition 1.11. A *category* is a graph $\mathcal{C}$, as above, equipped with a *composition* function $\mathrm{comp} \colon A \times_O A \to A$ and *unit* function $\mathrm{id} \colon O \to A$, denoted by $(f, g) \mapsto fg$ and $a \mapsto 1_a$, respectively, such that

(i) $\mathrm{dom} \cdot \mathrm{id} = 1_O = \mathrm{cod} \cdot \mathrm{id}$.
(ii) $\mathrm{dom}\,(fg) = \mathrm{dom}\, g$ and $\mathrm{cod}\,(fg) = \mathrm{cod}\, f$, for all composable (f, g).
(iii) $f \cdot 1_{\mathrm{dom}\, f} = f$ and $1_{\mathrm{cod}\, f} \cdot f = f$, for all arrows f.
(iv) $(fg)h = f(gh)$, whenever (f, g) and (g, h) are composable.

The arrows 1_a are called *identity* arrows. The first axiom says that the arrow 1_a has domain and codomain equal to a.

If we let $\mathcal{C}(a, b) = \{f \in \mathrm{Ar}(\mathcal{C}) \colon \mathrm{dom}\, f = a \text{ and } \mathrm{cod}\, f = b\}$, then (ii) means that composition restricts to a map $\mathcal{C}(b, c) \times \mathcal{C}(a, b) \to \mathcal{C}(a, c)$. It thus follows from the last two axioms that the set O together with the sets $\mathcal{C}(a, b)$, and the given composition law and identity arrows, forms a category according to the first definition. On the other hand, given a category $\mathcal{C}$ by our first definition, we let $\mathrm{Ar}(\mathcal{C})$ be the union of all the sets $\mathcal{C}(a, b)$, and let dom and cod be the usual domain and codomain functions, and we see that $\mathcal{C}$ satisfies the second definition.

Property (i) implies that the map $\mathrm{id} \colon O \to A$ is injective. Hence we may identify O with its image, thus regarding set of objects as a subset of the set of arrows, namely, as the set of identity arrows. We may thus regard a category as a generalization of a monoid, that is, as a set together with a partially defined binary operation and identity elements, satisfying certain properties (see Exercise 1).

It often will be convenient to adopt this perspective, viewing categories as abstract algebraic entities rather than as classes of mathematical objects having their own internal structure. For example, from this *arrows only* point of view, there is no difference between the categorical and traditional definitions of monoid and group, and a discrete category is precisely the same thing as a set. This approach facilitates the definition of many standard constructions in category theory (a good example of this is the definition of subcategory, below). From now on, we will shift between these two points of view as is convenient.

1.3. **Some Constructions.**

Definition 1.12. A *subcategory* of a category $\mathcal{C}$ is a subset $\mathcal{D}$ of $\mathcal{C}$ that is closed under composition and formation of domains and codomains. We write $\mathcal{D} \subseteq \mathcal{C}$ to indicate that $\mathcal{D}$ is a subcategory of $\mathcal{D}$.

Alternatively, we may define a category $\mathcal{D}$ to be a subcategory of $\mathcal{C}$ if $\mathrm{Ob}(\mathcal{D}) \subseteq \mathrm{Ob}(\mathcal{C})$ and $\mathrm{Ar}(\mathcal{D}) \subseteq \mathrm{Ar}(\mathcal{C})$ such that composition and identities in $\mathcal{D}$ correspond to those in $\mathcal{C}$.

The subcategory $\mathcal{D}$ of $\mathcal{C}$ is *full* if $\mathcal{D}(a,b) = \mathcal{C}(a,b)$, for all objects a and b of $\mathcal{D}$. Note that a full subcategory is completely determined by its set of objects. For example, **FinSet** and **Grp** are full subcategories of **Set** and **Mon**, respectively. The categories **Inj**, **Surj** and **Bij**, which have all sets as objects and, respectively, all injections, surjections and bijections as arrows, are subcategories of **Set** that are not full.

If $\{\mathcal{D}_i \colon i \in I\}$ is a collection of subcategories of $\mathcal{C}$ then the intersection $\bigcap_{i\in I} \mathcal{D}_i$ is also a subcategory. Note that

$$\mathrm{Ob}\left(\cap_{i\in I}\mathcal{D}_i\right) = \cap_{i\in I}\mathrm{Ob}(\mathcal{D}_i) \qquad \text{and} \qquad \mathrm{Ar}\left(\cap_{i\in I}\mathcal{D}_i\right) = \cap_{i\in I}\mathrm{Ar}(\mathcal{D}_i).$$

If S is any collection of arrows in $\mathcal{C}$, the subcategory *generated by* S, denoted by $\langle S\rangle$, is the intersection of all subcategories of $\mathcal{C}$ whose arrow sets contain S. Observe that

$$\mathrm{Ob}(\langle S\rangle) = \{\mathrm{dom}\, f \colon f \in S\} \cup \{\mathrm{cod}\, f \colon f \in S\}$$

and $\mathrm{Ar}(\langle S\rangle)$ is the set of all arrows that can be obtained by composing in $\mathcal{C}$ (composable) sequences of arrows belonging S.

Definition 1.13. If $\{\mathcal{D}_i \colon i \in I\}$ is a collection of subcategories of $\mathcal{C}$, then the *join* $\bigvee_{i\in I} \mathcal{D}_i$ is the subcategory of $\mathcal{C}$ generated by the union $\bigcup_{i\in I} \mathcal{D}_i$.

Example 1.14. Since any function $S \xrightarrow{f} T$ between sets factors canonically as a surjection (onto $f(S)$), followed by an injection (the inclusion of $f(S)$ into T), it follows that the category **Set** is equal to the join **Inj** $\vee$ **Surj**.

The intersection of a family $\{\mathcal{D}_i \colon i \in I\}$ of subcategories of $\mathcal{C}$ is the largest subcategory of $\mathcal{C}$ that is contained in each $\mathcal{D}_i$, and the join is the smallest subcategory that contains each $\mathcal{D}_i$. Hence the collection of all subcategories of a category is a complete lattice, partially ordered by inclusion, with least upper bounds given by joins and greatest lower bounds given by intersections.

Definition 1.15. If $\mathcal{C}$ is a category, then the *opposite* category $\mathcal{C}^{\mathrm{op}}$ has the same objects and arrows as $\mathcal{C}$, but with $\mathcal{C}^{\mathrm{op}}(a,b) = \mathcal{C}(b,a)$, for all objects a and b, and if $f \colon a \to b$ and $g \colon b \to c$ in $\mathcal{C}$, then the composition fg in $\mathcal{C}^{\mathrm{op}}$ defined to be the composition gf in $\mathcal{C}$.

We write f^{op} for an arrow $f \in \mathcal{C}(b,a)$ whenever we want to regard f as an arrow in $\mathcal{C}^{\mathrm{op}}$, so that

$$f^{\mathrm{op}} \colon a \to b \text{ in } \mathcal{C}^{\mathrm{op}} \quad \Leftrightarrow \quad f \colon b \to a \text{ in } \mathcal{C}.$$

The reader should bear in mind that f and f^{op} denote the same element of the set $\mathcal{C}(b,a) = \mathcal{C}^{\text{op}}(a,b)$, but when regarded as arrows in the categories $\mathcal{C}$ and $\mathcal{C}^{\text{op}}$, respectively, they differ because the domain of each is the codomain of the other. Using this notation, the definition of composition in $\mathcal{C}^{\text{op}}$ becomes much clearer; it is simply

$$f^{\text{op}} g^{\text{op}} = (gf)^{\text{op}}.$$

If P is a property that objects and arrows in a category may have, then the *dual* of P is the property of having P in the opposite category. For example, an object s is initial in $\mathcal{C}$ if and only if s is terminal in $\mathcal{C}^{\text{op}}$. Thus being terminal and initial are dual properties. Also, an arrow is monic in $\mathcal{C}$ if and only if it is epi in $\mathcal{C}^{\text{op}}$, and hence being monic and epi are dual properties. The property of being an isomorphism is self-dual.

The *duality principle* for category theory states that if a statement that is built from atomic statements (e.g., '$a = b$', '$f = g$', '1 is the identity of a', etc.) using the usual connectives and quantifiers, follows from the axioms (i.e. is true in any category), then the dual statement is also true in any category. In this manner, we obtain dual theorems for free. For example, having proved that initial objects are unique up to isomorphism, we have the dual result that terminal objects are unique up to isomorphism. Another example, is given by Propositions 1.20 and 1.21, which are dual to one another.

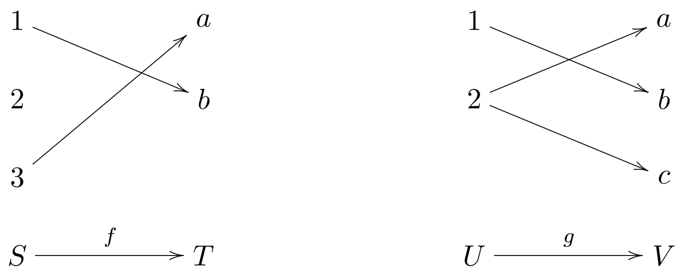
Beware, though, that if a statement is only true in some categories, and happens to be true in $\mathcal{C}$, then the dual statement, although true in $\mathcal{C}^{\text{op}}$, is not necessarily true in $\mathcal{C}$. For example, the statement '$\mathcal{C}$ has an initial object' may be true, while the dual '$\mathcal{C}$ has a terminal object' is not.

1.4. **The Category of Relations.** A *relation* is an ordered triple (f, S, T), where S and T are sets and f is a subset of the cartesian product $S \times T$. Given a relation (f, S, T), we say that f is a relation *from* S *to* T, and denote this fact by $f\colon S \to T$, as we do in the special case of functions. If $f\colon S \to T$ and $g\colon R \to S$ are relations, then the *composition* $fg\colon R \to T$ is defined by

$$fg = \{(r,t) \in R \times T \colon (r,s) \in g \text{ and } (s,t) \in f, \text{ for some } s \in S\}.$$

It is easy to check that composition of relations is an associative operation generalizing the usual composition of functions, and that the identity function, or *equality* relation, on a set is also the identity relation. Hence we have a category **Rel** with sets as objects and relations as morphisms, that contains **Set** as a subcategory.

For any relation $f\colon S \to T$, the *converse* f^{-1} is the relation $T \to S$ obtained from f by reversing all ordered pairs. If f happens to be a bijection, then the converse f^{-1} is just the inverse function. We call a relation a *coinjection* if it's converse is an injection, and a *cosurjection* if its converse is a surjection. For example, the relation f below is a coinjection and g is a cosurjection.



Coinjections and cosurjections comprise subcategories of **Rel** which may be identified with the dual categories $\mathbf{Inj}^{\mathrm{op}}$ and $\mathbf{Surj}^{\mathrm{op}}$ by identifying the converse of a relation f with the arrow f^{op}.

If $f\colon S \to T$ is a relation and $s \in S$ then the *degree* $\deg(s)$ of s is the cardinality of the set $\{t \in T\colon (s,t) \in f\}$, and the degree of $t \in T$ is the cardinality of the $\{s \in S\colon (s,t) \in f\}$. The *range* of f is the set $\{t \in T\colon \deg(t) \geq 1\}$ and the *corange* is the set $\{s \in S\colon \deg(s) \geq 1\}$. The *restriction* $f|R$ of f to a subset $R \subseteq S$, is the relation $R \to S$ given by

$$f|R = \{(s,t) \in f\colon s \in R\}.$$

A relation $f\colon S \to T$ is a coinjection if and only if it is a *partial bijection*, that is, if and only if the restriction of f to its corange is a bijection onto T. Coinjections $S \to T$ are also characterized by requirement that $\deg(s) \leq 1$, for all $s \in S$, and $\deg(t) = 1$, for all $t \in T$. On the other hand, cosurjections $S \to T$ are those relations with $\deg(s) \geq 1$, for all $s \in S$ and $\deg(t) = 1$, for all $t \in T$.

It is interesting to consider all the joins and intersections of the categories **Inj**, $\mathbf{Inj}^{\mathrm{op}}$, **Surj**, and $\mathbf{Surj}^{\mathrm{op}}$, that is, the sublattice of the lattice of all subcategories of **Rel** generated by these four categories. For example, the intersection of any two, or more, of them is the category **Bij** of bijections. We have already noted that the join of **Inj** and **Surj** is the category **Set** (to which we might better refer as the category of functions, to be consistent). The reader should check that $\mathbf{Inj}^{\mathrm{op}} \vee \mathbf{Inj}$, $\mathbf{Inj}^{\mathrm{op}} \vee \mathbf{Surj}$, and $\mathbf{Inj}^{\mathrm{op}} \vee \mathbf{Inj} \vee \mathbf{Surj}$ are the categories of partial injections, partial surjections and partial functions, respectively.

1.5. **Special Objects and Arrows.**

Definition 1.16. An object s in a category $\mathcal{C}$ is *initial* if, for each object $a \in \mathcal{C}$, there is precisely one arrow $s \longrightarrow a$. An object t is *terminal* if there is precisely one arrow $a \longrightarrow t$, for each a; and $z \in \mathcal{C}$ is *null* if it is both initial and terminal.

Examples 1.17. In **Set** the empty set $\emptyset$ is initial, any singleton set is terminal and there are no null objects. In $\mathbf{Set}_*$, any singleton set (with the only possible choice of base-point) is a null object. In **Grp**, any trivial group is null. If we assume that in the category **Rng**, all rings have unit element and all homomorphisms preserve units, then the ring of integers $\mathbb{Z}$ is initial. In a poset $\mathcal{P}$, an initial object is a minimum (not minimal!) element, and a terminal object is a maximum element.

Proposition 1.18. *In a category,*

(i) If s and s' are initial, then $s \cong s'$.
(ii) If t and t' are terminal, then $t \cong t'$.
(iii) If z and z' are null, then $z \cong z'$.

Proof. For (i), suppose that s and s' are initial; then there exist unique arrows $f\colon s \to s'$ and $g\colon s' \to s$. Also, there are unique $s \to s$ and $s' \to s'$, which must be identity arrows. Hence the compositions gf and fg must be equal to 1_s and $1_{s'}$, respectively; and so $s \cong s'$. Part (ii) is proved similarly, and (iii) follows directly from (i) and (ii). □

Since initial, terminal and null objects are all unique up to isomorphism (which is as unique as one could ever want in a category), we often refer to *the* initial, terminal, or null object of a category.

We now introduce classes of arrows that the generalize to an arbitrary category the notions of injective and surjective function.

Definition 1.19. An arrow m in a category $\mathcal{C}$ is *monic*, or a *monomorphism*, if the equality $mf = mg$ implies that $f = g$, for all arrows $f, g \in \mathcal{C}$. An arrow h is *epi*, or an *epimorphism*, if $fh = gh$ implies that $f = g$, for all $f, g \in \mathcal{C}$.

In other words, monics are arrows that are left cancellable, and epis are arrows that are right cancellable. We will see shortly that in **Set**, the monics are precisely the injections and the epis are precisely the surjections.

Proposition 1.20. *In a category $\mathcal{C}$,*

(i) The composition of two monics is monic.
(ii) If the composition hk is monic, then k is monic.

Proof. Suppose that we have the diagram $d \underset{g}{\overset{f}{\rightrightarrows}} a \xrightarrow{k} b \xrightarrow{h} c$.

(i) If h and k are monic and $(hk)f = (hk)g$, then $h(kf) = h(kg)$, which implies that $kf = kg$, by monotonicity of h. It then follows from the monotonicity of k that $f = g$; therefore hk is monic.
(ii) Suppose that hk is monic and that $kf = kg$. Then $h(kf) = h(kg)$, or equivalently $(hk)f = (hk)g$; and so $f = g$, by monotonicity of hk. Thus k is monic.

□

We state the corresponding result for epis without proof.

Proposition 1.21. *In a category $\mathcal{C}$,*

(i) The composition of two epis is epi.
(ii) If the composition hk is epi, then h is epi.

Corollary 1.22. *In any category, isomorphisms are both monic and epi.*

Proof. It is immediate from the definitions that identity arrows are monic and epi. Now f is an isomorphism if and only if there exists some g with fg and gf equal to identity arrows. It then follows from the second parts of Propositions 1.20 and 1.21 that f is monic and epi. □

We warn the reader that the converse is not true; in general, it is possible for an arrow to be both monic and epi, but not an isomorphism. Note that, if f is an isomorphism, then it follows from f being either monic or epi that there is a unique (two-sided) inverse g of f.

Part of the connection between the notions of monic and epi and that of injective and surjective in a concrete category is given by the following proposition.

Proposition 1.23. *In a concrete category,*

(i) Injective $\Rightarrow$ monic.
(ii) Surjective $\Rightarrow$ epi.

Proof. (i) Suppose that $m\colon A \to B$ is injective, $f, g\colon D \to A$ and $mf = mg$. Then, for any $x \in D$, we have $mf(x) = mg(x)$, which implies that $f(x) = g(x)$, by injectivity of m. Therefore $f = g$, and so m is monic. The proof of (ii) is similar. □

We emphasize that we must be working in a concrete category even to state Proposition 1.23, since the concepts of injectivity and surjectivity of an arrow are not defined in a general category.

In any given concrete category the converse of either part of Proposition 1.23 may or may not hold. In particular they both hold in **Set**, which we now prove.

Proposition 1.24. *In the category* **Set***,*

(i) Monic $\Rightarrow$ injective.
(ii) Epi $\Rightarrow$ surjective.

Hence, the monic and epi arrows in **Set** *are precisely the injections and surjections, respectively.*

Proof. (i) Let $t = \{*\}$ be a terminal object in **Set**. For any set S and $x \in S$, let $\bar{x}\colon t \to S$ be the function defined by $\bar{x}(*) = x$. The correspondence $x \leftrightarrow \bar{x}$ is thus a bijection $S \leftrightarrow \mathbf{Set}(t, S)$. Suppose that $m\colon S \to T$ is monic and that $m(x) = m(y)$, for some $x, y \in S$. Then $m\bar{x} = m\bar{y}$, which implies that $\bar{x} = \bar{y}$, by monotonicity of m. Thus $x = y$, and so m is injective.

(ii) If U is a subset of T, let $\chi_U : T \to \{0,1\}$ be the *characteristic function* of U, given by

$$\chi_U(x) = \begin{cases} 1, & \text{if } x \in U; \\ 0, & \text{otherwise.} \end{cases}$$

Now suppose that $h: S \to T$ is an epimorphism. The compositions $\chi_T h$ and $\chi_{h(S)} h$ are both equal to the constant function χ_S on S, and so $\chi_T = \chi_{h(S)}$, since h is epi. Therefore $h(S) = T$; that is, h is surjective. □

If $\mathcal{C}$ is any category that has a terminal object t, we define a *global element* of an object $a \in \mathcal{C}$ to be an arrow $t \to a$. Note that if we choose another terminal object t', then there is a canonical bijection between the sets of global elements $\mathcal{C}(t, a)$ and $\mathcal{C}(t', a)$; hence we speak of 'the' set of global elements of a. In the category **Set** we have seen that there is a canonical bijection between any set S and its set of global elements $\mathbf{Set}(t, S)$, and this fact implies that monics are injective in **Set**. The same phenomenom occurs in **Top**, implying in precisely the same manner that monics are injective there. We emphsasize that in most concrete categories, the set of global elements of an object is much smaller than the underlying set of the object. For example in $\mathbf{Set}_*$ and **Grp**, every object has only one global element.

Examples 1.25. For any group G, there is a bijection $\mathbf{Grp}(\mathbb{Z}, G) \leftrightarrow G$; for any ring R and R-module M, there is a bijection ${}_R\mathbf{Mod}(R, M) \leftrightarrow M$; and for any commutative ring K, there is a bijection $\mathbf{ComRng}(\mathbb{Z}[x], K) \leftrightarrow K$. Using these facts, the argument in part (i) of 1.24 may be modified in the obvious manner to show that monics are injective in the categories **Grp**, ${}_R\mathbf{Mod}$ and **ComRng**.

