Category theory in context

Emily Riehl

The aim of theory really is, to a great extent, that of systematically organizing past experience in such a way that the next generation, our students and their students and so on, will be able to absorb the essential aspects in as painless a way as possible, and this is the only way in which you can go on cumulatively building up any kind of scientific activity without eventually coming to a dead end.

M.F. Atiyah, “How research is carried out”

Contents

Preface 1
Preview 2
Notational conventions 2
Acknowledgments 2

Chapter 1. Categories, Functors, Natural Transformations 5
1.1. Abstract and concrete categories 6
1.2. Duality 11
1.3. Functoriality 14
1.4. Naturality 20
1.5. Equivalence of categories 25
1.6. The art of the diagram chase 32

Chapter 2. Representability and the Yoneda lemma 43
2.1. Representable functors 43
2.2. The Yoneda lemma 46
2.3. Universal properties 52
2.4. The category of elements 55

Chapter 3. Limits and Colimits 61
3.1. Limits and colimits as universal cones 61
3.2. Limits in the category of sets 67
3.3. The representable nature of limits and colimits 71
3.4. Examples 75
3.5. Limits and colimits and diagram categories 79
3.6. Warnings 81
3.7. Size matters 81
3.8. Interactions between limits and colimits 82

Chapter 4. Adjunctions 85
4.1. Adjoint functors 85
4.2. The unit and counit as universal arrows 90
4.3. Formal facts about adjunctions 93
4.4. Adjunctions, limits, and colimits 96
4.5. Existence of adjoint functors 100

Chapter 5. Monads and their Algebras 107
5.1. Monads from adjunctions 107
5.2. Adjunctions from monads 111
5.3. Free algebras and canonical presentations 117
5.4. Recognizing categories of algebras 121

5.5. Limits and colimits in categories of algebras 123

Chapter 6. All Concepts are Kan Extensions 131
6.1. Kan extensions 131
6.2. A formula for Kan extensions 134
6.3. Pointwise Kan extensions 136
6.4. All concepts 138

Epilogue: Theorems in category theory 141
E.1. Theorems in basic category theory 141
E.2. Coherence for monoidal categories 142
E.3. The universal property of the unit interval 144
E.4. A characterization of Grothendieck toposes 145
E.5. Embeddings of abelian categories 145

Bibliography 147

Glossary of Notation 149

Index 151

Preface

Atiyah described mathematics as the "science of analogy"; in this vein, the purview of category theory is *mathematical analogy*. Specifically, category theory provides a mathematical language that can be deployed to describe phenomena in any mathematical context. Perhaps surprisingly given this level of generality, these concepts are neither meaningless and nor in many cases so clearly visible prior to their advent. In part, this is accomplished by a subtle shift in perspective. Rather than characterize mathematical objects directly, the categorical approach emphasizes the morphisms, which give comparisons between objects of the same type. Structures associated to particular objects can frequently be characterized by their *universal properties*, i.e., by the existence of certain canonical morphisms to or from other objects of a similar form.

A great variety of constructions can be described in this way: products, kernels, and quotients for instance are all *limits* or *colimits* of a particular shape, a characterization that emphasizes the universal property associated to each construction. Tensor products, free objects, and localizations are also uniquely characterized by universal properties in appropriate categories. Important technical differences between particular sorts of mathematical objects can be described by the distinctive properties of their categories: that rings have all limits and colimits while fields have few, that certain classes maps are *monomorphisms* or *epimorphisms*. Constructions that take one type of mathematical object to objects of another type are often morphisms between categories, called *functors*. In contrast with earlier numerical invariants in topology, functorial invariants (the fundamental group, homology) tend both to be more easily computable and also provide more precise information. Functors can then be said to *preserve* particular categorical structures, or not. Of particular interest is when a functor describes an *equivalence* of categories, which means that objects of the one sort can be translated into and reconstructed from objects of another sort.

Category theory also contributes new proof techniques, such as *diagram chasing* or duality; Steenrod called these methods "abstract nonsense."[1] The aim of this course will be to introduce the language, philosophy, and basic theorems of category theory. A complementary objective will be to put this theory into practice: studying functoriality in algebraic topology, naturality in group theory, and universal properties in algebra.