Example 1.26. The inclusion $\mathbb{Z} \to \mathbb{Q}$ is an epimorphism in **Rng** that is not a surjection.

Example 1.27. In the category **Haus** of Hausdorff spaces and continuous maps, an arrow $f : S \to T$ is determined by its values on any dense subset of S. It follows that the epimorphisms in **Haus** are precisely the continuous mappings with dense image.

Example 1.28. By giving the set $\{0, 1\}$ the indiscrete topology, we readily may modify the proof of part (ii) of Proposition 1.24 to show that the epimorphisms in **Top** are precisely the surjective continuous maps.

Definition 1.29. If $f : a \to b$ and $g : b \to a$ are arrows in a category that satisfy $gf = 1_a$ then we say that f is a *section* of g, and g is a *retraction* of f.

In other words, a section of an arrow is a right inverse and a retraction is a left inverse; hence, in particular, an arrow is both a section and a retraction if and only if it is an isomorphism.

Proposition 1.30. *In a category,*

(i) *Every section is monic.*
(ii) *Every retraction is epi.*

Proof. The proof is immediate from the second parts of Propositions 1.20 and 1.21. □

The term 'retraction' originated in topology; if there exists a retraction $T \to S$ of an inclusion map $S \hookrightarrow T$ in **Top**, then the space S is said to be a *retract* of T. For example, the unit circle S^1 is a retract of the cylinder $C = S^1 \times [0,1]$, with retraction map $C \to S^1$ given by $(x,t) \mapsto x$. On the other hand, elementary techniques from algebraic topology may be used to show, for example, that the S^1 is not a retract of the unit disk $D^1 = \{x \in \mathbb{R}^2 \colon |x| \le 1\}$.

The term 'section' also first arose in topology. Suppose that $\{S_i \colon i \in I\}$ is a pairwise-disjoint family of nonempty sets. Let $S = \cup_{i \in I} S_i$, and let $p \colon S \to I$ be the *projection* map, defined by $p(x) = i$, for all $x \in S_i$. This structure is called a *bundle*, with *base space* I and *total space* S. For each element $i \in I$, the set S_i is called the *stalk*, or *fiber*, over i. A *section* of S is a section $s \colon I \to S$ of the projection map p. If we picture each fiber S_i as sitting 'above' the point i, then a section consists of a choice of a point in S above i, for each $i \in I$. Note that a section of S is precisely an element of the cartesian product $\prod_{i \in I} S_i$. In topology, one usually considers bundles in which S and I are topological spaces and each S_i is equipped with some algebraic structure, such as that of a group or vector space.

1.6. **Exercises.**

(1) For each $k \ge 2$, let $\mathbf{Surj}_k$ be the set of all arrows $f \colon S \to T$ in **FinSet** having the property that $|f^{-1}(x)| = 1$ modulo k, for all $x \in T$. Show that $\mathbf{Surj}_k$ is a subcategory of **FinSet** that contains **FinBij**.

(2) Find a "nice" description of all the subcategories of **FinSet** that contain **Bij**.

(3) In a category $\mathcal{C}$ having null objects, a *zero arrow* $a \to b$ is defined as the composition $a \to z \to b$ where z is null. Show that the zero arrow $a \to b$ is unique for each ordered pair of objects $a, b \in \mathcal{C}$.

(4) Are all epimorphisms surjections in the category of monoids **Mon**? Prove or give a counterexample.

(5) Find a concrete category $\mathcal{C}$ and a monic arrow in $\mathcal{C}$ that is not an injection.

(6) Recall that an abelian group A is *divisible* if, for every $y \in A$ and nonzero integer n, there exists $x \in A$ such that $y = nx$. Show that monic does not imply injective in the category **Div** of divisible abelian groups and group homomorphisms. (Hint: consider the quotient map $\pi\colon \mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$.)

(7) (a) If $gf = 1_a$ in a category $\mathcal{C}$, then f is called a *split* monic and g a *split* epi. Show that all monics and epis are split in **Set**. Is this also true in **Ab**?
(b) Suppose that $gf = 1_a$ in $\mathcal{C}$, and we let $h = fg$. Show that $h^2 = hh = h$; that is, h is an *idempotent* arrow. An idempotent with such a factorization is called *split*. Show that all idempotents are split in **Set**.

(8) An *endofunction* of a set S is a map $S \to S$. Given an endofunction f of S, let G_f be the directed graph having vertex set S and edge set consisting of all arrows $x \to f(x)$, for $x \in S$. Such a graph is called a *functional digraph*. Give a simple characterization of the functional digraphs on a finite set S.

(9) The category **End** has as objects all pairs (S, g), where S is a set and g is an endomorphism of S, with arrows $(S, g) \to (T, h)$ given by functions $f\colon S \to T$ such that $hf = fg$.
(a) Verify that **End** is a category.
(b) Describe the initial, terminal, and null objects of **End**, if there are any.
(c) Describe the global elements of an object (S, g) of **End**.
(d) Characterize the monics and epis in **End**.

(10) Show that the join $\mathbf{Inj} \vee \mathbf{Inj}^{\mathrm{op}} \vee \mathbf{Surj} \vee \mathbf{Surj}^{\mathrm{op}}$ in **Rel** is equal to **Rel**.

2. Functors and Natural Transformations

2.1. **Functors.**

Definition 2.1. If $\mathcal{C}$ and $\mathcal{D}$ are categories, a *functor* $F\colon \mathcal{C} \to \mathcal{D}$ consists of a functions $\mathrm{Ob}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{C})$ and $\mathrm{Ar}(\mathcal{C}) \to \mathrm{Ar}(\mathcal{D})$, also denoted by F, such that

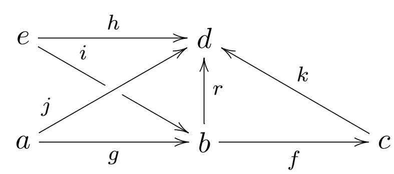
(i) $F\colon \mathcal{C}(a,b) \to \mathcal{D}(F(a),F(b))$, for all $a,b \in \mathcal{C}$.
(ii) $F(1_a) = 1_{F(a)}$, for all $a \in \mathcal{C}$.
(iii) $F(fg) = F(f)F(g)$, for all composable (f,g).

From the arrows-only perspective, a functor is just a function that preserves all of structure maps defining a category. Referring to definition (1.11), we see that (i),(ii) and (iii) above are equivalent to the following respective conditions:

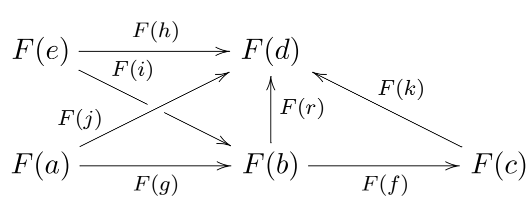
(1) $F \cdot \mathrm{dom} = \mathrm{dom} \cdot F$ and $F \cdot \mathrm{cod} = \mathrm{cod} \cdot F$;
(2) $F \cdot \mathrm{id} = \mathrm{id} \cdot F$;
(3) $F \cdot \mathrm{comp} = \mathrm{comp} \cdot (F \times F)$.

The functors whose study led to creation (discovery) of category theory, and in particular to the definition of functor itself, are the homology functors $H_n\colon \mathbf{Top} \to \mathbf{Grp}$, mentioned in the introduction. The reader may find the definitions of many variants of these functors in any textbook on algebraic topology.

We think of a functor $\mathcal{C} \to \mathcal{D}$ as giving a picture of $\mathcal{C}$ inside of the category $\mathcal{D}$. For example, a functor $F\colon \mathbf{1} \to \mathcal{D}$ simply consists of a choice of an object $F(0)$, or equivalently, an identity arrow $F(1_0)$ in $\mathcal{D}$; a functor $F\colon \mathbf{2} \to \mathcal{D}$ is just a choice of arrow $F(0) \to F(1)$; and a functor $F\colon \mathbf{3} \to \mathcal{D}$ consists of a composable pair of arrows $F(0) \to F(1) \to F(2)$ or, equivalently, a commutative triangle in $\mathcal{D}$. More generally, a functor maps all commutative diagrams to commutative diagrams. For example, if the diagram



commutes in $\mathcal{C}$ and $F\colon \mathcal{C} \to \mathcal{D}$ is a functor, then the diagram



commutes in $\mathcal{D}$. It should be noted that distinct objects and arrows in $\mathcal{C}$ may map to same objects and arrows in $\mathcal{D}$, and so some collapsing may take place when we map a diagram from $\mathcal{C}$ to $\mathcal{D}$. An extreme case of this occurs in the following example.

Example 2.2. If $\mathcal{C}$ and $\mathcal{D}$ are categories and d is an object of $\mathcal{D}$. Then there is a *constant* functor $\mathcal{C} \to \mathcal{D}$ that takes all objects in $\mathcal{C}$ to d, and all arrows to 1_d.

The following three examples provide evidence that functors are indeed the 'correct' arrows between categories.

Example 2.3. If $\mathcal{C}$ and $\mathcal{D}$ are discrete categories, then the functors from $\mathcal{C}$ to $\mathcal{D}$ correspond to the functions $\mathrm{Ob}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{D})$.

Example 2.4. If $\mathcal{C}$ and $\mathcal{D}$ are monoids (or groups), then a functor $\mathcal{C} \to \mathcal{D}$ is a monoid (or group) homomorphism.

Example 2.5. If $\mathcal{P}$ and $\mathcal{Q}$ are preorders or posets, then a functor $F\colon \mathcal{P} \to \mathcal{Q}$ is an order-preserving map; i.e., $a \leq b \Rightarrow F(a) \leq F(b)$.

It is easy to verify that the composition of two functors is a functor, and that the identity map on objects and arrows in a category $\mathcal{C}$ is a functor, called the *identity functor* on $\mathcal{C}$, and denoted by $1_{\mathcal{C}}$. Hence we may form the category **Cat** of all categories, having categories as objects and functors as arrows. A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an isomorphism in **Cat** if there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ such that the compositions FG and GF are the identity functors on $\mathcal{D}$ and $\mathcal{C}$, respectively. In the previous three examples, this notion of isomorphism corresponds to the usual one.

We often don't bother to distinguish between isomorphic categories, since they are identical up to a labelling of the objects and arrows. For example, the category of coinjections is isomorphic, but not identical, to the dual $\mathbf{Inj}^{\mathrm{op}}$ of the category of injections, but nonetheless, we usually refer to $\mathbf{Inj}^{\mathrm{op}}$ as the category of coinjections.

It turns out that the notion of isomorphism of categories is far too strict to be of much use; there are many pairs of categories that we consider as essentially the same, for which no isomorphism exists. As an example, consider the category **Set** of sets and functions, and the category **Disc** of discrete categories and functors. There is an evident functor $F\colon \mathbf{Disc} \to \mathbf{Set}$ that maps each discrete category to its object set, by throwing away the identity arrows, and takes each functor between discrete categories to the corresponding function between object sets. Now let's try to find a functor $G\colon \mathbf{Set} \to \mathbf{Disc}$ that is inverse to F. For each set S, we must choose a discrete category $G(S)$ having object set equal to S. This amounts to choosing a set $A(S)$ of identity arrows, together with a bijection $A(S) \leftrightarrow S$, and there are many possible choices for $A(S)$ and this bijection, all yielding isomorphic discrete categories. Hence, while the composition FG is the identity functor on **Set**, the composition GF cannot be the identity functor on **Disc**; for any discrete category $\mathcal{D}$, the category $GF(\mathcal{D})$ is isomorphic to, but in general not equal to, $\mathcal{D}$.

In most situations when we want to show that two categories are 'essentially the same', this is the best we can hope for: functors going both ways whose compositions with one another aren't identity functors, but map objects to isomorphic objects (in a nice, consistent manner). This should not be a surprise, given the general philosophy mentioned earlier that it is unnatural to talk about equality of objects in category theory, and better to use the isomorphism relation instead. In the next section, after we have introduced natural transformations of functors, we will give the formal definition of this type of equivalence of categories and look at a number of examples.

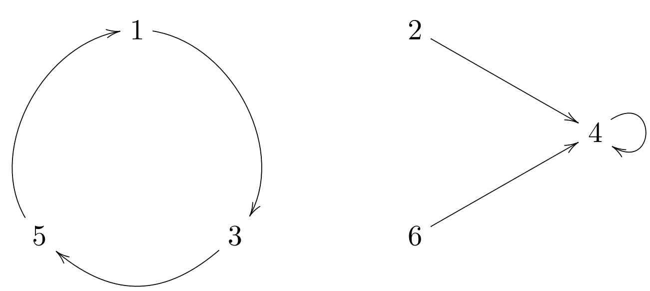
Here are some more specific examples of functors.

Example 2.6. The *power-set* functor $P\colon \mathbf{Set} \to \mathbf{Set}$ is defined by setting $P(S) = \{U\colon U \subseteq S\}$ for all sets S, and for any function $f\colon S \to T$, defining $P(f)\colon P(S) \to P(T)$ by $P(f)(U) = \{f(x)\colon x \in U\}$, for all $U \subseteq S$. In other words $P(f)(U)$ is just the image of U under f, which is usually written simply as $f(U)$.

Example 2.7. For any set S, let $\mathrm{End}(S)$ denote the set of all endofunctions (i.e., self-maps) of S. We make End into a functor $\mathbf{Bij} \to \mathbf{Set}$ by setting

$$\mathrm{End}(f)(\alpha) = f\alpha f^{-1},$$

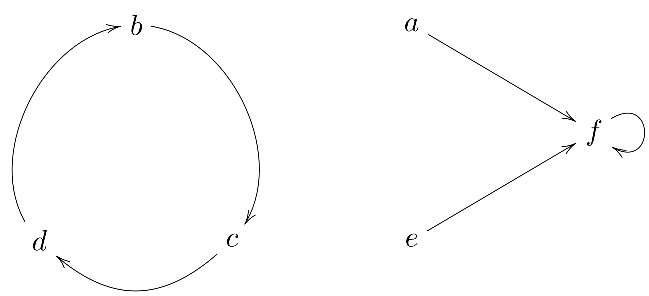
for all endofunctions α of S and bijections $f\colon S \to T$. The simplest way to understand the endomorphism $\mathrm{End}(f)(\alpha)$ is to think of it as the same as α, but where each element of S is given a new name, or label, from the set T via the bijection f. For example, suppose that α is the endofunction



of the set $S = \{1, 2, 3, 4, 5, 6\}$. If $f\colon S \to T = \{a, b, c, d, e, f\}$ is the bijection

$$1 \mapsto b, \quad 2 \mapsto a, \quad 3 \mapsto c, \quad 4 \mapsto f, \quad 5 \mapsto d, \quad 6 \mapsto e,$$

then $\mathrm{End}(f)(\alpha)$ is given by



It is a simple matter to extend the definition of End to a functor $\mathbf{Inj} \to \mathbf{Set}$, but there is no apparent way of extending it to **Surj** or to all of **Set**.

Example 2.8. Suppose that $\mathcal{G}$ is a group, that is, a category with one object $\{*\}$ in which all arrows are invertible, and write G for the set of arrows $\mathrm{Ar}(\mathcal{G})$, which is a group in the usual sense. A functor $F \colon \mathcal{G} \to \mathbf{Set}$ consists of a choice of a set $S = F(*)$, and a homomorphism (also denoted by F) from G into the group of permutations of S. In other words, F is a *permutation representation* of the group G.

A functor $F \colon \mathcal{G} \to \mathbf{Vect}_K$ consists of a choice of K-vector space $V = F(*)$ and a homomorphism from G into the group $\mathrm{G}L(V)$ of linear automorphisms of V. Hence F is a *linear representation*, or simply, a *representation*, of G. The study of such functors comprises the vast and important subject of *group representation theory*.

Example 2.9. For any abelian group M, there is a functor

$$ {}_M T \colon \mathbf{Ab} \to \mathbf{Ab}, $$

defined by

$$ {}_M T(N) = M \otimes_{\mathbb{Z}} N \quad \text{and} \quad {}_M T(f) = 1_M \otimes f, $$

for all abelian groups N and homomorphisms f. The functor ${}_M T$ is usually denoted by $M \otimes_{\mathbf{Z}} -$. Tensoring on the right by M yields another functor $\mathbf{Ab} \to \mathbf{Ab}$, denoted by T_M, or $- \otimes_{\mathbb{Z}} M$.

More generally, if R is any commutative ring and M is an R-module (recall that the notions of left and right R-module are equivalent for R commutative), then tensoring by M over R defines functors ${}_R\mathbf{Mod} \to {}_R\mathbf{Mod}$, denoted by $M \otimes_R -$ and $- \otimes_R M$.

The most general situation occurs when given two rings R and S, and M is an R-S-bimodule; in other words, M is simultaneously a left R-module and right S-module, with these structures compatible in the sense that $(rm)s = r(ms)$, for all $r \in R$, $s \in S$ and $m \in M$. Then tensoring with M on the left and right, over S and R, respectively, gives us functors

$$ M \otimes_S - : {}_S\mathbf{Mod} \to {}_R\mathbf{Mod} \quad \text{and} \quad - \otimes_R M \colon \mathbf{Mod}_R \to \mathbf{Mod}_S. $$

Recall that left R-modules and R-$\mathbb{Z}$-bimodules are the same thing, right S-modules and $\mathbb{Z}$-S-bimodules are the same thing, and that abelian groups are the same as either left or right $\mathbb{Z}$-modules. Hence, if M is a right S-module, then $M \otimes_S -$ takes values in $\mathbf{Ab}$, and if M is a left R-module, then $- \otimes_R M$ also takes values in $\mathbf{Ab}$.

Example 2.10. Suppose that a is some fixed object in a category $\mathcal{C}$. There is a functor $\mathcal{C}(a, -)\colon \mathcal{C} \to \mathbf{Set}$, called a *hom-functor*, defined by $\mathcal{C}(a, -)(b) = \mathcal{C}(a, b)$, for all objects $b \in \mathcal{C}$; and for any arrow $f\colon b \to c$,

$$\mathcal{C}(a, -)(f) = \mathcal{C}(a, f)\colon \mathcal{C}(a, b) \to \mathcal{C}(a, c)$$

is given by $\mathcal{C}(a, f)(g) = fg$, for all arrows $g\colon a \to b$. The function $\mathcal{C}(a, f)$ is usually denoted by f_*.

2.2. **Full and Faithful Functors.**

Definition 2.11. A functor $F\colon \mathcal{C} \to \mathcal{D}$ is *faithful* if $F\colon \mathcal{C}(a, b) \to \mathcal{D}(F(a), F(b))$ is injective for all objects $a, b \in \mathcal{C}$. The functor F is *full* if $F\colon \mathcal{C}(a, b) \to \mathcal{D}(F(a), F(b))$ is surjective for all $a, b \in \mathcal{C}$. A functor is an *embedding* if it is faithful and is an injective function on objects.

Example 2.12. If $\mathcal{C}$ is a subcategory of $\mathcal{D}$, then the inclusion $i\colon \mathcal{C} \to \mathcal{D}$ is a faithful functor. The subcategory $\mathcal{C}$ is full if and only if i is full.

If $\mathcal{C}$ is any concrete category, then there is an *underlying set functor* $U\colon \mathcal{C} \to \mathbf{Set}$, taking each object of $\mathcal{C}$ to its underlying set, and each arrow to the corresponding function between sets. Clearly U is a faithful functor, but usually not full.

The wary reader should have been a little unhappy with the definition of concrete category that we gave before; in particular, we did not say precisely what we meant by 'additional structure' on a set. We know that binary operations, scalar multiplications, topologies and base points are examples of it, but how do we define it in general? The existence of underlying set functors points to a way around this difficulty – we simply *define* a concrete category as one equipped with such a functor.

Definition 2.13. A *concrete* category is a category $\mathcal{C}$, together with a faithful functor $U\colon \mathcal{C} \to \mathbf{Set}$.

This definition is much more versatile, and precise, than our previous attempted definition. An object a in a concrete category $\mathcal{C}$ can be anything at all; it is the functor U that allows us to think of it as something that is 'built' on a set, namely $U(a)$. The functor U gives us a *representation* inside of $\mathbf{Set}$ of the category $\mathcal{C}$. The fact that U is faithful means that this representation contains all essential information about $\mathcal{C}$.

Underlying set functors are examples of *forgetful* functors, that is, functors defined on concrete categories that 'forget' part of the structure of an object. Other examples of forgetful functors include the functor $\mathbf{Set}_* \to \mathbf{Set}$, that forgets the base-point; the functor **Rng** $\to$ **Ab**, sending a ring to its underlying additive abelian group; and the functor **Rng** $\to$ **Semigrp**, that sends a ring to its underlying multiplicative semigroup. A more subtle example is the functor **Grp** $\to$ **Mon**, which sends each group to the corresponding monoid. You may think that this functor is not 'forgetting' anything at all, but it is really forgetting an operation, namely the unary operation of inversion. Mathematics is filled with forgetful functors which, trivial as they might seem, turn out to be of great importance.

Definition 2.14. Suppose that P is a property that arrows in any category may or may not have.

(i) A functor F *preserves* P if f having property P implies that $F(f)$ has property P.
(ii) A functor F *reflects* P if $F(f)$ having property P implies that f has property P.

Proposition 2.15.

(i) All functors preserve isomorphisms.
(ii) All full and faithful functors reflect isomorphisms.
(iii) All faithful functors reflect monics and epis.

Proof. We prove the first half of (iii) and leave the rest as an exercise. Suppose that $F\colon \mathcal{C} \to \mathcal{D}$ is a faithful functor and that $F(f)$ is monic in $\mathcal{D}$ for some arrow f in $\mathcal{C}$. By functoriality of F, if $fg = fh$ in $\mathcal{C}$, then $F(f)F(g) = F(f)F(h)$, which implies that $F(g) = F(h)$, because $F(f)$ is monic. Since F is faithful, this means that $g = h$; hence f is monic. □

Since faithful functors reflect monics and epics, which are just the injections and surjections in the category **Set**, we obtain immediately Proposition 1.23, that is, the fact that injections are monics and surjections are epis in any concrete category.

The simple fact that functors preserve isomorphisms is surprisingly powerful. In some categories it can be a relatively simple matter to show that two objects are isomorphic (by explicitly constructing an isomorphism between them, for example) while being difficult to show that two objects are not isomorphic (which means somehow showing that no isomorphism between them exists). Suppose that we have a functor F from such a category $\mathcal{C}$ to a category $\mathcal{D}$, in which it is easier to determine when objects are not isomorphic. The fact that F preserves isomorphisms gives us the ability to prove that objects a and b are not isomorphic in $\mathcal{C}$ by showing that $F(a)$ and $F(b)$ are not isomorphic in $\mathcal{D}$. The classic examples of this technique occur in algebraic topology, using the homology functors $H_n\colon$ **Top** $\to$ **Grp**.

2.3. Contravariant Functors.

Definition 2.16. A *contravariant functor* from a category $\mathcal{C}$ to a category $\mathcal{D}$ is a functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$.

We use the notation $F\colon \mathcal{C} \to \mathcal{D}$ to denote contravariant, as well as ordinary functors. Hence, the statement '$F\colon \mathcal{C} \to \mathcal{D}$ is a contravariant functor' means that F assigns an object $F(a)$ to each object a in $\mathcal{C}$, and F assigns to each arrow $f\colon a \to b$ of $\mathcal{C}$ an arrow $F(f)\colon F(b) \to F(a)$ of $\mathcal{D}$, such that $F(fg) = F(g)F(f)$, for all composable pairs of arrows (f, g) in $\mathcal{C}$. When we need to emphasize that a particular functor is not contravariant, we will call it a *covariant* functor.

Now every functor $F\colon \mathcal{C} \to \mathcal{D}$ corresponds to a functor $F^{\mathrm{op}}\colon \mathcal{C}^{\mathrm{op}} \to \mathcal{D}^{\mathrm{op}}$ having the same values as F on objects and satisfying $F^{\mathrm{op}}(f^{\mathrm{op}}) = (F(f))^{\mathrm{op}}$ on arrows. (The reader should verify that F^{op} is indeed a functor.) The correspondence $F \mapsto F^{\mathrm{op}}$ thus defines a bijection of arrow sets

$$\mathbf{Cat}(\mathcal{C}, \mathcal{D}) \to \mathbf{Cat}(\mathcal{C}^{\mathrm{op}}, \mathcal{D}^{\mathrm{op}}).$$

In particular, we may equally well have defined a contravariant functor $F\colon \mathcal{C} \to \mathcal{D}$ as a functor $\mathcal{C} \to \mathcal{D}^{\mathrm{op}}$.

Example 2.17. The *contravariant power-set functor* $\bar{P}\colon \mathbf{Set} \to \mathbf{Set}$ is defined by letting $\bar{P}(S)$ be the set of all subsets of S, for any set S, and for any function $f\colon S \to T$, letting $\bar{P}(f)\colon \bar{P}(T) \to \bar{P}(S)$ be the preimage map, sending $U \subseteq T$ to the set $f^{-1}(U) = \{s \in S\colon f(s) \in U\}$.

Example 2.18. Given a topological space T, let $\mathcal{O}_T$ be the partially ordered set having as objects all open sets in T, ordered by inclusion (so that there is a single arrow $U \to V$, whenever $U \subseteq V$ are open in T). For every open set U in T, let $C(U)$ be the set of all real-valued continuous functions defined on U. Whenever $U \subseteq V$ is an inclusion of open sets in T, restriction to V defines a function $C(U) \to C(V)$, and it is readily verified that C is thus a contravariant functor $\mathcal{O}_T \to \mathbf{Set}$. In fact, the set of continuous functions $C(U)$ has the structure of an $\mathbb{R}$-algebra (with pointwise operations) and the restriction maps are $\mathbb{R}$-algebra maps; thus C is a functor from $\mathcal{O}_T$ to the category $\mathbf{Alg}_{\mathbb{R}}$ of $\mathbb{R}$-algebras. The functor C is called the *presheaf* of continuous functions on T.

Presheaves of continuous functions are among the classical examples of contravariant functors. Borrowing from their example, category theorists refer to any contravariant set-valued functor $\mathcal{C}^{\mathrm{op}} \to \mathbf{Set}$ as a *presheaf on* $\mathcal{C}$.

Example 2.19. Corresponding to any object a in a category $\mathcal{C}$ there is a functor $\mathcal{C}(-, a)\colon \mathcal{C}^{\mathrm{op}} \to \mathbf{Set}$, called a (*contravariant*) *hom-functor* on $\mathcal{C}$, defined by $\mathcal{C}(-, a)(b) = \mathcal{C}(b, a)$, for all objects $b \in \mathcal{C}$, and for any arrow $f\colon b \to c$,

$$\mathcal{C}(-, a)(f) = \mathcal{C}(f, a)\colon \mathcal{C}(c, a) \to \mathcal{C}(b, a)$$

given by $\mathcal{C}(a, f)(g) = gf$, for all arrows $g\colon c \to a$. The function $\mathcal{C}(f, a)$ is usually denoted by f^*.

2.4. **Products of Categories.**

Definition 2.20. The *product* of two categories $\mathcal{C}$ and $\mathcal{D}$ is the category $\mathcal{C} \times \mathcal{D}$, defined by $\mathrm{Ob}(\mathcal{C} \times \mathcal{D}) = \mathrm{Ob}(\mathcal{C}) \times \mathrm{Ob}(\mathcal{D})$ and $\mathrm{Ar}(\mathcal{C} \times \mathcal{D}) = \mathrm{Ar}(\mathcal{C}) \times \mathrm{Ar}(\mathcal{D})$, with

$$(\mathcal{C} \times \mathcal{D})((a, a'), (b, b')) = \mathcal{C}(a, b) \times \mathcal{D}(a', b'),$$

for all objects $a, a' \in \mathrm{Ob}(\mathcal{C})$ and $b, b' \in \mathrm{Ob}(\mathcal{D})$, and composition given by $(f, g)(f', g') = (ff', gg')$, for all arrows $f, f' \in \mathrm{Ar}(\mathcal{C})$ and $g, g' \in \mathrm{Ar}(\mathcal{D})$.

Hence, in the product category $\mathcal{C} \times \mathcal{D}$, an arrow $(a, a') \to (b, b')$ is just a pair of arrows (f, g), where $f\colon a \to b$ and $g\colon a' \to b'$.

If $F_1\colon \mathcal{C}_1 \to \mathcal{D}_1$ and $F_2\colon \mathcal{C}_2 \to \mathcal{D}_2$ are functors, then the *product* functor $(F_1 \times F_2)\colon \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{D}_1 \times \mathcal{D}_2$ is defined on objects by $(F_1 \times F_2)(a_1, a_2) = (F_1(a_1), F_2(a_2))$, and on arrows by $(F_1 \times F_2)(f_1, f_2) = (F_1(f_1), F_2(f_2))$.

Example 2.21. For any category $\mathcal{C}$, there is a *diagonal* functor $\Delta\colon \mathcal{C} \to \mathcal{C} \times \mathcal{C}$, defined on objects by $\Delta(a) = (a, a)$ and on arrows by $\Delta(f) = (f, f)$. For any $\mathcal{C}_1$ and $\mathcal{C}_2$ there is a *twist* functor Tw, defined on objects by $\mathrm{Tw}(a_1, a_2) = (a_2, a_1)$, and on arrows by $\mathrm{Tw}(f_1, f_2) = (f_2, f_1)$; there are also *projection* functors $P_1\colon \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{C}_1$ and $P_2\colon \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{C}_2$, defined by $P_i(a_1, a_2) = a_i$ and $P_i(f_1, f_2) = f_i$, for $i = 1, 2$.

The following proposition is easy to prove.

Proposition 2.22. *For all categories $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{D}$ and any pair of functors $F_1\colon \mathcal{D} \to \mathcal{C}_1$ and $F_2\colon \mathcal{D} \to \mathcal{C}_2$, there exists a unique functor*

$$G\colon \mathcal{D} \to \mathcal{C}_1 \times \mathcal{C}_2$$

such that $F_1 = P_1 G$ and $F_2 = P_2 G$.

Observe that Proposition 2.22 can be interpreted as the statement that the map

$$\mathbf{Cat}(\mathcal{D}, \mathcal{C}_1 \times \mathcal{C}_2) \longrightarrow \mathbf{Cat}(\mathcal{D}, \mathcal{C}_1) \times \mathbf{Cat}(\mathcal{D}, \mathcal{C}_2),$$

defined by $G \mapsto (P_1 G, P_2 G)$, is a bijection.

Example 2.23. Cartesian product Prod is a functor from the product category $\mathbf{Set} \times \mathbf{Set}$ to $\mathbf{Set}$. We write $S \times T$, instead of $\mathrm{Prod}(S, T)$, for all sets S and T; and if $(f, g)\colon (S, T) \to (S', T')$ is an arrow in $\mathbf{Set} \times \mathbf{Set}$, we write $f \times g$ for the function

$$\mathrm{Prod}((f, g))\colon S \times T \to S' \times T',$$

which is defined by $(f \times g)((x, y)) = (f(x), g(y))$, for all $x \in S$ and $y \in T$.

Functors defined on a product of categories $\mathcal{C}_1 \times \mathcal{C}_2$ are often referred to as *bifunctors*, or *functors of two variables*. If we are given a bifunctor $F\colon \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{D}$, then fixing an object $a \in \mathcal{C}_1$ defines a functor $F(a, -)\colon \mathcal{C}_2 \to \mathcal{D}$ determined by $c \mapsto F(a, c)$ and $f \mapsto F(1_a, f)$, for objects a and arrows f in $\mathcal{C}_2$. Similarly, fixing an object b in $\mathcal{C}_2$ defines a functor $F(-, b)\colon \mathcal{C}_1 \to \mathcal{D}$.

Example 2.24. For any category $\mathcal{C}$, there is a hom-(bi)functor

$$\mathcal{C}(-,-)\colon \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathbf{Set}$$

that satisfies $\mathcal{C}(-,-)(a,b) = \mathcal{C}(a,b)$, for all objects a and b in $\mathcal{C}$. An arrow $(a,b) \to (a',b')$ in $\mathcal{C}^{\mathrm{op}} \times \mathcal{C}$ consists of an ordered pair (f,g), where $f\colon a' \to a$ and $g\colon b \to b'$ in $\mathcal{C}$. The function

$$\mathcal{C}(-,-)(f,g) = \mathcal{C}(f,g)\colon \mathcal{C}(a,b) \to \mathcal{C}(a',b')$$

is given by $\mathcal{C}(f,g)(f,g)(h) = ghf$, for all $h \in \mathcal{C}(a,b)$.

By fixing an object a on the left, as above, we obtain the covariant hom-functor $\mathcal{C}(a,-)\colon \mathcal{C} \to \mathbf{Set}$ described in Example 2.10. In that example, we used the notation $\mathcal{C}(a,f)$ for the value of the functor $\mathcal{C}(a,-)$ on an arrow f, while according to discussion above, this value is denoted by $\mathcal{C}(1_a,f)$. This notational inconsistency is remedied if we adopt the arrows-only point of view, for then we have $a = 1_a$.

We may also obtain the contravariant hom-functors, described in Example 2.19, by fixing objects on the right in the functor $\mathcal{C}(-,-)$.

In many categories, the sets of arrows come equipped with additional structure, and hence the hom-functors take values in categories other than **Set**. For example:

(1) The functor $\mathbf{Ab}(-,-)$ takes values in **Ab**.
(2) For any ring R, the functor ${}_R\mathbf{Mod}(-,-)$ takes values in **Ab**. If R is commutative, then ${}_R\mathbf{Mod}(-,-)$ takes values in ${}_R\mathbf{Mod}$.
(3) Let **Pos** denote the category of all partially ordered sets (posets) and order-preserving maps. If P and Q are posets, then the set $\mathbf{Pos}(P,Q)$ is partially ordered by setting $f \leq g$ whenever $f(x) \leq g(x)$, for all $x \in P$. Hence the hom-functor $\mathbf{Pos}(-,-)$ maps **Pos** to itself.

For each category $\mathcal{C}$ in the above list of examples, not only does the hom-bifunctor takes values in some category $\mathcal{D}$ other than **Set**, but for all objects a,b,c, the composition $\mathcal{C}(a,b) \times \mathcal{C}(b,c) \to \mathcal{C}(a,c)$, is an arrow in $\mathcal{D}$ rather than a function between sets. In this situation we refer to $\mathcal{C}$ as a $\mathcal{D}$-*enriched* category. Strictly speaking, enriched categories are not necessarily categories, but include ordinary categories as a special case, because in general the arrow-objects $\mathcal{C}(a,b)$ are not assumed to be sets. It takes a little work to give a precise definition of enriched category; in particular, we no longer may refer to elements of $\mathcal{C}(a,b)$ and so we must take care to formulate the definition of identity arrows in an arrows-only fashion.

Example 2.25. Given rings R and S, we denote by ${}_R\mathbf{Mod}_S$ the category of all R-S-bimodules and R-S-linear maps. It is easy to check that tensoring over S defines a functor $\otimes_S\colon {}_R\mathbf{Mod}_S \times {}_S\mathbf{Mod}_T \to {}_R\mathbf{Mod}_T$, for all rings R, S and T. We write $M \otimes_S N$ instead of $\otimes_S(M,N)$ for modules S and T, and write $f \otimes g$ instead of $\otimes_S(f,g)$, for maps f and g. All of the special cases of Example 2.9 can be obtained from the functor $\otimes_S$ by fixing variables and choosing R, S and T appropriately.

We now introduce what may seem like a completely bizarre notational convention and write $M\colon R \to S$, whenever M is an R-S-bimodule. Then if we are given bimodules $M\colon R \to S$ and $N\colon S \to T$, we obtain a bimodule $M \otimes_S N\colon R \to T$ by tensoring over S. This suggests that if we regard bimodules as arrows between rings, then tensor product behaves alot like composition of arrows. In fact, we get something very close to a category by taking rings as objects, bimodules as arrows, and tensor product as the composition operation. This 'category' is enriched in the sense that, rather than arrow-*sets*, we have the arrow-*categories* ${}_R\mathbf{Mod}_S$, for all rings R and S; and the composition law is a functor, rather than just a function. The reason why the word 'category' is in quotes in the previous sentence is that this composition law is not strictly associative, and there are no (strict) identity arrows. However, the canonical isomorphisms $(M \otimes_S N) \otimes_T P \cong M \otimes_S (N \otimes_T P)$ and $M \otimes_S S \cong M$ and $R \otimes_R M \cong M$ mean that associativity of composition and the identity axiom hold up to isomorphism; these isomorphisms are in fact natural isomorphisms of functors.

3. Natural Transformations

3.1. **Definition and Some Examples.** When we defined the category **Cat**, having categories as objects and functors as arrows, we moved up to a new level of abstraction, since the objects of this category themselves consist of objects and arrows. We now proceed to move up one more level, by defining a notion of arrow between functors, which will thus make the set of functors $\mathbf{Cat}(\mathcal{C}, \mathcal{D})$ the object set of a category for any categories $\mathcal{C}$ and $\mathcal{D}$.

Definition 3.1. Suppose that $F \colon \mathcal{C} \to \mathcal{D}$ and $G \colon \mathcal{C} \to \mathcal{D}$ are functors. A *natural transformation* τ from F to G is comprised of a set of arrows $\{\tau_a \colon F(a) \to G(a) \colon a \in \mathrm{Ob}(\mathcal{C})\}$ in $\mathcal{D}$, indexed by objects in $\mathcal{C}$, such that for any $f \colon a \to b$ in $\mathcal{C}$, the diagram

$$\begin{array}{ccc} F(a) & \xrightarrow{\tau_a} & G(a) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(b) & \xrightarrow{\tau_b} & G(b) \end{array} \tag{3.2}$$

commutes. The individual arrows τ_a are the *components* of τ, and the commutativity of the square (3.2) is called *naturality.* The natural transformation τ is a *natural isomorphism* if each component τ_a is an isomorphism in $\mathcal{D}$; in this case the functors F and G are called *isomorphic* and we write $F \cong G$.

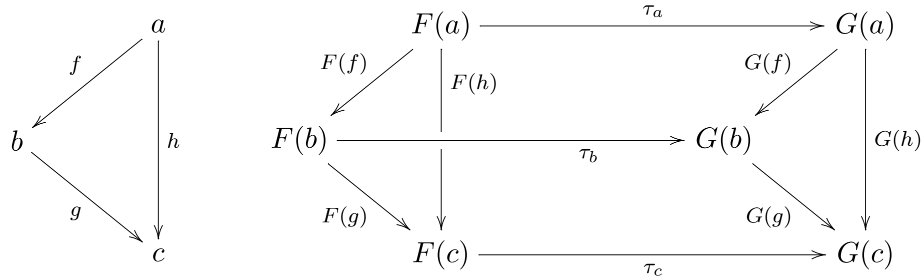
We write $\tau \colon F \Rightarrow G$ to indicate that τ is a natural transformation from F to G, and we represent τ by the diagram

$$\mathcal{C} \underset{G}{\overset{F}{\rightrightarrows}} \mathcal{D}, \quad (\Downarrow \tau)$$

if F and G have domain $\mathcal{C}$ and codomain $\mathcal{D}$.