Practitioners often assert that the hard part of category theory is to state the correct definitions. Once these are established and the categorical style of argument is sufficiently internalized, proving the theorems tends to be relatively easy.[2] The relative simplicity of the proofs of major theorems occasionally leads detractors to assert that there are no theorems in category theory. This is not at all the case! Counterexamples abound in the text that

[1]Contrary to popular belief, this was not intended as an epithet.

[2]A famous exercise in Serge Lang's *Algebra* asks the reader to "Take any book on homological algebra, and prove all the theorems without looking at the proofs given in that book" [**Lan84**, p. 175]. Homological algebra is the subject whose development induced Eilenberg and Mac Lane to introduce the general notions of category, functor, and natural transformation.

follows. A list of further major theorems, beyond the scope of this course, will appear eventually in an epilogue.

Preview

It is difficult to preview the main theorems in category theory because we are not yet fluent in the language needed to state them. Instead, here are a few corollaries, results in other areas of mathematics that follow trivially, as special cases of general categorical results.

Corollary. *In a path connected space, any choice of basepoint yields an isomorphic fundamental group.*

Corollary. *Every row operation on matrices with k rows is defined by left multiplication by some $k \times k$ matrix.*

Corollary. *For any pair of sets X and Y and any function $f\colon X \times Y \to \mathbb{R}$*

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) \leq \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

The following three results are corollaries of the same theorem:

Corollary. *For any function $f\colon A \to B$, the inverse image $f^{-1}\colon P(B) \to P(A)$, a function between the powersets of A and B, preserves both unions and intersections, while the direct image $f_*\colon P(A) \to P(B)$ only preserves unions.*

Corollary. *For any vector spaces U, V, W, $U \otimes (V \oplus W) \cong (U \otimes V) \oplus (U \otimes W)$.*

Corollary. *The free group on the set $X \coprod Y$ is the free product of the free groups on the sets X and Y.*

Corollary. *Any bijective continuous function between compact Hausdorff spaces is a homeomorphism.*

Notational conventions

An arrow symbol "→", either in a display or in text, will only ever be used to denote a morphism in the appropriate category. The symbol "↦", read as "maps to," will appear occasionally when defining a function between sets by specifying its action on particular elements. The symbol "⇝" will be used in a less technical sense to mean something along the lines of "yields" or "leads to" or "can be used to construct." We use "⇉" as an abbreviation for a **parallel pair** of morphisms, i.e., for a pair of morphisms with common domain and codomain.

One usage of italics will be to highlight technical terms; boldface will signal that a technical term is being defined by its surrounding text.

Acknowledgments

I have been consulting many sources while preparing these notes; a complete bibliography will appear here eventually.

Thanks to the following people who responded to a call for examples on the *n*-Category Café and the categories mailing list: David Ellerman, Kimmo Rosenthal, Fred Linton, John Terilla, Tyler Bryson, Tom LaGatta, Samuel Dean, Ronnie Brown, Enrico Vitale, Aaron Mazel-Gee, Anders Kock, Mark Johnson, David White, Richard Garner, Ross Street, Paul Levy, Peter Smith, Qiaochu Yuan, Mike Shulman, Tim Campion, John Baez, Todd Trimble, Gejza Jenča, Graham White, Adrian Clough, Jesse McKeown, Mozibur Ullah, Josh

Drum, Tom Ellis, Arnaud Spiwack, Sameer Gupta, Arnaud Spiwack, Yemon Choi, Martin Brandenburg.

In particular, Anders Kock suggested a more general formulation of “the chain rule expresses the functoriality of the derivative” than appears in Example 1.3.2.(viii). The expression of the fundamental theorem of Galois theorem as an isomorphism of categories that appears as Example 1.3.13 is a favorite exercise of Peter May. Peter Haine suggested Example 1.4.6, expressing the Riesz representation theorem as a natural isomorphism of Banach space valued functors. John Baez reminded me that the groupoid of finite sets is a categorification of the natural numbers, providing a suitable framework in which to prove certain basic equations in elementary arithmetic; see Example 1.4.9 and Corollary 4.4.6. Juan Climent Vidal suggested using the axiom of regularity to define the non-trivial part of the equivalence of categories presented in Example 1.5.4. Samuel Dean suggested Corollary 1.5.10, Fred Linton suggested Corollary 2.2.7, and Ronnie Brown suggested Example 3.4.4. Peter Haine formulated Example 3.8.4. I learned of the non-natural pointwise isomorphism appearing in Example 3.5.5 from Martin Brandenburg. He also suggested the description of the real exponential function as a Kan extension; see Example 6.2.4. Mozibur Ullah suggested Exercise 3.4.3. Andrew Putman pointed out that Lang’s *Algebra* constructs the free group on a set using the construction of the General Adjoint Functor Theorem (Example 4.5.5). Paul Levy suggested using affine spaces to motivate the category of algebras over a monad, as discussed in Section 5.2; a similar example was suggested by Enrico Vitale. Marina Lehner, while writing her undergraduate senior thesis, showed me that an entirely satisfactory account of Kan extensions can be given without the calculus of ends and coends. I have enjoyed and this book has been enriched by several years of impromptu categorical conversations with Omar Antolín Camarena. Finally, I am very grateful for careful readings undertaken by Peter Smith and Darij Grinberg, who each sent detailed lists of corrections.