If think of functors $F, G \colon \mathcal{C} \to \mathcal{D}$ as giving pictures of $\mathcal{C}$ inside of $\mathcal{D}$, then we may regard a natural transformation $\tau \colon F \Rightarrow G$ as a process that takes us from one of these pictures to the other using the arrows of $\mathcal{D}$. Specifically, F and G map any commutative diagram D in $\mathcal{C}$ to commutative diagrams $F(D)$ and $G(D)$ in $\mathcal{D}$, and τ provides an arrow $\tau_a \colon F(a) \to G(a)$, from each vertex of $F(D)$ to the corresponding vertex of $G(D)$ in such a way that the union of the diagrams $F(D)$ and $G(D)$, together with all of these arrows, is itself a commutative diagram. We think of this big commutative diagram as a cylinder, having the diagram $F(D)$ at one end and the diagram $G(D)$ at the other, that transforms $F(D)$ into $G(D)$, using the arrows of $\mathcal{D}$. For example, the commutative triangle in $\mathcal{C}$, below left, corresponds to the cylinder in $\mathcal{D}$

on the right.



Example 3.3. Consider the functors End, $P\colon \mathbf{Bij} \to \mathbf{Set}$, where End is the endomorphism functor defined in Example 2.7, and P is the power-set functor, defined in Example 2.6, restricted to the category of bijections **Bij**. There is a natural transformation $\varphi\colon \text{End} \Rightarrow P$, with $\varphi_S\colon \text{End}(S) \to P(S)$ defined for all sets S by

$$\varphi_S(h) = \{s \in S\colon x \text{ is a fixed point of } h\},$$

for all $h \in \text{End}(S)$. To verify naturality, we must check that

$$\varphi_T(\text{End}(f)(h)) = P(f)(\varphi_S(h)),$$

for all bijections $f\colon S \to T$ and endomorphisms h of S. But this is equivalent to the equality of sets

$$\{\text{fixed points of } fhf^{-1}\} = f(\{\text{fixed points of } h\}),$$

which is apparent.

The naturality of $\varphi\colon \text{End} \Rightarrow P$ means that whenever we choose an endomorphism h of S, relabel it via a bijection $f\colon S \to T$, and then take it's set of fixed points, we obtain the same subset of T that we get from taking the set of fixed points of h first, then relabelling. This means that the process of passing from an endomorphism to its set of fixed points does not depend on any particular labelling of the underlying set; whether we change labels before or afterwards makes no difference. Thus we may regard φ as a transformation between the endomorphism and power-set *constructions*, rather than just as an assortment of mappings $\text{End}(S) \to P(S)$ for all sets S.

Example 3.4. Let Π and Eqrel be the functors $\mathbf{Bij} \to \mathbf{Set}$ that take a set U to the set of all partitions of U, and to the set of all equivalence relations on U, respectively, and have the obvious definition on arrows. (Recall that a *partition* of U is a set π of nonempty, pairwise disjoint subsets of U, called *blocks*, whose union is U.) For every set S, define $\tau_S\colon \Pi(S) \to \text{Eqrel}(S)$ by letting $\tau_S(\pi)$ be the relation

$$\{(s,t) \in S \times S\colon\ s,t \in B, \text{ for some block } B \in \pi\}$$

on S, for all partitions π of S. It is readily verified that the maps τ_S are the components of a natural isomorphism $\tau\colon \Pi \Rightarrow \text{Eqrel}$, whose inverse is defined by letting $\tau_S^{-1}(R)$ be the set of equivalence classes of R, for any equivalence relation R on a set S.

Example 3.5. The definition given in Example 2.23 for the cartesian product functor $\text{Prod}\colon \mathbf{Set} \times \mathbf{Set} \to \mathbf{Set}$ is in fact ambiguous. Even though we may all agree that the set $S \times T$ consists of all ordered pairs (s,t) with $s \in S$ and $t \in T$, we may not agree on what, precisely, is meant by an ordered pair. One common definition is to let (s,t) be the set $\{s, \{s,t\}\}$. Another approach, that is readily generalizable to arbitrary families of sets, is to define an ordered pair in $S \times T$ as a function $\alpha\colon \{1,2\} \to S \cup T$ that satisfies $\alpha(1) \in S$ and $\alpha(2) \in T$. The functions $\tau_{S,T}$ defined by $\alpha \mapsto \{\alpha(1), \{\alpha(1), \alpha(2)\}\}$ are the components of a natural isomorphism from the second to the first versions of the cartesian product functor. Of course the index set $\{1,2\}$ in the second definition may be replaced with any other totally ordered two-element set, resulting in yet another isomorphic, but distinct, functor.

We see in the previous example a recurring theme in category theory; things we wish to define are often only meaningfully specified up to isomorphism. The reader may ask why we don't simply choose a particular version of the functor Prod to work with. For some purposes it may be useful to do so, but in Section 93, we shall meet a general definition of product of objects in a category (which may or may not exist in any particular category), of which cartesian product of sets is a special case, and by it's very nature, this definition only determines a product up to (unique) isomorphism.

Example 3.6. For any field K, there is a contravariant functor $(-)^*$ from the category $\mathbf{Vect}$ of K-vector spaces to itself, taking a vector space V to its dual V^*, and a linear map $f\colon V \to W$ to its *transpose* (or *dual*) $f^*\colon W^* \to V^*$, given by $f^*(h) = hf$, for all $h \in W^*$. Note that $(-)^*$ is just the contravariant hom-functor $\mathbf{Vect}(-, K)$. Composing $(-)^*$ with itself, we obtain the (*co*variant) functor $(-)^{**}$ that takes each space V to its double dual V^{**} and each linear map $f\colon V \to W$ to the linear map $f^{**}\colon V^{**} \to W^{**}$, given by $f^{**}(\alpha) = \alpha f^*$, for all $\alpha \in V^{**}$.

For each vector space V, there is an injective linear map $\tau_V\colon V \to V^{**}$, given by $x \mapsto \tilde{x}$, where $\tilde{x} \in V^{**}$ is defined by setting $\tilde{x}(f) = f(x)$, for each $x \in V$ and $f \in V^*$. The map τ_V is called the canonical inclusion of V into its double dual. We claim that the maps τ_V are the components of a natural transformation $\tau\colon 1_{\mathbf{Vect}} \Rightarrow (-)^{**}$, which justifies the use of the word 'canonical' in the previous sentence. It is worth verifying naturality carefully here, in order to get some practice untangling the definitions of f^* and f^{**};

thus we must check that the square

$$\begin{array}{ccc} V & \xrightarrow{\tau_V} & V^{**} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f^{**}} \\ W & \xrightarrow{\tau_W} & W^{**} \end{array}$$

commutes, for all linear maps $f\colon V \to W$. Now for any $x \in V$ and $h \in W^*$, we have $(\widetilde{f(x)})(h) = hf(x)$ and

$$\begin{aligned} f^{**}(\tilde{x})(h) &= \tilde{x}(f^*(h)) \\ &= \tilde{x}(hf) \\ &= hf(x), \end{aligned}$$

and hence the square commutes.

If we restrict attention to the category **fVect** of finite-dimensional vector spaces over K, then τ is a natural isomorphism. The usual proof of this goes as follows: choose a basis $\{e_1, \dots, e_n\}$ of V, and define a map $\gamma_v\colon V \to V^*$ by $e_i \mapsto e_i^*$, where $e_i^* \in V^*$ is determined by

$$e_i^*(e_j) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise,} \end{cases}$$

for $1 \le i, j \le n$. It is easy to see that γ_V is an isomorphism, but it is not natural because the definition of each map γ_V depends on a choice of basis for V. However, if we choose a basis $\{e_1, \dots, e_n\}$ for a vector space V, and map V into the double dual V^{**}, by $e_i \mapsto e_i^{**} = (e_i^*)^*$, then we obtain the map $\tau_V\colon V \to V^{**}$, which is thus an isomorphism. Note that, while the map $e_i \mapsto e_i^*$ depends on the choice of basis, the map $e_i \mapsto e_i^{**}$ does not.

Example 3.7. Suppose that $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{D}$ are categories, and that we are given a functor $F\colon \mathcal{C}_1 \times \mathcal{C}_2 \to \mathcal{D}$. Recall that, for each object $a \in \mathcal{C}_1$, the functor $F(a, -)\colon \mathcal{C}_2 \to \mathcal{D}$ is defined by $F(a,-)(x) = F(a,x)$ and $F(a,-)(f) = F(1_a, f)$, for all objects x and arrows f in $\mathcal{C}_2$. Each arrow $g\colon a \to b$ in $\mathcal{C}_1$ induces a natural transformation

$$F(g,-)\colon F(a,-) \Rightarrow F(b,-),$$

having components

$$F(g,-)_x = F(g, 1_x)\colon F(a,x) \to F(b,x),$$

for each $x \in \mathcal{C}_2$. The naturality of $F(g, -)$ is equivalent to the commutativity of the square

$$\begin{array}{ccc} F(a,x) & \xrightarrow{F(g,1_x)} & F(b,x) \\ {\scriptstyle F(1_a,f)}\downarrow & & \downarrow{\scriptstyle F(1_b,f)} \\ F(a,y) & \xrightarrow[F(g,1_y)]{} & F(b,y), \end{array}$$

for each arrow $f\colon x \to y$ in $\mathcal{C}_2$, which is clear. Similarly, each arrow $f\colon x \to y$ in $\mathcal{C}_2$ gives rise to a natural transformation

$$F(-\,, f)\colon F(-\,, x) \Rightarrow F(-\,, y),$$

of functors from $\mathcal{C}_1$ to $\mathcal{D}$ whose naturality corresponds to the same commutative square.

In the special case of the hom-functor $\mathcal{C}(-\,,-)\colon \mathcal{C} \times \mathcal{C} \to \mathcal{C}$, for each arrow $f\colon a \to b$, we write f^* for the natural transformation

$$\mathcal{C}(f, -)\colon \mathcal{C}(b, -) \Rightarrow \mathcal{C}(a, -),$$

as well as for each component $\mathcal{C}(f, x)\colon \mathcal{C}(b, x) \to \mathcal{C}(a, x)$, which is given by composing with f on the right; and we write f_* for the natural transformation

$$\mathcal{C}(-\,, f)\colon \mathcal{C}(-\,, a) \Rightarrow \mathcal{C}(-\,, b),$$

as well as for each component $\mathcal{C}(x, f)\colon \mathcal{C}(x, a) \to \mathcal{C}(x, b)$, which is given by composing with f on the left.

It is useful to compare the definition of natural transformation with that of homotopy between continuous functions. Recall that if U and V are topological spaces and $f, g\colon U \to V$ are continuous functions, then a *homotopy* from f to g is a continuous map $H\colon U \times [0,1] \to V$ satisfying

$$H(x, 0) = f(x) \qquad \text{and} \qquad H(x, 1) = g(x),$$

for all $x \in U$. Hence, if we define $H_t\colon U \to V$, for each $\in [0,1]$ to be the map $x \mapsto H(x,t)$, then $\{H_t\}$ is a continuously parameterized family of continous maps from U to V, with $H_0 = f$ and $H_1 = g$. If there is a homotopy from f to g, we say that f and g are *homotopic*, and write $f \cong g$. It is easy to check that being homotopic is an equivalence relation on the set of all continuous maps from U to V.

Now suppose $f, g\colon U \to V$ are continuous, and H is a homotopy from f to g. For each element x of U, the map $[0,1] \to V$, given by $t \mapsto H_t(x)$, is a path in V from $f(x)$ to $g(x)$. Hence we have a very close analogy: topological spaces correspond to categories with points as objects and paths as arrows, continuous maps between spaces correspond to functors, and homotopies between maps correspond to natural transformations. It is an interesting exercise to try make these correspondences more precise, so that they are no longer merely analogies.

3.2. Some Natural Transformations Involving the Cartesian Product Functor.

Example 3.8 (**Unit**)**.** Choose a singleton set $T = \{*\}$, and let

$$L_T \colon \mathbf{Set} \to \mathbf{Set}$$

be the functor defined on objects by $L_T(S) = T \times S$ and on arrows by $L_T(f) = 1_T \times f$. Then there is a natural isomorphism τ from the identity functor $1_{\mathbf{Set}}$ to L_T, with components defined by $\tau_S(s) = (*, s)$, for all $s \in S$.

Example 3.9 (**Commutativity**)**.** The *twist* maps

$$\tau_{S,T} \colon S \times T \to T \times S,$$

defined for all sets S and T by $(s,t) \mapsto (t,s)$, are the components of a natural isomorphism from the cartesian product functor

$$\mathrm{Prod} \colon \mathbf{Set} \times \mathbf{Set} \to \mathbf{Set}$$

to the composition $\mathrm{Prod} \cdot \mathrm{Tw}$, where $\mathrm{Tw} \colon \mathbf{Set} \times \mathbf{Set} \to \mathbf{Set} \times \mathbf{Set}$ is the twist functor, which was defined in general in Example 2.21.

Example 3.10 (**Associativity**)**.** The *associators*

$$\alpha_{S,T,U} \colon ((S \times T) \times U) \to (S \times (T \times U)),$$

given for all sets S, T and U by $((s,t),u) \mapsto (s,(t,u))$, are the components of a natural isomorphism

$$\mathrm{Prod} \cdot (\mathrm{Prod} \times 1_{\mathbf{Set}}) \Rightarrow \mathrm{Prod} \cdot (1_{\mathbf{Set}} \times \mathrm{Prod})$$

of functors $\mathbf{Set} \times \mathbf{Set} \times \mathbf{Set} \to \mathbf{Set}$.

It is interesting to compare the category **Set** with a commutative monoid. In a commutative monoid, the associativity and commutativity of the binary operation, and the unit property of the identity element, are specified by equations among elements. In the category **Set**, with binary operation on objects given by cartesian product, these properties are given by natural isomorphisms, and the identity element itself is defined only up to isomorphism. This is as to be expected, since when working in a category, isomorphism is as close to equality as we can reasonably expect to get. We thus view **Set**, equipped with cartesian product, as a kind of *categorified* commutative monoid, with defining equations replaced by appropriate isomorphisms. Such categories are called *symmetric monoidal* categories. We will take a close look at them and the more general *monoidal* categories in Section n^2.

3.3. **Equivalence of Categories.** Recall that a functor $F\colon \mathcal{C} \to \mathcal{D}$ is an isomorphism if there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ such that

$$FG = 1_{\mathcal{D}} \qquad \text{and} \qquad GF = 1_{\mathcal{C}}. \tag{3.11}$$

As we mentioned in Section 2.1, Conditions (3.11) seldom hold in practice; usually the best we can hope for is that either of these compositions applied to an object yields an isomorphic, not equal, object. Indeed, given our general philosophy, it would be unnatural to require equality here. Now that we have natural transformations at our disposal, we can define the desired notion of equivalence concisely.

Definition 3.12. A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an *equivalence* if there exists a functor $G\colon \mathcal{D} \to \mathcal{D}$ and natural isomorphisms $FG \cong 1_{\mathcal{D}}$ and $GF \cong 1_{\mathcal{C}}$. If there exists an equivalence from $\mathcal{C}$ to $\mathcal{D}$, then $\mathcal{C}$ and $\mathcal{D}$ are called *equivalent* categories.

Example 3.13. The category **Disc** of discrete categories is equivalent, but not isomorphic, to the category **Set** of sets.

Example 3.14. The category of monoids and monoid homomorphisms is equivalent to the category of one-object categories and functors. If we adopt the arrows-only definition of category, then these categories are isomorphic. Similar comments hold for **Grp**.

Definition 3.15. A *skeleton* of a category $\mathcal{C}$ is a full subcategory $\mathcal{K}$ such that each object of $\mathcal{C}$ is isomorphic to precisely one object in $\mathcal{K}$.

It is not difficult to see that any category is equivalent to any of its skeletons, and two categories are equivalent if and only if they have isomorphic skeletons.

Proposition 3.16. *For any functor $F\colon \mathcal{C} \to \mathcal{D}$, the following are equivalent:*

(i) F is an equivalence;
(ii) F is full and faithful, and each object $d \in \mathcal{D}$ is isomorphic to $F(c)$, for some object $c \in \mathcal{C}$.

Proof. $(i) \Rightarrow (ii)$ is easy. $(ii) \Rightarrow (i)$ is also easy, but relies on the axiom of choice. $\square$

3.4. **Categories of Functors.** Suppose that $\mathcal{C}$ and $\mathcal{D}$ are categories and that F G and H are functors $\mathcal{C} \to \mathcal{D}$. Given natural transformations $\sigma\colon F \Rightarrow G$ and $\tau\colon G \Rightarrow H$, we define the composition $\tau\sigma$ to be the natural transformation $F \Rightarrow H$ having components $(\tau\sigma)_a = \tau_a\sigma_a$, for all objects

$a \in \mathcal{C}$. To verify that $\tau\sigma$ is indeed natural, consider the diagram

$$\begin{array}{ccccc} F(a) & \xrightarrow{\sigma_a} & G(a) & \xrightarrow{\tau_a} & H(a) \\ {\scriptstyle F(f)}\downarrow & & {\scriptstyle G(f)}\downarrow & & \downarrow{\scriptstyle H(f)} \\ F(b) & \xrightarrow[\sigma_b]{} & G(b) & \xrightarrow[\tau_b]{} & H(b), \end{array}$$

for some arrow $f: a \to b$ in $\mathcal{C}$. Naturality of σ and τ means that the left and right squares commute. It follows that composing the labels along each of the three paths from $F(a)$ to $H(b)$ yields the same arrow in $\mathcal{D}$, and therefore the whole diagram commutes. In particular, the outside rectangle commutes, which is the naturality condition for $\tau\sigma$.

Associativity of composition of natural transformations follows immediately from the associativity of composition of arrows in $\mathcal{D}$. For every functor $F: \mathcal{C} \to \mathcal{D}$, the natural transformation $1_F: F \Rightarrow F$, defined by $(1_F)_a = 1_{F(a)}$, for all objects $a \in \mathcal{C}$, is an identity for composition. Hence we have a category with all functors from $\mathcal{C}$ to $\mathcal{D}$ as objects and natural transformations between functors as arrows. We write either $\mathbf{Cat}(\mathcal{C}, \mathcal{D})$ or $\mathcal{D}^{\mathcal{C}}$ for this category. For functors $F, G \in \mathbf{Cat}(\mathcal{C}, \mathcal{D})$, we denote the set of natural transformations from F to G by $\mathrm{Nat}(F, G)$, instead of the more cumbersome $\mathbf{Cat}(\mathcal{C}, \mathcal{D})(F, G)$, or more cryptic $\mathcal{D}^{\mathcal{C}}(F, G)$.

Example 3.17. For any category $\mathcal{C}$, a functor $F: \mathbf{1} \to \mathcal{C}$ is determined by choosing a single object $F(0)$ in $\mathcal{C}$. If F and G are such functors, then a natural transformation $\tau: F \Rightarrow G$ consists of a single arrow $\tau_0: F(0) \to G(0)$ in $\mathcal{C}$, and composition of natural transformations corresponds to composition of arrows. Hence the category $\mathcal{C}^{\mathbf{1}}$ is (canonically) isomorphic to $\mathcal{C}$.

Example 3.18. A functor $F: \mathbf{2} \to \mathcal{C}$ is determined by choosing an arrow $f: F(0) \to F(1)$ in $\mathcal{C}$. Given functors $F, G: \mathbf{2} \to \mathcal{C}$, with G corresponding to the arrow g, a natural transformation $\tau: F \Rightarrow G$ consists of a pair of arrows τ_0 and τ_1 in $\mathcal{C}$, making the square

$$\begin{array}{ccc} F(0) & \xrightarrow{f} & F(1) \\ {\scriptstyle \tau_0}\downarrow & & \downarrow{\scriptstyle \tau_1} \\ G(0) & \xrightarrow{g} & G(1) \end{array}$$

commute. The category $\mathcal{C}^{\mathbf{2}}$ is referred to as the *category of arrows* of $\mathcal{C}$.