I would also like to thank the Department of Mathematics at Harvard University for giving me the opportunity to teach this course and my colleagues there, who have created an extremely pleasant working environment. I am grateful for financial support from the National Science Foundation Division of Mathematical Sciences DMS-1509016.

CHAPTER 1

Categories, Functors, Natural Transformations

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

$$0 \to G \to E \to H \to 0.$$[1]

A pair of group extensions E and E' of G and H are considered to be equivalent whenever there is an isomorphism $E \cong E'$ that *commutes with* the inclusions of G and quotient maps to H, in a sense that we will make precise in §1.6. When we restrict consideration to extensions E that are abelian groups, the set of equivalence classes of extensions of H by G defines an abelian group $\mathrm{Ext}(H, G)$.

In 1941, Saunders Mac Lane gave a lecture at the University of Chicago in which he computed that $\mathrm{Ext}(G_p, \mathbb{Z}) = \mathbb{Z}_p$, the group of p-adic integers, where G_p is the abelian group generated by elements $g_1, g_2, \ldots$ so that $pg_{n+1} = g_n$, for p a fixed prime. When he explained this result to Samuel Eilenberg, who had missed the lecture, Eilenberg recognized the calculation as the homology of the p-adic solenoid, a space formed as the infinite intersection of a sequence of solid tori, each wound around p times inside the preceding torus. In teasing apart this connection, the pair of them discovered what is now known as the **universal coefficient theorem** in algebraic topology, which relates the *homology* H_* and *cohomology groups* H^* associated to a space X via a group extension [**ML05**]:

$$0 \to \mathrm{Ext}(H_{n-1}(X), G) \to H^n(X, G) \to \mathrm{Hom}(H_n(X), G) \to 0. \tag{1.0.1}$$

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

The adjective "natural" had been used colloquially by mathematicians to mean something on the lines of "defined without respect to particular choices." For instance, to define an isomorphism between a finite dimensional vector space V and its *dual*, the vector space of linear maps from V to the ground field $\Bbbk$, requires a choice of basis. However, there is an isomorphism between V and its double dual that requires no choice of basis; the latter, but not the former, is *natural*.

[1] The zeros appearing on the ends provide no additional data. Instead, the first zero implicitly asserts that the map $G \to E$ is an inclusion and the second that the map $E \to H$ is a surjection. More precisely, the displayed sequence of group homomorphisms is *exact*, meaning that the kernel of each homomorphism equals the image of the preceeding homomorphism.

To give a rigorous proof that their particular natural isomorphisms extended to inverse and direct limits, Eilenberg and Mac Lane sought to give a mathematically precise definition of the informal concept of "naturality." To that end, they introduced the notion of a *natural transformation*, a parallel collection of homomorphisms between abelian groups in this instance. To characterize the source and target of a natural transformation, they introduced the notion of a *functor*.[2] And to define the source and target of a functor in the greatest generality, they introduced the concept of a *category*. This work, described in "The general theory of natural equivalences" [**EM45**], published in 1945, marked the birth of category theory.

While categories and functors were first conceived as auxiliary notions, needed to give a precise meaning to the concept of naturality, they have grown into interesting and important concepts in their own right. Categories suggest a particular perspective to be used in the study of mathematical objects, which we will gradually explore. Functors, which translate mathematical objects of one type into object of another, have a more immediate utility. For instance, the Brouwer fixed point theorem translates a seemingly intractable problem in topology to a trivial one ($0 \neq 1$) in algebra. It is to these topics that we now turn.