3.5. **The 2-Category of all Categories.** Suppose that $\mathcal{C}$ and $\mathcal{D}$ are categories and that F, G and H are functors $\mathcal{C} \to \mathcal{D}$. In Section 3.4 we showed how to compose natural transformations $\sigma\colon F \Rightarrow G$ and $\tau\colon G \Rightarrow H$ in order to obtain $\tau\sigma\colon F \Rightarrow H$, thus equipping the set $\mathbf{Cat}(\mathcal{C}, \mathcal{D})$ of all functors from $\mathcal{C}$ to $\mathcal{D}$ with the stucture of a category. If we regard functors as one-dimensional cells (i.e., arrows), going horizontally between objects of $\mathbf{Cat}$, then we may picture natural transformations as two-dimensional cells going between arrows, with a composable pair of natural transformations σ and τ stacked vertically as in the diagram

$$\begin{array}{ccc} & \overset{F}{\frown\!\!\!\to} & \\ & \Downarrow\sigma & \\ \mathcal{C} & \xrightarrow{\quad G \quad} & \mathcal{D}. \\ & \Downarrow\tau & \\ & \underset{H}{\smile\!\!\!\to} & \end{array}$$

Hence we refer to this composition operation as *vertical composition* of natural transformations. We now describe another way of composing natural transformations; given the situation

$$\mathcal{B} \underset{K}{\overset{G}{\rightrightarrows}} \!\!{\scriptstyle \Downarrow\sigma}\; \mathcal{C} \underset{H}{\overset{F}{\rightrightarrows}} \!\!{\scriptstyle \Downarrow\tau}\; \mathcal{D},$$

the *horizontal composition* $\tau * \sigma$ is the natural transformation $FG \Rightarrow HK$ with components $(\tau * \sigma)_a\colon FG(a) \to HK(a)$ defined as the diagonal of the square

$$\begin{array}{ccc} FG(a) & \xrightarrow{\tau_{G(a)}} & HG(a) \\ {\scriptstyle F(\sigma_a)}\Big\downarrow & \searrow^{(\tau*\sigma)_*} & \Big\downarrow{\scriptstyle H(\sigma_a)} \\ FK(a) & \xrightarrow[\tau_{K(a)}]{} & HK(a), \end{array}$$

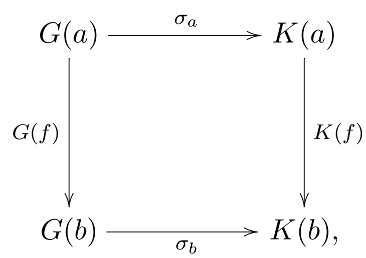
for all objects $a \in \mathcal{C}$. Note that this square commutes by naturality of τ. In other words $(\tau * \sigma)_a$ is given by

$$(\tau * \sigma)_a = H(\sigma_a) \cdot \tau_{G(a)} = \tau_{K(a)} \cdot F(\sigma_a), \tag{3.19}$$

for all $a \in \mathcal{C}$. In order to verify naturality of $\tau * \sigma$, suppose that $f\colon a \to b$ is an arrow in $\mathcal{C}$, and consider the diagram

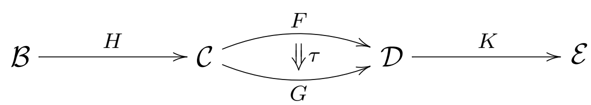
$$\begin{array}{ccccc} FG(a) & \xrightarrow{F(\sigma_a)} & FK(a) & \xrightarrow{\tau_{K(a)}} & HK(a) \\ {\scriptstyle FG(f)}\Big\downarrow & & {\scriptstyle FK(f)}\Big\downarrow & & \Big\downarrow{\scriptstyle HK(f)} \\ FG(b) & \xrightarrow[F(\sigma_b)]{} & FK(b) & \xrightarrow[\tau_{K(b)}]{} & HK(b). \end{array}$$

The square on the right commutes by naturality of τ, and the square on the left is the image of the square



which commutes by naturality of σ, and thus commutes itself. Therefore the whole diagram commutes; in particular, the outside rectangle commutes, which is naturality of $\tau * \sigma$.

Note that in the situation



the horizontal compositions $\tau * 1_H \colon FH \Rightarrow GH$ and $1_K * \tau \colon KF \Rightarrow KG$ are given by

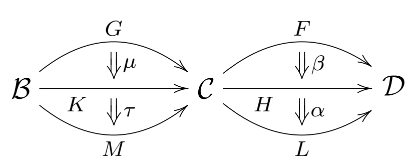
$$(\tau * 1_H)_b = \tau_{H(b)} \quad \text{and} \quad (1_K * \tau)_c = K(\tau_c), \tag{3.20}$$

for all $b \in \mathcal{B}$ and $c \in \mathcal{C}$. In particular, if we write $\iota_{\mathcal{C}}$ for the identity natural transformation $1_{1_{\mathcal{C}}}$ of the identity functor $1_{\mathcal{C}}$, then

$$\tau * \iota_{\mathcal{C}} = \iota_{\mathcal{D}} * \tau = \tau,$$

so that the identity natural transformations are identities for both the vertical and horizontal composition operations. It is easy to check that horizontal composition is associative; hence functors and natural transformations form a category with the horizontal, as well as with vertical, composition. As is the case with most algebraic systems having more than one binary operation, these operations are compatible with each other; specifically, we have the following proposition.

Proposition 3.21 (**Interchange law).** *Given functors and natural transformations as below*



the identity

$$(\alpha\beta) * (\tau\mu) = (\alpha * \tau) \cdot (\beta * \mu) \tag{3.22}$$

holds.

Proof. Exercise (fun diagrams). □

Corollary 3.23. *For all categories $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$, there is a functor*

$$\text{Comp}\colon \mathbf{Cat}(\mathcal{B},\mathcal{C}) \times \mathbf{Cat}(\mathcal{C},\mathcal{D}) \to \mathbf{Cat}(\mathcal{B},\mathcal{D})$$

given by $(G,F) \mapsto FG$ and $(\mu,\tau) \mapsto \tau * \mu$, for functors F, G, and natural transformations τ, μ.

Proof. The interchange law (3.22) is precisely the statement that Comp preserves composition. Given functors , we use the definition (3.19) to compute

$$\begin{aligned}(\iota_F * \iota_G)_a &= F((\iota_G)_a) \cdot (\iota_F)_{G(a)}\\ &= F(1_{G(a)}) \cdot 1_{F(G(a))}\\ &= 1_{F(G(a))} \cdot 1_{F(G(a))}\\ &= 1_{F(G(a))}\\ &= (\iota_{FG})_a,\end{aligned}$$

for all objects $a \in \mathcal{B}$. Hence Comp preserves identity arrows and is thus a functor. □

Corollary 3.23 explains the appearance of this other, mysterious, horizontal composition of natural transformations; if composition of functors is itself to be a functor, then horizontal composition of natural transformations is what this functor does to arrows. This result also provides a simple way to remember the interchange law; just think of what it means for the functor Comp to preserve composition of arrows.

The fact that each $\mathbf{Cat}(\mathcal{C},\mathcal{D})$ is itself an object of **Cat**, together with Corollary 3.23, means that the category **Cat** is enriched over **Cat**, just each of the categories **Ab**, **Pos**, and $_R\mathbf{Mod}$ (for R commutative) is enriched over itself (see the discussion immediately following Example 2.10).

3.6. **The Yoneda Embeddings.** In this section we will construct, for any category $\mathcal{C}$, an embedding of categories, called the *Yoneda functor*,

$$Y^*\colon \mathcal{C}^{\mathrm{op}} \to \mathrm{svf}(\mathcal{C}),$$

where $\mathrm{svf}(\mathcal{C})$ denotes the category $\mathbf{Set}^{\mathcal{C}}$ of set-valued functors on $\mathcal{C}$. Since an embedding is injective on both hom-sets and object sets, the functor Y^* embeds $\mathcal{C}^{\mathrm{op}}$ as a subcategory of $\mathrm{svf}(\mathcal{C})$, or equivalently, embeds $\mathcal{C}$ as a subcategory of $(\mathrm{svf}(\mathcal{C}))^{\mathrm{op}}$. The main result of the section, which is probably the single most useful tool in all of category theory, is the *Yoneda Lemma*, which implies in particular that the functor Y^* is full, and hence that $\mathcal{C}^{\mathrm{op}}$ embeds as a full subcategory of $\mathrm{svf}(\mathcal{C})$.

The primary reason that the Yoneda embedding is useful is that the category of set-valued functors $\mathrm{svf}(\mathcal{C})$ inherits many nice properties from **Set**. We will see in Section 2^n that many of the constructions available to us in **Set**, such as products and disjoint unions of objects, are also available in $\mathrm{svf}(\mathcal{C})$, regardless of the what the category $\mathcal{C}$ is. This is much the same as, say, the manner in which the set of all real-valued functions on an arbitrary set inherits algebraic and topological structure from $\mathbb{R}$.

The idea of embedding an object in some bigger object having desired properties is used in most branches of mathematics. Some examples include: compactifications of topological spaces; completions of metric spaces; and fields of quotients of integral domains.

A particularly close parallel occurs in functional analysis, with the construction of spaces of distributions. In this theory, one embeds certain spaces of functions into spaces of linear functionals defined on functions. To explain the analogy to the Yoneda embedding we simply assert at this point that the category **Set**, equipped with the operations of cartesian product and disjoint union, plays a role in **Cat** that is in many ways similar to that played by the field of scalars K as an object in the category **Vect** of all K-vector spaces (or more generally, that played by R in the category ${}_R\mathbf{Mod}$ of modules over a commutative ring R). Thus set-valued functors may be viewed as analogs of linear functionals, and the Yoneda functor as an embedding of a category $\mathcal{C}$ into a category of 'linear functionals' on $\mathcal{C}$.

More generally, if we regard an inner product $\langle -\,, - \rangle\colon V \times V \to K$ on a vector space V as a kind of hom-bifunctor on V, then the mapping $x \mapsto \langle x, - \rangle$ of V into its dual V^* (which is an embedding if $\langle -\,, - \rangle$ is nondegenerate) is the analog of the Yoneda functor.

The reader probably can guess how the Yoneda functor Y^* is defined, for we already know of a distinguished set-valued functor associated to each object a of a category $\mathcal{C}$, namely, the hom-functor $\mathcal{C}(a, -)$; furthermore, we saw in Example 3.7 that each arrow $f\colon a \to b$ in $\mathcal{C}$ induces a natural transformation

$$f^* = \mathcal{C}(f, -)\colon \mathcal{C}(b, -) \Rightarrow \mathcal{C}(a, -),$$

given by $g \mapsto gf$, for each arrow g with domain b. We thus may define $Y^*\colon \mathcal{C}^{\mathrm{op}} \to \mathrm{svf}(\mathcal{C})$ by

$$Y^*(a) = \mathcal{C}(a, -\,) \quad \text{and} \quad Y^*(f) = f^*, \tag{3.24}$$

for all objects a and arrows f of $\mathcal{C}$. It is easy to check that Y^* preserves identity arrows and composition, and is thus a functor. The disjointness property of arrow-sets implies that Y^* is injective on objects. To see that Y^* is faithful note that $f^*(1_b) = f$, for all arrows f with codomain b; hence, if $\operatorname{cod} f = \operatorname{cod} g$ and $f \neq g$, then $f^* \neq g^*$.

By duality, we obtain a functor $Y_*\colon \mathcal{C} \to \mathrm{svf}(\mathcal{C}^{\mathrm{op}})$, called the *dual Yoneda functor*, which is simply the functor Y^* applied to $\mathcal{C}^{\mathrm{op}}$. Recall that $\mathcal{C}^{\mathrm{op}}(a, -\,) = \mathcal{C}(-\,, a)$ and $\mathcal{C}^{\mathrm{op}}(f, -\,) = \mathcal{C}(-\,, f) = f_*$; hence Y_* satisfies

$$Y_*(a) = \mathcal{C}(-\,, a) \quad \text{and} \quad Y_*(f) = f_*, \tag{3.25}$$

for all objects a and arrows f of $\mathcal{C}$. It should be emphasized that the embeddings $Y_*\colon \mathcal{C} \to \mathrm{svf}(\mathcal{C}^{\mathrm{op}})$ and $(Y^*)^{\mathrm{op}}\colon \mathcal{C} \to (\mathrm{svf}(\mathcal{C}))^{\mathrm{op}}$ are distinct, for the categories $\mathrm{svf}(\mathcal{C}^{\mathrm{op}})$ and $(\mathrm{svf}(\mathcal{C}))^{\mathrm{op}}$ are not isomorphic in general.

Theorem 3.26 (The Yoneda lemma). *Suppose that $\mathcal{C}$ is a category, $F\colon \mathcal{C} \to \mathbf{Set}$ is a functor, and a is an object of $\mathcal{C}$. Then the mapping*

$$\theta = \theta_{F,a}\colon \mathrm{Nat}(\mathcal{C}(a, -\,), F) \to F(a),$$

defined by $\theta_{F,a}(\tau) = \tau_a(1_a)$, for all $\tau\colon \mathcal{C}(a, -\,) \Rightarrow F$, is a bijection.

Proof. We prove the result by constructing a map

$$\beta = \beta_{F,a}\colon F(a) \to \mathrm{Nat}(\mathcal{C}(a, -\,), F)$$

that is inverse to θ. Fix $x \in F(a)$ and, for any $b \in \mathcal{C}$, let $\beta(x)_b\colon \mathcal{C}(a, b) \to F(b)$ be defined by the rule $g \mapsto F(g)(x)$. We claim that the maps $\beta(x)_b$, for $b \in \mathcal{C}$, are the components of a natural transformation $\beta(x)\colon \mathcal{C}(a, -\,) \Rightarrow F$. To verify naturality, we must check that the square

$$\begin{array}{ccc}
\mathcal{C}(a,b) & \xrightarrow{\beta(x)_b} & F(b) \\
{\scriptstyle f_*=\mathcal{C}(a,f)}\big\downarrow & & \big\downarrow{\scriptstyle F(f)} \\
\mathcal{C}(a,c) & \xrightarrow[\beta(x)_c]{} & F(c)
\end{array}$$

commutes for all arrows $f\colon b \to c$. In order to this, we let $g \in \mathcal{C}(a, b)$ and compute

$$\begin{aligned}
F(f) \cdot \beta(x)_b(g) &= F(f) \cdot F(g)(x) \\
&= F(fg)(x) \\
&= \beta(x)_c(fg) \\
&= \beta(x)_c \cdot f_*(g);
\end{aligned}$$

hence $\beta(x)$ is a natural transformation. Now, for any $x \in F(a)$, we have

$$\begin{aligned}\theta\beta(x) &= \beta(x)_a(1_a)\\ &= F(1_a)(x)\\ &= 1_{F(a)}(x)\\ &= x,\end{aligned}$$

and so β is a right inverse for θ. On the other hand, given a natural transformation $\tau\colon \mathcal{C}(a,-) \Rightarrow F$, we need to show that $\beta\theta(\tau)_b = \tau_b$, for all objects b. But for any arrow $f\colon a \to b$, we have

$$\begin{aligned}\beta\theta(\tau)_b(f) &= \beta(\tau_a(1_a))_b(f)\\ &= F(f)(\tau_a(1_a))\\ &= \tau_b \cdot f_*(1_a)\\ &= \tau_b(f),\end{aligned}$$

where the third equality is by the naturality of τ. Hence β is also left inverse to θ. □

In the special case that F is the hom-functor $\mathcal{C}(b,-)$, we have the following corollary.

Corollary 3.27. *For all objects a and b of a category $\mathcal{C}$, the mapping*

$$\mathcal{C}(b,a) \to \mathrm{Nat}(\mathcal{C}(a,-), \mathcal{C}(b,-)),$$

defined by $f \mapsto f^*$, is a bijection.

Proof. Letting $F = \mathcal{C}(b,-)$, we obtain the bijection

$$\beta\colon \mathcal{C}(b,a) \to \mathrm{Nat}(\mathcal{C}(a,-), \mathcal{C}(b,-)),$$

defined in the proof of the Yoneda lemma, that is inverse to $\theta = \theta_{F,a}$. For each $f \in \mathcal{C}(b,a)$, the natural transformation $\beta(f)$ has components

$$\beta(f)_c\colon \mathcal{C}(a,c) \to \mathcal{C}(b,c),$$

defined for each object $c \in \mathcal{C}$ by $g \mapsto \mathcal{C}(b,g)(f) = f^*(g)$, and hence $\beta(f) = f^*$. □

Corollary 3.28. *The Yoneda functor $Y^*\colon \mathcal{C}^{op} \to \mathrm{svf}(\mathcal{C})$ and the dual functor $Y_*\colon \mathcal{C} \to \mathrm{svf}(\mathcal{C}^{op})$ are full embeddings.*

Proof. We have already observed that Y^* is an embedding. The fact that it is full is a direct consequence of Corollary 3.27. The result for Y_* follows by duality. □

To summarize, the Yoneda functor Y^* embeds a category $\mathcal{C}$ contravariantly in $\mathrm{svf}(\mathcal{C})$, as the full subcategory of covariant hom-functors on $\mathcal{C}$, and the dual functor Y_* embeds $\mathcal{C}$ covariantly in $\mathrm{svf}(\mathcal{C}^{\mathrm{op}})$, as the full subcategory of contravariant hom-functors on $\mathcal{C}$.

The categorical-minded reader should be a little unsatisfied with the statement of the Yoneda lemma as it stands. The lemma asserts that each of the maps $\theta_{F,a}$ is a bijection between sets whose definitions depend on the choice of $F \in \mathrm{svf}(\mathcal{C})$ and $a \in \mathcal{C}$. To be true to the spirit of category theory, the $\theta_{F,a}$ shouldn't be just some collection of unrelated bijections, but ought to be natural in the variables F and a. In order to state this naturality requirement precisely, we define functors

$$\mathrm{Ev}\colon \mathrm{svf}(\mathcal{C}) \times \mathcal{C} \to \mathbf{Set} \quad \text{and} \quad N\colon \mathrm{svf}(\mathcal{C}) \times \mathcal{C} \to \mathbf{Set},$$

whose values on objects (F, a) are $F(a)$ and $\mathrm{Nat}(\mathcal{C}(a, -\,), F)$, respectively. The functor Ev is called *evaluation*, since $\mathrm{Ev}(F, a)$ is just F evaluated at a; we don't bother to name N. We now define Ev and N on arrows. Suppose that $(\tau, f)\colon (F, a) \to (G, b)$ is an arrow in $\mathrm{svf}(\mathcal{C}) \times \mathcal{C}$, in other words, $\tau\colon F \Rightarrow G$ is a natural transformation of set-valued functors on $\mathcal{C}$ and $f\colon a \to b$ is an arrow in $\mathcal{C}$. There are two obvious choices of maps $F(a) \to G(b)$ that can be associated to the pair (τ, f), namely $\tau_b \cdot F(f)$ and $G(f) \cdot \tau_a$. But these are equal by naturality of τ; hence we define $\mathrm{Ev}(\tau, f)\colon F(a) \to G(b)$ by

$$\mathrm{Ev}(\tau, f) = \tau_b \cdot F(f) = G(f) \cdot \tau_a.$$

It is easy to verify that Ev is thus a functor. Next, we define

$$N(\tau, f)\colon \mathrm{Nat}(\mathcal{C}(a, -\,), F) \to \mathrm{Nat}(\mathcal{C}(b, -\,), G),$$

by setting $N(\tau, f)(\alpha)$ equal to the composition $\tau\alpha f^*$ of natural transformations

$$\mathcal{C}(b, -\,) \overset{f^*}{\Longrightarrow} \mathcal{C}(a, -\,) \overset{\alpha}{\Longrightarrow} F \overset{\tau}{\Longrightarrow} G,$$

for all $\alpha \in \mathrm{Nat}(\mathcal{C}(a, -\,), F)$, thus making N a functor.

Theorem 3.29 (**Yoneda lemma, revisited).** *For any category $\mathcal{C}$, the mappings $\theta_{F,a}\colon \mathrm{Nat}(\mathcal{C}(a, -\,), F) \to F(a)$, defined by $\tau \mapsto \tau_a(1_a)$, for all functors $F\colon \mathcal{C} \to \mathbf{Set}$ and objects $a \in \mathcal{C}$, are the components of a natural isomorphism $\theta\colon N \Rightarrow Ev$ of functors* $\mathrm{svf}(\mathcal{C}) \times \mathcal{C} \to \mathbf{Set}$.

Proof. We need to show that the square

$$\begin{array}{ccc}
\mathrm{Nat}(\mathcal{C}(a, -\,), F) & \xrightarrow{\theta_{F,a}} & F(a) \\
\Big\downarrow {\scriptstyle N(\tau,f)} & & \Big\downarrow {\scriptstyle \mathrm{Ev}(\tau,f)} \\
\mathrm{Nat}(\mathcal{C}(b, -\,), G) & \xrightarrow[\theta_{G,b}]{} & G(b)
\end{array}$$

commutes, for all arrows $(\tau, f)\colon (F, a) \to (G, b)$ in $\mathrm{svf}(\mathcal{C}) \times \mathcal{C}$. Suppose that $\alpha \in \mathrm{Nat}(\mathcal{C}(a, -\,), F)$, and compute

$$\begin{aligned}
\mathrm{Ev}(\tau, f) \cdot \theta_{F,a}(\alpha) &= \mathrm{Ev}(\tau, f) \cdot \alpha_a(1_a) \\
&= \tau_b \cdot F(f) \cdot \alpha_a(1_a).
\end{aligned}$$

On the other hand,

$$\begin{aligned}\theta_{G,b} \cdot N(\tau, f)(\alpha) &= (\tau\alpha f^*)_b(1_b) \\ &= \tau_b \cdot \alpha_b \cdot f_b^*(1_b) \\ &= \tau_b \cdot \alpha_b\,(f),\end{aligned}$$

which is equal to $\tau_b \cdot F(f) \cdot \alpha_a(1_a)$, by naturality of α. Hence the square commutes. □

3.7. **Representable Functors.**

Definition 3.30. A *representation* of a set-valued functor $F\colon \mathcal{C} \to \mathbf{Set}$ consists of a pair (γ, r), where r is an object of $\mathcal{C}$ and $r\colon \mathcal{C}(r, -\,) \Rightarrow F$ is a natural isomorphism. If such a representation exists, the functor F is called *representable* and r is called a *representing object* for F.

Because a representable functor F is completely determined by any representing object r, questions about F often can be translated into questions about r, which may be easier to answer. If a functor is representable, its representing object is not unique (of course), but is unique up to isomorphism. The next proposition shows that an even stronger form of uniqueness holds.

Proposition 3.31. *If (γ, r) and (β, s) are representations for a functor $F\colon \mathcal{C} \to \mathbf{Set}$ then there exists a unique isomorphism $f\colon s \to r$ in $\mathcal{C}$ such that $\gamma = \beta \cdot f^*$.*

Proof. The composition $\beta^{-1}\gamma$ is a natural isomorphism from $\mathcal{C}(r, -\,)$ to $\mathcal{C}(s, -\,)$. It follows from Corollary 3.27 that this natural isomorphism is equal to f^*, for a unique isomorphism $f\colon s \to r$. □

Example 3.32. Recall the underlying set functor $U\colon {}_R\mathbf{Mod} \to \mathbf{Set}$, that takes an R-module to its underlying set and a homomorphism $M \to N$ to the corresponding function $U(M) \to U(N)$. For any fixed set S, let $G\colon {}_R\mathbf{Mod} \to \mathbf{Set}$ be the functor defined on objects M by setting $G(M)$ equal to the set of all functions from S into the $U(M)$, and on arrows $f\colon M \to N$ by $G(f)(h) = fh$, for all $h \in G(M)$. In other words, G is equal to the composition of functors

$$ {}_R\mathbf{Mod} \xrightarrow{\ U\ } \mathbf{Set} \xrightarrow{\mathbf{Set}(S,-)} \mathbf{Set}, $$

which we denote by $\mathbf{Set}(S, U(-\,))$. Let F_S denote the free R-module on the set S. (Actually, there are many different ways of constructing F_S, let's just think of it as formal R-linear combinations of elements of S for now, whatever *they* are!) For any left R-module M, there is a bijection

$$\gamma_M\colon {}_R\mathbf{Mod}(F_S, M) \to \mathbf{Set}(S, U(M)) = G(M),$$

given by $f \mapsto U(f) \cdot i$, where $i\colon S \to U(F_S)$ is the inclusion map. It is straightforward to show that these bijections are natural in M, that is, they are the components of a natural isomorphism $\gamma\colon {}_R\mathbf{Mod}(F_S, -\,) \Rightarrow G$. Hence G is a representable functor, represented by the object F_S.

Example 3.33. Suppose that K is a normal subgroup of a group H. Define a functor $F = F_{H,K}\colon \mathbf{Grp} \to \mathbf{Set}$ by setting

$$F(G) = \{f \in \mathbf{Grp}(H, G) : \ker f \geq K\},$$

for all groups G, and $F(f) = f_*$, for all homomorphisms f. The fundamental homomorphism theorem of group theory gives us a bijection

$$\tau_G\colon \mathbf{Grp}(H/K, G) \to F(G)$$
$$g \mapsto g\pi,$$

for each group G, where $\pi\colon H \to H/K$ is the canonical surjection. The maps τ_G are the components of a natural isomorphism

$$\mathbf{Grp}(H/K, -\,) \Rightarrow F,$$

and thus the functor F is representable, with representing object H/K.

Example 3.34. Suppose that M and N are R-modules, where R is a commutative ring. Define a functor $B = B_{M,N}\colon {}_R\mathbf{Mod} \to \mathbf{Set}$ by letting $B(P)$ equal the set $\mathrm{Bil}(M \times N, P)$ of all bilinear maps $M \times N \to P$, for all R-modules P, and for any homomorphism of R-modules $g\colon P \to Q$, letting $B(g)\colon B(P) \to B(Q)$ be given by $f \mapsto gf$. The defining property of the tensor product $M \otimes_R N$ is that the map

$$\text{(3.35)} \qquad {}_R\mathbf{Mod}(M \otimes_R N, P) \to \mathrm{Bil}(M \times N, P),$$

given by $f \mapsto ft$ is a bijection, where $t\colon M \times N \to M \otimes_R N$ is the universal bilinear map determined by $t(x, y) = x \otimes y$. Once again, it is easy to check that these maps are the components of a natural isomorphism ${}_R\mathbf{Mod}(M \otimes_R N, -\,) \Rightarrow B$, and so B is representable, with representing object $M \otimes_R N$.

We remark that the set $\mathrm{Bil}(M \times N, P)$ is an R-module with pointwise sum and scalar multiplication, as is the set ${}_R\mathbf{Mod}(M \otimes_R N, P)$, and the map (3.35) is an R-module isomorphism. Hence B is actually a representable functor ${}_R\mathbf{Mod} \to {}_R\mathbf{Mod}$.

Example 3.36. For any set U, define a contravariant functor

$$R = R_U\colon \mathbf{Set} \to \mathbf{Set}$$

by setting $R(S) = \mathbf{Rel}(S, U)$, and $R(f)(h) = hf$, for all sets S and functions $f\colon S \to T$. In other words, R is just the restriction of the contravariant hom-functor $\mathbf{Rel}(-\,, U)$ to the subcategory $\mathbf{Set}$ of $\mathbf{Rel}$. For any $f \in \mathbf{Rel}(S, U)$, define a function $\bar{f}$ from S to the power-set $P(U)$ of U by

$$\bar{f}(x) = \{y \in U\colon (x, y) \in f\},$$

for all $x \in S$. The mappings $\mathbf{Rel}(S, U) \to \mathbf{Set}(S, P(U))$ are the components of a natural isomorphism $R \Rightarrow \mathbf{Set}(-\,, P(U))$, and so R is a representable contravariant functor, with representing object $P(U)$.

For any category $\mathcal{C}$, we denote by rsvf($\mathcal{C}^{\text{op}}$) the full subcategory of svf($\mathcal{C}^{\text{op}}$) consisting of all representable set-valued functors, or representable *presheaves*, on $\mathcal{C}$. The Yoneda functor $Y_*\colon \mathcal{C} \to \text{svf}(\mathcal{C})$ takes values in rsvf($\mathcal{C}$), hence by restricting the codomain of Y_*, we obtain a functor $\mathcal{C} \to \text{rsvf}(\mathcal{C}^{\text{op}})$, also denoted by Y_*, that embeds $\mathcal{C}$ as a full subcategory of rsvf($\mathcal{C}$).

Now, even though each object in rsvf($\mathcal{C}^{\text{op}}$) is a representable functor, it is not necessarily equal to a hom-functor, and thus is not necessarily in the image of Y_*. Therefore Y_* is not an isomorphism. However, since Y_* is full and faithful, and every object in rsvf($\mathcal{C}$) is isomorphic to an object in its image, we know by Propostion 3.16 that Y_* is an equivalence.

There are many ways to define an equivalence $R\colon \text{rsvf}(\mathcal{C}^{\text{op}}) \to \mathcal{C}$ left inverse to Y_*. In fact, to define R we must choose a representation (γ_F, r_F) for each $F \in \text{rsvf}(\mathcal{C})$ and set $R(F) = r_F$, and then for each natural transformation $\alpha\colon F \Rightarrow G$, set $R(\alpha)$ equal to the composition

$$\mathcal{C}(-\,, r_F) \xRightarrow{\gamma_F} F \xRightarrow{\alpha} G \xRightarrow{\gamma_G^{-1}} \mathcal{C}(-\,, r_G).$$

It then follows immediately that R is a functor, $Y_* \cdot R = 1_{\mathcal{C}}$ and $1_{\mathcal{C}} \cdot R \cong 1_{\text{rsvf}(\mathcal{C})}$.

3.8. **Exercises.**

(1) Show by example that condition (ii) in the definition (2.1) of functor is not redundant.

(2) Suppose that the categories $\mathcal{C}$ and $\mathcal{D}$ are groups, and that F and G are functors (i.e., homomorphisms) from $\mathcal{C}$ to $\mathcal{D}$. Describe the natural transformations from F to G.

(3) For any category $\mathcal{D}$, let Dom and Cod be the functors from the category of arrows $\mathcal{D}^{\mathbf{2}}$ to $\mathcal{D}$ defined on objects

$$\mathrm{Dom}(f) = \mathrm{dom}\, f \quad \text{and} \quad \mathrm{Cod}(f) = \mathrm{cod}\, f,$$

and on arrows by

$$\mathrm{Dom}(g, h) = g \quad \text{and} \quad \mathrm{Cod}(g, h) = g.$$

Show that a natural transformation $\tau\colon F \to G$ is the same thing as a functor $\tau\colon \mathcal{C} \to \mathcal{D}^{\mathbf{2}}$ satisfying $\mathrm{Dom} \cdot \tau = F$ and $\mathrm{Cod} \cdot \tau = G$.

(4) Show that the contravariant hom- and power-set functors

$$\mathbf{Set}(-\,, \{0, 1\})\colon \mathbf{Set} \to \mathbf{Set} \quad \text{and} \quad \bar{P}\colon \mathbf{Set} \to \mathbf{Set},$$

described in Examples 2.19 and 2.17, are naturally isomorphic.

(5) Suppose that the category $\mathcal{G}$ is a group. Describe the Yoneda functor $Y_*\colon \mathcal{G} \to \mathrm{svf}(\mathcal{G}^{\mathrm{op}})$ in detail, using the language of group theory. To which group-theoretic facts do the facts that Y_* is full and faithful correspond?

(6) Show, by example, that the categories $\mathrm{svf}(\mathcal{C}^{\mathrm{op}})$ and $(\mathrm{svf}(\mathcal{C}))^{\mathrm{op}}$ are not isomorphic. Are they equivalent?

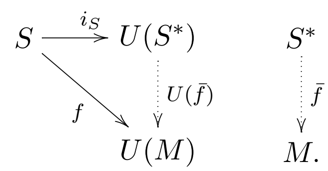
(7) Prove the interchange law (3.22).

(8) Suppose that $\mathcal{C}$ is any category, and that $1_{\mathcal{C}}\colon \mathcal{C} \to \mathcal{C}$ is the identity functor on $\mathcal{C}$. Show that the set $M = \mathrm{Nat}(1_{\mathcal{C}}, 1_{\mathcal{C}})$, equipped with either vertical or horizontal composition of natural transformations, is a *commutative* monoid. (Hint: the interchange law (3.22) is useful here.)

4. Adjoint Functors and Limits

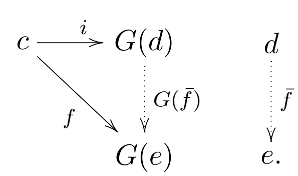
4.1. **Adjoint Functors.** Let's begin with an example. For any set S, the *free monoid* on S is the set S^* of all sequences $(s_1, \dots, s_k)$ of elements of S, for $k \geq 0$; the binary operation on S^* is concatenation of sequences, and the identity element is the empty sequence. We let $i_S \colon S \to S^*$ denote the injection given by $s \mapsto (s)$, for all $s \in S$. The defining, 'universal', property of the pair (S, i_S) is usually stated as follows: for any monoid M and any function $f \colon S \to M$, there exists a unique homomorphism of monoids $\bar{f} \colon S^* \to M$ satisfying $f = \bar{f} i_S$. The homomorphism $\bar{f}$ maps the sequence $(s_1, \dots, s_k)$ to the product $f(s_1) \cdots f(s_k)$ in M.

Now in category theory it's not very good form to consider arrows, such as i_S above, between objects belonging to different categories. To state this universal property correctly, we let $U \colon \mathbf{Mon} \to \mathbf{Set}$ denote the underlying set functor, and note that i_S is really a function $S \to U(S^*)$; then we have: for any monoid M and function $f \colon S \to U(M)$, there exists a unique monoid homomorphism $\bar{f} \colon S^* \to M$ such that $f = U(\bar{f}) \cdot i_S$. In a diagram:



In words: every arrow from S to an object in the image of U factors uniquely through i_S by an arrow in the image of U. We make the following general definition.

Definition 4.1. Suppose that $G \colon \mathcal{D} \to \mathcal{C}$ is a functor and c is an object of $\mathcal{C}$. An arrow $i \colon c \to G(d)$, for some $d \in \mathcal{D}$, is *universal from c to G*, if each arrow $f \colon c \to G(e)$, for $e \in \mathcal{D}$, factors uniquely as $f = G(\bar{f}) \cdot i$; in other words, if the diagram



commutes, for some unique $\bar{f} \colon d \to e$ in $\mathcal{D}$.

Hence the arrow $i_S \colon S \to S^*$ above is universal from S to U.

We now define a functor $F \colon \mathbf{Set} \to \mathbf{Mon}$, by setting $F(S) = S^*$, for all sets S, and defining $F(f) \colon F(S) \to F(T)$ by $(s_1, \dots s_k) \mapsto (f(s_1), \dots f(s_k))$, for every function $f \colon S \to T$. The universal property of free monoids now can be formulated by the statement that the maps

$$\gamma_{S,M} \colon \mathbf{Mon}(F(S), M) \to \mathbf{Set}(S, U(M)) \qquad g \mapsto U(g) \cdot i_S \tag{4.2}$$

are bijections, for all sets S and monoids M. Furthermore, these bijections are natural in the variables S and M; that is, for all functions $f\colon S' \to S$ and homomorphisms $g\colon M \to M'$, the square

$$\begin{array}{ccc} \mathbf{Mon}(F(S), M) & \xrightarrow{\gamma_{S,M}} & \mathbf{Set}(S, U(M)) \\ \downarrow{\scriptstyle \mathbf{Mon}(F(f),g)} & & \downarrow{\scriptstyle \mathbf{Set}(f,U(g))} \\ \mathbf{Mon}(F(S'), M') & \xrightarrow[\gamma_{S',M'}]{} & \mathbf{Set}(S', U(M')) \end{array}$$

commutes. In particular, naturality in M means that for each set S, the bijections $\gamma_{S,M}$ are the components of a natural isomorphism

$$\gamma_{S,-}: \mathbf{Mon}(F(S), -\,) \Rightarrow \mathbf{Set}(S, U(-\,));$$

in other words, the pair $(\gamma_{S,-}, F(S))$ is a representation of the functor $\mathbf{Set}(S, U(-\,))\colon \mathbf{Mon} \to \mathbf{Set}$. On the other hand, naturality in S means that, for each monoid M, we have a natural isomorphism

$$\gamma_{-\,,M}\colon \mathbf{Mon}(F(-\,), M) \Rightarrow \mathbf{Set}(-\,, U(M)),$$

that is, the pair $((\gamma_{-\,,M})^{-1}, U(M))$ is a representation of the contravariant functor $\mathbf{Mon}(F(-\,), M)\colon \mathbf{Set} \to \mathbf{Set}$.

A similar situation occurs for any 'free' construction; for example, the free R-module on a set, the free group on a set, the free category on a graph. Indeed, as we shall see, a pair of functors related as are F and U above is always lurking behind the scenes whenever any universal construction is made. We thus make the following definition.

Definition 4.3. Let $\mathcal{C}$ and $\mathcal{D}$ be categories. An *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is a triple (F, G, γ), where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ are functors and

$$\gamma\colon \mathcal{D}(F(-\,), -\,) \Rightarrow \mathcal{C}(-\,, G(-\,))$$

is a natural isomorphism of set-valued functors on $\mathcal{C}^{\mathrm{op}} \times \mathcal{D}$. The ordered pair (F, G) is an *adjoint pair* of functors, with F a *left adjoint* for G, or equivalently, G a *right adjoint* for F. The natural isomorphism γ is the *adjugant isomorphism* of the adjunction. We write $F \dashv G$ to denote that (F, G) is an adjoint pair of functors and $\gamma\colon F \dashv G$ to indicate that γ is the adjugant isomorphism of this adjunction.

The natural isomorphism γ consists of isomorphisms

$$\gamma_{c,d}\colon \mathcal{D}(F(c), d) \to \mathcal{C}(c, G(d)) \tag{4.4}$$

such that, for all arrows $f\colon c' \to c$ in $\mathcal{C}$ and $g\colon d \to d'$ in $\mathcal{D}$, the square

$$
\begin{array}{ccc}
\mathcal{D}(F(c),d) & \xrightarrow{\gamma_{c,d}} & \mathcal{C}(c,G(d)) \\
{\scriptstyle \mathcal{D}(F(f),g)}\downarrow & & \downarrow{\scriptstyle \mathcal{C}(f,G(g))} \\
\mathcal{D}(F(c'),d') & \xrightarrow[\gamma_{c',d'}]{} & \mathcal{C}(c',G(d'))
\end{array}
\tag{4.5}
$$

commutes; in other words, such that the equality

$$\gamma_{c',d'}(g \cdot h \cdot F(f)) = G(g) \cdot \gamma_{c,d}(h) \cdot f$$

holds for all $h\colon F(c) \to d$ in $\mathcal{D}$. The commutativity of the square (4.5) is equivalent to that of both the squares

$$
\begin{array}{ccc}
\mathcal{D}(F(c),d) & \xrightarrow{\gamma_{c,d}} & \mathcal{C}(c,G(d)) \\
{\scriptstyle g_*}\downarrow & & \downarrow{\scriptstyle G(g)_*} \\
\mathcal{D}(F(c),d') & \xrightarrow[\gamma_{c,d'}]{} & \mathcal{C}(c,G(d'))
\end{array}
\qquad
\begin{array}{ccc}
\mathcal{D}(F(c),d) & \xrightarrow{\gamma_{c,d}} & \mathcal{C}(c,G(d)) \\
{\scriptstyle F(f)^*}\downarrow & & \downarrow{\scriptstyle f^*} \\
\mathcal{D}(F(c'),d) & \xrightarrow[\gamma_{c',d}]{} & \mathcal{C}(c',G(d)),
\end{array}
$$

that is, the naturality of γ in the variables c and d is equivalent to γ being natural in each variable separately.