The axioms defining a category, introduced in §1.1, when suitably interpretted, are self-dual. This duality, in analogy with the duality in projective plane geometry, can be formulated precisely in first-order logic. Thus, as we describe in §1.2, for any proof of a theorem about all categories from these axioms, there is a dual proof of the dual theorem obtained by a syntactic process that is interpreted as "turning around all the arrows."

Functors and natural transformations are introduced in §1.3 and §1.4 with examples intended to shed light on the linguistic and practical utility of these concepts. The category-theoretic notions of *isomorphism*, *monomorphism*, and *epimorphism* are invariant under certain classes of functors, including in particular the *equivalences of categories*, introduced in §1.5. At a high level, an equivalence of categories provides a precise expression of the intuition that one type of mathematical objects are "the same as" objects of another variety.

At its origins, category theory, in addition to providing a new language to describe emerging mathematical phenomena, also introduced a new proof technique: that of the diagram chase. The introduction to the influential book [**ES52**] introduces the notion of *commutative diagram*, a new technique of proof appropriate for their axiomatic treatment of homology theory. §1.6 provides an introduction to the art of the diagram chase. These techniques are put to use to define the vertical and horizontal composition operations for natural transformations.

1.1. Abstract and concrete categories

> It frames a possible template for any mathematical theory: the theory should have *nouns* and *verbs*, i.e., objects, and morphisms, and there should be an explicit notion of composition related to the morphisms; the theory should, in brief, be packaged by a category.
>
> Barry Mazur, "When is one thing equal to some other thing?"

[2]A brief account of functors and natural isomorphisms in group theory appeared in a 1942 paper [**EM42b**].

DEFINITION 1.1.1. A **category** consists of

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

so that:

- Each morphism has a specified **domain** and **codomain** among the collection of objects. The notation $f\colon X \to Y$ signifies that f is a morphism with domain X and codomain Y.
- Each object has a designated **identity morphism** $1_X\colon X \to X$.
- For any **composable** pair of morphisms, i.e., for any pair f, g with the codomain of f equal to the domain of g, there exists a specified **composite morphism**[3] gf whose domain is equal to the domain of f and whose codomain is equal to the codomain of g, i.e.:

$$f\colon X \to Y, \quad g\colon Y \to Z \qquad \rightsquigarrow \qquad gf\colon X \to Z.$$

This data is subject to the following two axioms:

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

$$f\colon X \to Y, \quad g\colon Y \to Z, \quad h\colon Z \to W \qquad \rightsquigarrow \qquad hgf\colon X \to W.$$

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

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

It is traditional to name a category after its objects; in most cases, the appropriate accompanying choice of structure-preserving morphisms is clear. However, this practice is somewhat contrary to the basic philosophy of category theory: that mathematical objects should always be considered in tandem with the morphisms between them. As we have seen, the category can be recovered from the algebra of morphisms so of the two, the objects and morphisms, the morphisms take primacy.

EXAMPLES 1.1.2. Many familiar varieties of mathematical objects assemble into a category.

(i) Set has sets as its objects and functions, with specified domain and codomain,[4] as its morphisms.
(ii) Top has topological spaces as its objects and continuous functions as its morphisms.
(iii) Set_* and Top_* have sets or spaces with a specified basepoint[5] as objects and basepoint-preserving (continuous) functions as morphisms.

---

[3]The composite may also be written less concisely as $g \cdot f$ when this adds typographical clarity.

[4][**EM45**, p 239] emphasizes that the data of a function should include specified sets of inputs and potential outputs, a perspective that was somewhat radical at the time.

[5]A **basepoint** is simply a chosen distinguished point in the set or space.

(iv) Group has groups as objects and group homomorphisms as morphisms. This example lent the general term "morphisms" to the data of an abstract category. The categories Ring of rings and ring homomorphisms and Field of fields and field homomorphisms are defined similarly.

(v) For a fixed ring R, Mod_R is the category of R-modules and R-module homomorphisms. We typically write $\mathsf{Vect}_{\Bbbk}$ when the ring happens to be a field $\Bbbk$, and write Ab for $\mathsf{Mod}_{\mathbb{Z}}$, as a $\mathbb{Z}$-module is precisely an abelian group.

(vi) Graph has graphs as objects and graph morphisms (functions carrying vertices to vertices and edges to edges, preserving incidence relations) as morphisms. We will also be interested in the variant DirGraph whose objects are directed graphs, whose edges are now depicted as arrows, and whose morphisms are directed graph morphisms, which must preserve sources and targets.