Example 4.6. The free monoid functor $F\colon \mathbf{Set} \to \mathbf{Mon}$ is a left adjoint of the underlying set functor $U\colon \mathbf{Mon} \to \mathbf{Set}$. The adjugant isomorphism $\gamma\colon F \dashv U$ is given by (4.2).

An adjunction $\gamma\colon F \dashv G$ may be pictured by the following two-dimensional diagram, or *two-cell*,

$$
\begin{array}{ccc}
\mathcal{C}^{\mathrm{op}} \times \mathcal{D} & \xrightarrow{F^{\mathrm{op}} \times 1_{\mathcal{D}}} & \mathcal{D}^{\mathrm{op}} \times \mathcal{D} \\
{\scriptstyle 1_{\mathcal{C}^{\mathrm{op}}} \times G}\downarrow & \swarrow\!\!\!\!\swarrow \gamma & \downarrow{\scriptstyle \mathcal{D}(-,-)} \\
\mathcal{C}^{\mathrm{op}} \times \mathcal{C} & \xrightarrow[\mathcal{C}(-,-)]{} & \mathbf{Set}
\end{array}
\tag{4.7}
$$

in **Cat**, which is not a commutative square, but indicates that γ is a natural transformation from the composition

$$\mathcal{D}(F(-),-) = \mathcal{D}(-,-) \cdot (F^{\mathrm{op}} \times 1_{\mathcal{D}})$$

to

$$\mathcal{C}(-\,,G(-)) = \mathcal{C}(-\,,-\,)\cdot(1_{\mathcal{C}^{op}})\times G.$$

Since γ is a natural isomorphism, we say that the square commutes *up to natural isomorphism*.

The origin of the terminology 'adjoint pair' comes from linear algebra. Suppose that V and W are K-vector spaces, each with inner product denoted by $\langle -\,,-\,\rangle$. Then the adjoint of a linear map $f\colon V\to W$ is the unique map $g\colon W\to V$ satisfying $\langle f(x),y\rangle = \langle x, g(y)\rangle$, for all $x\in V$ and $y\in W$; in other words, such that the square

$$\begin{array}{ccc}
V\times W & \xrightarrow{1\times g} & V\times V \\
{\scriptstyle f\times 1}\downarrow & & \downarrow{\scriptstyle \langle -\,,-\,\rangle} \\
W\times W & \xrightarrow[\langle -\,,-\,\rangle]{} & K
\end{array}$$

commutes. Once again, we observe close analogies between inner products and hom-bifunctors, and between the field of scalars and the category **Set**.

Proposition 4.8. *Suppose we are given categories and functors*

$$\mathcal{B} \underset{H}{\overset{G}{\rightleftarrows}} \mathcal{C} \underset{K}{\overset{F}{\rightleftarrows}} \mathcal{D}\,,$$

with $G \dashv H$ *and* $F \dashv K$. *Then* $FG \dashv HK$.

Proof. Suppose that $\gamma\colon G\dashv H$ and $\alpha\colon F\dashv K$. For all $b\in\mathcal{B}$ and $d\in\mathcal{D}$, define $\beta_{b,d}\colon \mathcal{D}(FG(b),d)\to\mathcal{B}(b,HK(d))$ as the composition

$$\mathcal{D}(FG(b),d) \xrightarrow{\alpha_{G(b),d}} \mathcal{C}(G(b),K(d)) \xrightarrow{\gamma_{b,K(d)}} \mathcal{B}(b,HK(d)).$$

Then $\beta\colon \mathcal{D}(FG(-\,),-\,)\Rightarrow\mathcal{B}(-\,,HK(-\,))$ is a natural isomorphism. □

Proposition 4.9. *If* F *and* G *are functors, then* $F\dashv G$ *if and only if* $G^{op}\dashv F^{op}$.

Proof. Replace $\mathcal{C}$, $\mathcal{D}$, F and G in the diagram (4.7) by $\mathcal{C}^{\mathrm{op}}$, $\mathcal{D}^{\mathrm{op}}$, F^{op} and G^{op}, respectively. □

Hence, the dual of the concept of left adjoint is right adjoint, and *vice-versa*.

Proposition 4.10. *Let* $\mathcal{C}$ *and* $\mathcal{D}$ *be categories.*

(i) A functor $F\colon\mathcal{C}\to\mathcal{D}$ *has a right adjoint if and only if the functor* $\mathcal{D}(F(-\,),d)\colon\mathcal{C}\to\mathbf{Set}$ *is a representable, for all* $d\in\mathcal{D}$.

(ii) A functor $G\colon\mathcal{D}\to\mathcal{C}$ *has a left adjoint if and only if the functor* $\mathcal{C}(c,G(-\,))\colon\mathcal{D}\to\mathbf{Set}$ *is representable, for all* $c\in\mathcal{C}$.

Proof. We prove the second statement; the proof of the first is dual. Suppose that F is a left adjoint for $G\colon \mathcal{D} \to \mathcal{C}$, with $\gamma\colon F \dashv G$. For every $c \in \mathcal{C}$ there is a natural isomorphism

$$\gamma_{c,-} : \mathcal{D}(F(c), -) \Rightarrow \mathcal{C}(c, G(-)),$$

and so $\mathcal{C}(c, G(-))$ is representable, with representing object $F(c)$.

For the converse, suppose that each $\mathcal{C}(c, G(-))$ is representable, and for each $c \in \mathcal{C}$ choose a representing object $r_c \in \mathcal{D}$ and natural isomorphism $\rho_c\colon \mathcal{D}(r_c, -) \Rightarrow \mathcal{C}(c, G(-))$. For all arrows $f\colon c' \to c$ in $\mathcal{C}$, the composition

$$\mathcal{D}(r_c, -) \xrightarrow{\rho_c} \mathcal{D}(c, G(-)) \xrightarrow{f^*} \mathcal{D}(c', G(-)) \xrightarrow{\rho_{c'}^{-1}} \mathcal{D}(r_{c'}, -)$$

is a natural transformation $\mathcal{D}(r_c, -) \Rightarrow \mathcal{D}(r_{c'}, -)$ and thus by the Yoneda corollary (3.27), must be equal to $(\bar{f})^*$, for some $\bar{f}\colon r_{c'} \to r_c$. Hence we define $F\colon \mathcal{C} \to \mathcal{D}$ by setting

$$F(c) = r_c \quad \text{and} \quad F(f) = \bar{f},$$

for all objects c and arrows f in $\mathcal{C}$; it is readily verified that F is indeed a functor. The bijections

$$\gamma_{c,d} = (\rho_c)_d\colon \mathcal{D}(F(c), d) \to \mathcal{C}(c, G(d))$$

are natural in d because $\gamma_{c,-} = \rho_c$ is a natural transformation, for all c. Now, for all $f\colon c' \to c$ in $\mathcal{C}$, the square

$$\begin{array}{ccc} \mathcal{D}(F(c), d) & \xrightarrow{\gamma_{c,d}} & \mathcal{C}(c, G(d)) \\ \downarrow{\scriptstyle F(f)^*} & & \downarrow{\scriptstyle f^*} \\ \mathcal{D}(F(c'), d') & \xrightarrow[\gamma_{c',d'}]{} & \mathcal{C}(c', G(d')) \end{array}$$

commutes, by definition of $F(f)$. Hence the family $\gamma_{c,d}$ is natural in c as well, and so $\gamma\colon F \dashv G$. □

Proposition 4.11. *If $F \dashv G$ and $F' \dashv G$, then F and F' are naturally isomorphic. If $F \dashv G$ and $F \dashv G'$, then G and G' are naturally isomorphic.*

Proof. We prove the first statement; the second is dual. Suppose that $\gamma\colon F \dashv G$ and $\gamma'\colon F' \dashv G$. Then $(\gamma_{c,-}, F(c))$ and $(\gamma_{c',-}, F(c'))$ are representations of the functor $\mathcal{C}(c, G(-))\colon \mathcal{D} \to \mathbf{Set}$, for all $c \in \mathcal{C}$, and so by Proposition 3.31 there is a unique isomorphism $\tau_c\colon F'(c) \to F(c)$ such that $\gamma'_{c,-} \cdot (\tau_c)^* = \gamma_{c,-}$. It is not difficult to verify that the maps τ_c are natural in c.

□

4.2. **The Unit and Counit of an Adjunction.** So far, we have considered three different levels of similarity between categories $\mathcal{C}$ and $\mathcal{D}$, namely, equality, isomorphism and equivalence. Each of these relations is a special case of the next; to say that $\mathcal{C}$ and $\mathcal{D}$ are equivalent means that there exist functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ and natural isomorphisms $1_{\mathcal{C}} \Rightarrow GF$ and $FG \Rightarrow 1_{\mathcal{D}}$, that is, F and G are *weak inverses* of one another. If $\mathcal{C}$ and $\mathcal{D}$ are isomorphic then these natural transformations can be taken to be identities, while if $\mathcal{C}$ and $\mathcal{D}$ are equal, the functors F and G themselves can be taken to be identities. We will now show that the existence of an adjunction is a further generalization of equivalence between categories. Specifically, if F and G above are an adjoint pair, with $\gamma\colon F \dashv G$, then there are natural transformations

$$\eta\colon 1_{\mathcal{C}} \Rightarrow GF \quad \text{and} \quad \epsilon\colon FG \Rightarrow 1_{\mathcal{D}},$$

called the *unit* and *counit* of the adjunction, respectively, each of which determines γ uniquely, and such that F and G are weak inverses if and only if η and ϵ are natural isomorphisms. We shall also see that η and ϵ satisfy a certain pair of identities, and that any pair of natural transformations $\eta\colon 1_{\mathcal{C}} \Rightarrow GF$ and $\epsilon\colon FG \Rightarrow 1_{\mathcal{D}}$ satifying these identities must be the unit and counit of a unique adjunction $\gamma\colon F \dashv G$.

Supppose that $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{D}$ are an adjoint pair of functors, with adjugant isomorphism $\gamma\colon F \dashv G$. By choosing $d = F(c)$ in Equation 4.4, we obtain a bijection

$$\gamma_{c,F(c)}\colon \mathcal{D}(F(c), F(c)) \to \mathcal{C}(c, GF(c)),$$

for each object $c \in \mathcal{C}$.

Proposition 4.12. *The arrows $\eta_c\colon c \to GF(c)$ defined for all $c \in \mathcal{C}$ by*

$$\eta_c = \gamma_{c,F(c)}(1_{F(c)}), \tag{4.13}$$

are the components of a natural transformation $\eta\colon 1_{\mathcal{C}} \Rightarrow GF$.

Proof. Suppose $f\colon c \to c'$ is an arrow in $\mathcal{C}$. We need to show that the square

$$\begin{array}{ccc} c & \xrightarrow{\eta_c} & GF(c) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle GF(f)} \\ c' & \xrightarrow[\eta_{c'}]{} & GF(c') \end{array} \tag{4.14}$$

commutes; that is,

$$\gamma_{c',F(c')}(1_{F(c')}) \cdot f = GF(f) \cdot \gamma_{c,F(c)}(1_{F(c)}). \tag{4.15}$$

Using the naturality of γ in both variables, we obtain the commutative diagram

$$\begin{array}{ccccc}
\mathcal{D}(F(c),F(c)) & \xrightarrow{F(f)_*} & \mathcal{D}(F(c),F(c')) & \xleftarrow{F(f)^*} & \mathcal{D}(F(c'),F(c')) \\
\Big\downarrow {\scriptstyle \gamma_{c,F(c)}} & & \Big\downarrow {\scriptstyle \gamma_{c,F(c')}} & & \Big\downarrow {\scriptstyle \gamma_{c',F(c')}} \\
\mathcal{C}(c,GF(c)) & \xrightarrow[GF(f)_*]{} & \mathcal{C}(c,GF(c')) & \xleftarrow[f^*]{} & \mathcal{C}(c',GF(c')).
\end{array}$$

Equation 4.15 is the statement that the identity arrows $1_{F(c')}$ and $1_{F(c)}$ in the upper corners of this diagram map along the perimeter to the same element of $\mathcal{C}(c, GF(c'))$. Since the diagram commutes, this is equivalent to the equality

$$\gamma_{c,F(c')} \cdot F(f)_*(1_{F(c)}) = \gamma_{c,F(c')} \cdot F(f)^*(1_{F(c')}),$$

which holds because $F(f)_*(1_{F(c)}) = F(f) \cdot 1_{F(c)} = F(f)$ and $F(f)^*(1_{F(c')}) = 1_{F(c')} \cdot F(f) = F(f)$. $\square$

Definition 4.16. The natural transformation $\eta\colon 1_{\mathcal{C}} \Rightarrow GF$ defined by Equation 4.13 is the *unit* of the adjunction $\gamma\colon F \dashv G$.

Proposition 4.17. *Suppose that η is the unit of an adjunction $\gamma\colon F \dashv G$, where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$. Then, for all $c \in \mathcal{C}$ and $d \in \mathcal{D}$, the bijection $\gamma_{c,d}\colon \mathcal{D}(F(c), d) \to \mathcal{C}(c, G(d))$ is given by*

$$\gamma_{c,d}(f) = G(f) \cdot \eta_c, \tag{4.18}$$

for all arrows $f\colon F(c) \to d$ in $\mathcal{D}$.

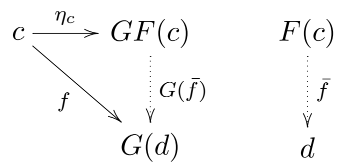
Proof. Follow the identity arrow $1_{\mathcal{D}}$ around the commutative square

$$\begin{array}{ccc}
\mathcal{D}(F(c),F(c)) & \xrightarrow{\gamma_{c,F(c)}} & \mathcal{C}(c,GF(c)) \\
\Big\downarrow {\scriptstyle f_*} & & \Big\downarrow {\scriptstyle G(f)_*} \\
\mathcal{D}(F(c),d) & \xrightarrow[\gamma_{c,d}]{} & \mathcal{C}(c,G(d)).
\end{array}$$

$\square$

Corollary 4.19. *Suppose that η is the unit of an adjunction $\gamma\colon F \dashv G$, where $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$. Then the arrow $\eta_c\colon c \to GF(c)$ is universal from c to G, for all $c \in \mathcal{C}$.*

Proof. The corollary states that, for every arrow $f: c \to G(d)$ in $\mathcal{C}$, for $d \in \mathcal{D}$, there exists a unique arrow $\bar{f}: F(c) \to d$ in $\mathcal{D}$ such that $G(\bar{f}) \cdot \eta_c = f$, that is, such that the triangle below



commutes. By Proposition 4.17 we may take $\bar{f} = \gamma^{-1}_{c,d}(f)$. □

Example 4.20. Suppose that $F: \mathbf{Set} \to \mathbf{Mon}$ is the free monoid functor and that $U: \mathbf{Mon} \to \mathbf{Set}$ is the underlying set functor. The unit η of the adjunction $F \dashv U$ has components $\eta_S = i_S$, where $i_S: S \to UF(S)$ is given for all sets S by $s \mapsto (s)$. Indeed the adjugant isomorphism (4.2) was defined in terms of the unit $\eta = i$, by Equation 4.18.

We now examine the dual notion, the *counit* of an adjunction. Suppose, as above, that $F: \mathcal{C} \to \mathcal{D}$ and $G: \mathcal{D} \to \mathcal{C}$ are adjoint, with $\gamma: F \dashv G$. Choosing $c = G(d)$ in Equation 4.4, we obtain a bijection

$$\gamma_{G(d),d}: \mathcal{D}(FG(d), d) \to \mathcal{C}(G(d), G(d)),$$

for each $d \in \mathcal{D}$.

Proposition 4.21. *The arrows $\epsilon_d: FG(d) \to d$ defined for all $d \in \mathcal{D}$ by*

$$\epsilon_d = \gamma^{-1}_{G(d),d}(1_{G(d)}), \tag{4.22}$$

are the components of a natural transformation $\epsilon: FG \Rightarrow 1_{\mathcal{D}}$.

Proof. The proof is dual to that of Proposition 4.12. □

Definition 4.23. The natural transformation $\epsilon: FG \Rightarrow 1_{\mathcal{D}}$ defined by Equation 4.22 is the *counit* of the adjunction $\gamma: F \dashv G$.

Proposition 4.24. *Suppose that ϵ is the counit of an adjunction $\gamma: F \dashv G$, where $F: \mathcal{C} \to \mathcal{D}$ and $G: \mathcal{D} \to \mathcal{C}$. Then, for all $c \in \mathcal{C}$ and $d \in \mathcal{D}$, the inverse of the bijection $\gamma_{c,d}: \mathcal{D}(F(c), d) \to \mathcal{C}(c, G(d))$ is given by*

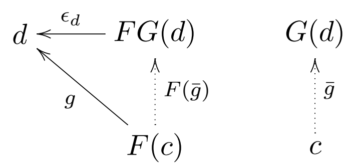
$$\gamma^{-1}_{c,d}(g) = \epsilon_d \cdot F(g), \tag{4.25}$$

for all arrows $g: c \to G(d)$ in $\mathcal{C}$.

Proof. Dual to the proof of Proposition 4.17. □

Corollary 4.26. *Suppose that ϵ is the counit of an adjunction $\gamma: F \dashv G$, where $F: \mathcal{C} \to \mathcal{D}$ and $G: \mathcal{D} \to \mathcal{C}$. Then the arrow $\epsilon_d: FG(d) \to d$ is universal from F to d, for all $d \in \mathcal{D}$.*

Proof. The corollary states that, for every arrow $g\colon F(c) \to d$ in $\mathcal{D}$, for $c \in \mathcal{C}$, there exists a unique arrow $\bar{g}\colon c \to G(d)$ in $\mathcal{C}$ such that $\epsilon_d \cdot F(\bar{g}) = g$, that is, such that the triangle below



commutes. By Proposition 4.24 we may take $\bar{g} = \gamma^{-1}_{c,d}(g)$. ∎

Example 4.27. Once again let's consider the adjunction $F \dashv U$, where $F\colon \mathbf{Set} \to \mathbf{Mon}$ and $U\colon \mathbf{Mon} \to \mathbf{Set}$ are the free monoid and underlying set functors. For any monoid M, the component ϵ_M of the counit is the homomorphism $FU(M) \to M$ that takes a sequence $(m_1, \ldots, m_k)$ of elements of M to their product $m_1 \cdots m_k$ in M. Let's use Equation 4.25, to describe the bijection $\gamma^{-1}_{S,M}\colon \mathbf{Set}(S, U(M)) \to \mathbf{Mon}(F(S), M)$ in terms of ϵ: for $g\colon S \to U(M)$ and $(s_1, \ldots, s_k) \in F(S)$, we have

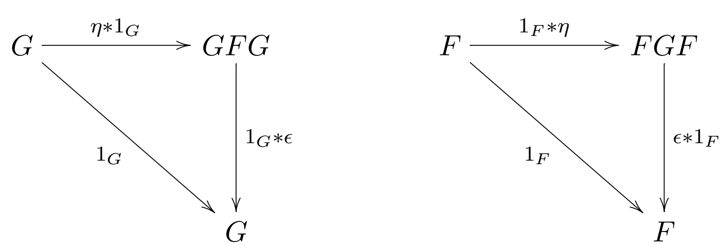
$$\begin{aligned}\gamma^{-1}_{S,M}(g)(s_1, \ldots, s_k) &= \epsilon_M \cdot F(g)(s_1, \ldots, s_k)\\ &= \epsilon_M(g(s_1), \ldots, g(s_k))\\ &= g(s_1) \cdots g(s_k),\end{aligned}$$

which is the usual 'extension of g to $F(S)$'.