(vii) Man has smooth (i.e., infinitely differentiable) manifolds as objects and smooth functions as morphisms.

(viii) Poset has partially-ordered sets as objects and order-preserving functions as morphisms.

(ix) For any *first-order language* $\mathcal{L}$, specifying constant, function, and relation symbols, there is a category $\mathsf{Model}_{\mathcal{L}}$ whose objects are *models of* $\mathcal{L}$, sets equipped with appropriate constants, relations, and functions. Morphisms are functions that preserve the specified constants, relations, and functions, in the usual sense. Special cases include (iv), (v), (vi), and (viii).

The preceding are all examples of *concrete categories*, those whose objects have underlying sets and whose morphisms are functions between these underlying sets, subject to appropriate structure-preservation conditions; a more precise definition of a concrete category will be given in 1.6.19. However, not all categories have this form.

Examples 1.1.3.

(i) For $\Bbbk$ a field (or unital ring), write $\mathsf{Mat}_{\Bbbk}$ for the category whose objects are positive integers and in which the set of morphisms from n to m is the set of $m \times n$ matrices with values in $\Bbbk$. Composition is by matrix multiplication with identity matrices serving as the identity morphisms.

(ii) A group G (or more generally a monoid[6]) defines a category $\mathsf{B}G$ with a single object.[7] The group elements are its morphisms, with composition given by multiplication. The identity element acts as the identity morphism for the unique object in this category.

(iii) A poset P (or more generally a preorder[8]) may be regarded as a category. The elements of P are the objects of the category and there exists a unique morphism $x \to y$ if and only if $x \leq y$. Transitivity of the relation "$\leq$" implies that the required composite morphisms exist. Reflexivity implies that identity morphisms do.

---

[6]A **monoid** is a set M equipped with an associative binary multiplication operation $M \times M \xrightarrow{\cdot} M$ and an identity element $e \in M$ serving as a two-sided identity. In other words, a monoid is precisely a one-object category.

[7]The notation "$\mathsf{B}G$" comes from topology, as the category $\mathsf{B}G$ serves as a model for the *classifying space* of the group G.

[8]A **preorder** is a set with a binary relation $\leq$ that is reflexive and transitive. In other words, a preorder is precisely a category in which there are no parallel pairs of distinct morphisms between any fixed pair of objects. A **poset** is a preorder that is additionally antisymmetric: $x \leq y$ and $y \leq x$ implies that $x = y$.

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

$$0 \to 1 \to 2 \to 3 \to \cdots$$

in the sense that every non-identity morphism can be uniquely factored as a composite of morphisms in the displayed graph; a precise definition of this notion will be given in Example 4.1.12.

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

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

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

Thus, the philosophy of category theory is extended. The categories listed in Examples 1.1.2 suggest that mathematical objects ought to be considered together with the appropriate notion of morphism between them. The categories listed in Examples 1.1.3 illustrate that these morphisms are not always functions.[9] The morphisms in a category are also called **arrows** or **maps**, particularly in the contexts of the examples of 1.1.3 and 1.1.2, respectively.

Remark 1.1.4. Russell's paradox implies that there is no set whose elements are "all sets." This is the reason why we have used the vague word "collection" in Definition 1.1.1. Indeed, in each of the examples listed in 1.1.2, the collection of objects is not a set. Eilenberg and Mac Lane address this potential area of concern as follows:

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

The set-theoretical issues that confront us while defining the notion of a category will compound as we further develop category theory. For that reason, common practice among category theorists is to work in an extension of the usual Zermelo-Fraenkel axioms of set theory, with new axioms allowing one to distinguish between "small" and "large" sets, or

[9] Miles Reid's *Undergraduate algebraic geometry* emphasizes that the morphisms are not always functions, writing "Students who disapprove are recommended to give up at once and take a reading course in category theory instead" [**Rei88**, p 4].

between sets and classes. The search for the most useful set-theoretical foundations for category theory is a large and fascinating area of study that we will unfortunately not have time to explore.[10] For now, we will sweep these foundational issues under the rug, not because these issues are not serious or interesting, but because they distract from the task at hand.