Proposition 4.28. *The unit $\eta\colon 1_{\mathcal{C}} \Rightarrow GF$ and counit $\epsilon\colon FG \Rightarrow 1_{\mathcal{D}}$ of an adjunction $\gamma\colon F \dashv G$ satisfy the identities*

$$(4.29)\qquad (1_G * \epsilon) \cdot (\eta * 1_G) = 1_G \quad \textit{and} \quad (\epsilon * 1_F) \cdot (1_F * \eta) = 1_F,$$

that is, the triangles



commute in $\mathbf{Cat}(\mathcal{D}, \mathcal{C})$ and $\mathbf{Cat}(\mathcal{C}, \mathcal{D})$, respectively.

Proof. By the definition of ϵ, Equation 4.18 and Equation 3.20, we have

$$\begin{aligned}(1_G)_d = 1_{G(d)} &= \gamma_{G(d),d}(\epsilon_d)\\ &= G(\epsilon_d) \cdot \eta_{G(d)}\\ &= (1_G * \epsilon)_d \cdot (\eta * 1_G)_d,\end{aligned}$$

for all $d \in \mathcal{D}$, and so $1_G = (1_G * \epsilon) \cdot (\eta * 1_G)$ On the other hand, the definition of η, Equation 4.25 and Equation 3.20 give us

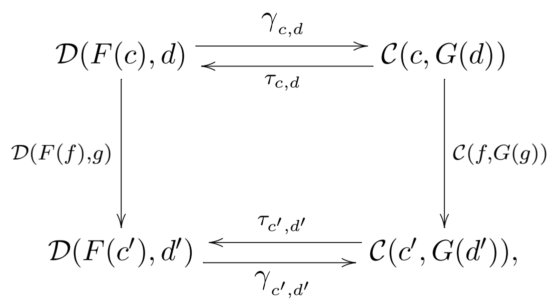
$$\begin{aligned}(1_F)_c = 1_{F(c)} &= \gamma^{-1}_{c,F(c)}(\eta_c) \\ &= \epsilon_{F(c)} \cdot F(\eta_c) \\ &= (\epsilon * 1_F)_c \cdot (1_F * \eta)_c,\end{aligned}$$

for all $c \in \mathcal{C}$, and hence $1_F = (\epsilon * 1_F) \cdot (1_F * \eta)$. □

The converse of Proposition 4.28 is also true; any pair of natural transformations η and ϵ, as above, that satisfies the triangle identities (4.29) consists of the unit and counit of a unique adjunction $\gamma \colon F \dashv G$.

Proposition 4.30. *Suppose that $F \colon \mathcal{C} \to \mathcal{D}$ and $G \colon \mathcal{D} \to \mathcal{C}$ are functors, and that $\eta \colon 1_\mathcal{C} \Rightarrow GF$ and $\epsilon \colon FG \Rightarrow 1_\mathcal{D}$ are natural transformations satisfying the triangle identies (4.29). Then (F, G) is an adjoint pair and the maps $\gamma_{c,d} \colon \mathcal{D}(F(c), d) \to \mathcal{C}(c, G(D))$, defined for all $c \in \mathcal{C}$ and $d \in \mathcal{D}$ by $f \mapsto G(f) \cdot \eta_c$, are the components of the adjugant isomorphism $\gamma \colon F \dashv G$. The inverse isomorphism $\tau = \gamma^{-1}$ is determined by $\tau_{c,d}(g) = \epsilon_d \cdot F(g)$, for all $g \colon c \to G(d)$.*

Proof. First we show that γ and τ defined as above are natural; then we show that they are mutually inverse. In other words, we will show that, for all $f \colon c' \to c$ in $\mathcal{C}$ and $g \colon d \to d'$ in $\mathcal{D}$, the diagram



(with an invisible identity arrows at each corner) commutes.

For any $h \in \mathcal{D}(F(c), d)$, we have

$$\begin{aligned}\mathcal{C}(f, G(g)) \cdot \gamma_{c,d}(h) &= \mathcal{C}(f, G(g)) \cdot G(h) \cdot \eta_c \\ &= G(g) \cdot G(h) \cdot \eta_c \cdot f.\end{aligned}$$

On the other hand,

$$\begin{aligned}\gamma_{c',d'} \cdot \mathcal{D}(F(f), g)(h) &= \gamma_{c',d'}(g \cdot h \cdot F(f)) \\ &= G(g \cdot h \cdot F(f)) \cdot \eta_{c'} \\ &= G(g) \cdot G(h) \cdot GF(f) \cdot \eta_{c'},\end{aligned}$$

which is equal to $G(g) \cdot G(h) \cdot \eta_c \cdot f$, by naturality (4.14) of η. Therefore γ is natural.

Now suppose that $k \in \mathcal{C}(c, G(d))$; then

$$\begin{aligned}\mathcal{D}(F(f), g) \cdot \tau_{c,d}(k) &= \mathcal{D}(F(f), g) \cdot \epsilon_d \cdot F(k) \\ &= g \cdot \epsilon_d \cdot F(k) \cdot F(f),\end{aligned}$$

while

$$\begin{aligned}\tau_{c',d'} \cdot \mathcal{C}(f, G(g))(k) &= \tau_{c',d'}(G(g) \cdot k \cdot f) \\ &= \epsilon_{d'} \cdot F(G(g) \cdot k \cdot f) \\ &= \epsilon_{d'} \cdot FG(g) \cdot F(k) \cdot F(f) \\ &= g \cdot \epsilon_d \cdot F(k) \cdot F(f),\end{aligned}$$

where the last equality is by naturality of ϵ. Hence τ is natural. Now

$$\begin{aligned}\tau_{c,d} \cdot \gamma_{c,d}(h) &= \tau_{c,d}(G(h) \cdot \eta_c) \\ &= \epsilon_d \cdot F(G(h) \cdot \eta_c) \\ &= \epsilon_d \cdot FG(h) \cdot F(\eta_c),\end{aligned}$$

and by naturality of ϵ, the square

$$\begin{array}{ccc} FGF(c) & \xrightarrow{\epsilon_{F(c)}} & F(c) \\ {\scriptstyle FG(h)}\downarrow & & \downarrow{\scriptstyle h} \\ FG(d) & \xrightarrow[\epsilon_d]{} & d \end{array}$$

commutes; hence,

$$\begin{aligned}\tau_{c,d} \cdot \gamma_{c,d}(h) &= h \cdot \epsilon_{F(c)} \cdot F(\eta_c) \\ &= h \cdot (\epsilon * 1_F)_c \cdot (1_F * \eta)_c,\end{aligned}$$

which is equal to h by the first triangle identity. Finally, we compute

$$\begin{aligned}\gamma_{c,d} \cdot \tau_{c,d}(k) &= \gamma_{c,d}(\epsilon_d \cdot F(k)) \\ &= G(\epsilon_d \cdot F(k)) \cdot \eta_c \\ &= G(\epsilon_d) \cdot GF(k) \cdot \eta_c \\ &= G(\epsilon_d) \cdot \eta_{G(d)} \cdot k \\ &= (1_G * \epsilon)_d \cdot (\eta * 1_G)_d \cdot k \\ &= k,\end{aligned}$$

where the fourth equality follows from naturality of η. Therefore τ and γ are mutually inverse. □

Proposition 4.31. *Suppose that $F\colon \mathcal{C} \to \mathcal{D}$ is an equivalence of categories, with weak inverse $G\colon \mathcal{D} \to \mathcal{C}$, so there exist natural isomorphisms $\eta\colon 1_{\mathcal{C}} \Rightarrow GF$ and $\epsilon\colon FG \Rightarrow 1_{\mathcal{D}}$, not necessarily satisfying the triangle identies. Then there exist adjunctions $\gamma\colon F \dashv G$ and $\beta\colon F \dashv G$ having unit η and counit ϵ, respectively. There exists an adjunction $G \dashv F$ having ϵ^{-1} as unit, and one having η^{-1} as counit.*

Proof. Define $\gamma_{c,d}\colon \mathcal{D}(F(c),d) \to \mathcal{C}(c, G(D))$, for each $c \in \mathcal{C}$ and $d \in \mathcal{D}$, as the composition

$$\mathcal{D}(F(c),d) \xrightarrow{G} \mathcal{C}(GF(c), G(d)) \xrightarrow{\eta_c^*} \mathcal{C}(c, G(d)),$$

so that $\gamma_{c,d}(f) = G(f)\cdot\eta_c$, for all $f\colon F(c) \to d$. The same argument as in the proof of Proposition 4.30 shows that the maps $\gamma_{c,d}$ are natural in c and d. They are bijections because G is full and faithful and η is an isomorphism. Hence $\gamma\colon F \dashv G$. The proof of Proposition 4.17 shows that η is the unit of this adjunction.

Defining $\tau_{c,d}\colon \mathcal{D}(F(c),d) \to \mathcal{C}(c, G(d))$ by $g \mapsto \epsilon_d \cdot F(g)$, we see as above that the $\tau_{c,d}$ are the components of a natural isomorphism τ, and hence $\alpha = \tau^{-1}\colon F \dashv G$ has counit ϵ. The other statements follow by symmetry. □

We have now shown that the notion of adjunction between categories is a generalization of that of equivalence, so we may think of a left or right adjoint of a functor as a kind of inverse, weaker than a weak inverse. In a more precise sense, though, adjoint functors should be thought of as *duals* of one another. To see why, let's recall some facts from linear algebra. Suppose that V is a vector space over a field K and that $V^* = \mathbf{Vect}(V, K)$ is the dual space. There is a bilinear map $\epsilon\colon V^* \otimes V \to K$, called *evaluation*, given by $(f, x) \mapsto f(x)$. If V is finite-dimensional, there is also a somewhat lesser-known map $\eta\colon K \to V \otimes V^*$, called *coevaluation*. We give two descriptions of η, in increasing order of concreteness. First, η is obtained from the transpose, or *dual* map $\epsilon^*\colon K^* \to (V^* \otimes V)^*$, by composing with the natural isomorphisms

$$K \cong K^* \quad \text{and} \quad (V^* \otimes V)^* \cong V^{**} \otimes V^* \cong V \otimes V^*$$

(finite dimensionality is needed for the last two isomorphisms). For the second description, we use the isomorphism $\alpha\colon V \otimes V^* \to \mathbf{Vect}(V, V)$ given by $x \otimes f \mapsto f_x$, where $f_x(v) = f(v)x$, for all $v \in V$ (finite dimensionality is needed for this isomorphism). Then $\eta\colon K \to V \otimes V^*$ is defined by $r \mapsto r\alpha^{-1}(1_V)$; in other words, η is the linear map that takes 1 to the element of $V \otimes V^*$ corresponding to the identity map on V. If $\{e_i\}$ is a basis for V and $\{e^j\}$ is the dual basis, defined by $e^j(e_i) = \delta_{ij}$, for all i and j, then η satifies

$$\eta(r) = \sum_i r e_i \otimes e^i.$$

For any vector space V, there are natural isomorphisms $V \cong K \otimes V$ and $V \cong V \otimes K$ determined by $x \leftrightarrow 1 \otimes x$ and $x \leftrightarrow x \otimes 1$. Regarding these isomorphisms as identities, as is usually done, we claim that the equalities

$$(1_V \otimes \epsilon) \cdot (\eta \otimes 1_V) = 1_V \quad \text{and} \quad (\epsilon \otimes 1_{V^*}) \cdot (1_{V^*} \otimes \eta) = 1_{V^*}$$

hold. Look familiar? We verify the first of these by computing

$$\begin{aligned}(1_V \otimes \epsilon) \cdot (\eta \otimes 1_V)(x) &= (1_V \otimes \epsilon) \cdot (\eta \otimes 1_V)(1 \otimes x) \\ &= (1_V \otimes \epsilon) \left(\sum_i e_i \otimes e^i \otimes x \right) \\ &= \sum_i e_i \otimes e^i(x) \\ &= \sum_i e^i(x) \cdot e_i \\ &= x,\end{aligned}$$

for all $x \in V$. The second equation is checked similarly.

Like most analogies in category theory, the parallel between adjoint functors and duals of finite-dimensional vector spaces is in fact more than an analogy. Indeed, we saw in the discussion immediately following Example 2.25 that there is a context in which we can regard bimodules as arrows between rings, with tensor product as composition. From this point of view, the dual of a finite dimensional K-vector space (i.e., K-K-bimodule) V is precisely the adjoint of the arrow V.

4.3. **Examples of adjunctions.** As we observed in the previous section, the free monoid functor $F\colon \mathbf{Set} \to \mathbf{Mon}$ is a left adjoint of the underlying set, or *forgetful* functor $U\colon \mathbf{Mon} \to \mathbf{Set}$. Here is a list containing some more examples of forgetful functors and their left adjoints:

Forgetful functor U	Left adjoint F
${}_R\mathbf{Mod} \to \mathbf{Set}$	$S \mapsto R\{S\}$ = Free R-module, basis S
$\mathbf{Grp} \to \mathbf{Set}$	$S \mapsto F(S)$ = Free group generated by S
$\mathbf{Grp} \to \mathbf{Ab}$	$G \mapsto G/[G,G]$ = Commutator factor group
${}_R\mathbf{Mod} \to \mathbf{Ab}$	$A \mapsto R \otimes_{\mathbb{Z}} A$ = Extension of scalars to R
$\mathbf{ComRng} \to \mathbf{Set}$	$S \mapsto \mathbb{Z}[S]$ = Polynomial ring
$\mathbf{Rng} \to \mathbf{Mon}$	$M \mapsto \mathbb{Z}\{M\}$ = Monoid algebra of M over $\mathbb{Z}$
$\mathbf{Alg}_R \to \mathbf{Mon}$	$M \mapsto R\{M\}$ = Monoid algebra of M over R
$\mathbf{Rng} \to \mathbf{Set}$	$S \mapsto \mathbb{Z}\{S^*\}$ = Free ring on S
$\mathbf{Alg}_R \to \mathbf{Set}$	$S \mapsto R\{S^*\}$ = Free R-algebra on S
$\mathbf{Alg}_R \to \mathbf{Mod}_R$	$M \mapsto T(M)$ = Tensor algebra of M over R
$\mathbf{Pos} \to \mathbf{Set}$	$S \mapsto \{S, \text{with trivial ordering}\}$
$\mathbf{Top} \to \mathbf{Set}$	$S \mapsto \{S, \text{with indiscrete topology}\}$
$\mathbf{Fld} \to \mathbf{IntDom}_m$	$D \mapsto$ Field of fractions of D

Remarks:

(i) By $\mathbf{IntDom}_m$ we mean the category whose objects are integral domains and arrows given by *injective* ring homomorphisms. The forgetful functor from **Fld** to the category of integral domains and all ring homomorphisms has no left adjoint.

(ii) When we speak of the category $\mathbf{Alg}_R$ of R-algebras, we assume that the ring R is commutative.

(iii) For R a commutative ring and M a monoid, the monoid algebra $R\{M\}$ is equal to the free R-module on M, equipped with unique multiplication extending that of M; in other words,

$$(\sum_i r_i \cdot m_i)(\sum_j s_j \cdot n_j) = \sum_{i,j} r_i s_j \cdot m_i n_j,$$

for $r_i, s_j \in R$ and $m_i, n_j \in M$.

(iv) In all these examples except the third and fourth, the unit map η_S is the natural injection. In the third example η_G is the natural projection, and in the fourth, η_A is given by $x \mapsto 1 \otimes x$.

Example 4.32. The forgetful functor $\mathbf{Top} \to \mathbf{Set}$ also has a right adjoint; namely the functor that puts the indiscrete topology on a set.

Example 4.33. Fix some set A. Then for all sets S and T, there is a bijection

$$\gamma_{S,T} \colon \mathbf{Set}(A \times S, T) \to \mathbf{Set}(S, \mathbf{Set}(A, T)),$$

that takes a function $f \colon A \times S \to T$ to the function $\bar{f} \colon S \to \mathbf{Set}(A, T)$, defined by $\bar{f}(s) = f(-\,, s)$, for all $s \in S$. In other words, for all $s \in S$,

$$\gamma_{S,T}(f)(s)(a) = f(a, s),$$

for all $a \in A$. These bijections are natural in S and T, and thus $A \times -$ and $\mathbf{Set}(A, -)$ are an adjoint pair of functors from **Set** to itself, with $\mathbf{Set}(A, -)$ right adjoint to $A \times -$. If we denote the set of functions $\mathbf{Set}(X, Y)$ by X^Y, for all sets X and Y, the above bijection can be written as the *exponential formula*

$$T^{A \times S} \cong (T^A)^S,$$

for all sets A and S and T. The unit η has components $\eta_S \colon S \to (A \times S)^A$, satisfying $\eta_S(s)(a) = (a, s)$, for all $s \in S$ and $a \in A$. The component $\epsilon_T \colon A \times T^A \to T$ of the counit is the evaluation map, given by $(a, f) \mapsto f(a)$.

Example 4.34. Suppose that R, S and T are rings, and that $P \in {}_R\mathbf{Mod}_S$, $M \in {}_S\mathbf{Mod}_T$ and $N \in {}_R\mathbf{Mod}_T$ are bimodules. The abelian group $P \otimes_S M$ has an R-T-bimodule structure, with left and right scalar multiplication given by

$$r(x \otimes y) = (rx) \otimes y \quad \text{and} \quad (x \otimes y)t = x \otimes (yt),$$

for all $x \otimes y \in P \otimes_S M$, $r \in R$ and $t \in T$. Also, the abelian group ${}_R\mathbf{Mod}(P, N)$ has an S-T-bimodule structure, with sf and ft defined, for all $f \in {}_R\mathbf{Mod}(P, N)$, $s \in S$ and $t \in T$, by

$$(sf)(x) = f(xs) \quad \text{and} \quad (ft)(x) = (f(x))t,$$

for all $x \in P$. Hence we have functors

$$P \otimes_S - : {}_S\mathbf{Mod}_T \to {}_R\mathbf{Mod}_T \quad \text{and} \quad {}_R\mathbf{Mod}(P, -)\colon {}_R\mathbf{Mod}_T \to {}_S\mathbf{Mod}_T$$

The defining property of tensor product implies that there is a bijection (in fact, an isomorphism of abelian groups)

$${}_R\mathbf{Mod}_T\,(P \otimes_S M, N) \to {}_S\mathbf{Mod}_T\,(M, {}_R\mathbf{Mod}(P, N)),$$

determined by $f \mapsto \bar{f}$, where $\bar{f}(m) = f(- \otimes m)$, for all $m \in M$. The inverse bijection is given by $g \mapsto g'$, where $g'(p \otimes m) = g(m)(p)$, for all $p \in P$ and $m \in M$. These bijections are natural; hence $P \otimes_S -$ is left adjoint to ${}_R\mathbf{Mod}(P, -)$. For all N, the counit map

$$\epsilon_N\colon P \otimes {}_R\mathbf{Mod}(P, N) \to N$$

is the evaluation homomorphism $p \otimes f \mapsto f(p)$.

There are two special cases worth mentioning. First, suppose that $R = T = \mathbb{Z}$, so that ${}_R\mathbf{Mod}_S = \mathbf{Mod}_S$, ${}_S\mathbf{Mod}_T = {}_S\mathbf{Mod}$ and ${}_R\mathbf{Mod}_T = \mathbf{Ab}$, and P is a right S-module. Then we have adjoint functors

$$P \otimes_S - : {}_S\mathbf{Mod} \to \mathbf{Ab} \quad \text{and} \quad \mathbf{Ab}(P, -)\colon \mathbf{Ab} \to {}_S\mathbf{Mod}.$$

The second special case is when $R = S = T$ is commutative and P is a left (or, equivalently, right) R-module. Then $P \otimes_R -$ and ${}_R\mathbf{Mod}(P, -)$ constitute an adjoint pair of functors from ${}_R\mathbf{Mod}$ to itself.

E-mail address: wschmitt@gwu.edu