For the reasons just discussed, it is important to introduce adjectives that explicitly address the size of a category. A category is **small** if it has only a set's worth of arrows; this implies that it has only a set's worth of objects. None of the categories in Example 1.1.2 are small, but they all satisfy a common useful set-theoretical property. A category is **locally small** if between any pair of objects there is only a set's worth of morphisms. In this case, it is traditional to write $\mathsf{C}(X, Y)$ or $\mathrm{Hom}(X, Y)$ for the set of morphisms from X to Y in a locally small category C.[11] The set of arrows between a pair of fixed objets in a locally small category is typically called a **hom-set**, whether or not it is a set of "homomorphisms" of any particular kind.

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

Definition 1.1.5. An **isomorphism** in a category is a morphism $f: X \to Y$ for which there exists a morphism $g: Y \to X$ so that $gf = 1_X$ and $fg = 1_Y$. The objects X and Y are **isomorphic** whenever there exists an isomorphism between X and Y, in which case one writes $X \cong Y$.

Whenever f and g are morphisms so that gf is an identity, we say that g is a **left inverse** to or a **retraction** of f and, equivalently, that f is a **right inverse** to or a **section** of g. A morphism can admit multiple left or right inverses, but if f is an isomorphism, its inverse isomorphism is unique; see Exercise 1.1.1.

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

Examples 1.1.6.

(i) The isomorphisms in Set are precisely the *bijections*.
(ii) The isomorphisms in Group, Ring, Field, or Mod_R are the bijective homomorphisms.
(iii) The isomorphisms in the category Top are the *homeomorphisms*, i.e., the continuous functions with continuous inverse, which is a stronger property than merely being a bijective continuous function.
(iv) The isomorphisms in the category Htpy are the *homotopy equivalences*.

Examples 1.1.6.(ii) and (iii) suggest the following general question: in a concrete category, when are the isomorphisms precisely those maps in the category that induce bijections between the underlying sets? We will see an answer in Lemma 5.5.3.

Definition 1.1.7. A **groupoid** is a category in which every morphism is an isomorphism.

Examples 1.1.8.

[10] The preprint [**Shu08**] gives an excellent summary, though it is perhaps better read after Chapters 1–4.

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

(i) A **group** is a groupoid with one object.[12]
(ii) For any space X, its **fundamental groupoid** $\Pi_1(X)$ is a category whose objects are the points of X and whose morphisms are basepoint-preserving homotopy classes of paths.

A **subcategory** D of a category C is defined by restricting to a subcollection of objects and subcollection of morphisms subject to the requirements that the subcategory D contains the domain and codomain of any morphism in D, the identity morphism of any object in D, and the composite of any composable pair of morphisms in D.

Example 1.1.9. Any category C has a **maximal subgroupoid** containing all of the objects and only those morphisms that are isomorphisms.

For instance, $\mathsf{Fin}_{\mathrm{iso}}$, the category of finite sets and bijections, is the maximal subgroupoid of the category Fin of finite sets and all functions. We will explore the relationship between this groupoid and the laws of elementary arithmetic in Example 1.4.9.

Exercises.

Exercise 1.1.1. Show that any left inverse and any right inverse to a common morphism necessarily coincide. Conclude that a morphism can have at most one inverse isomorphism.

Exercise 1.1.2. Let C be a category. Show that the collection of isomorphisms in C defines a subcategory, the **maximal groupoid** inside C.

1.2. Duality

> The dual of any axiom for a category is also an axiom . . . A simple metamathematical argument thus proves the *duality principle*. If any statement about a category is deducible from the axioms for a category, the dual statement is likely deducible.
>
> Saunders Mac Lane, "Duality for groups"

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

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

- the same objects as in C, and
- the same morphisms as in C. When necessary for notational clarity, we write f^{op} for the morphism in C^{op} corresponding to a morphism f in C.

The remaining structure of the category C^{op} is given as follows:

- The domain of f^{op} is defined to be the codomain of f and the codomain of f^{op} is defined to be the domain of f.
- For each object X, the arrow 1_X^{op} serves as its identity in C^{op}.

[12] This is not simply an example; it is a definition.

- To define composition, observe that a pair of morphisms f^{op}, g^{op} in C^{op} is composable precisely when the pair g, f is composable in C, i.e., precisely when the codomain of g equals the domain of f. We then define $g^{op} \cdot f^{op}$ to be $(f \cdot g)^{op}$.

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

Examples 1.2.2.

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

(ii) When a preorder $(P, \leq)$ is regarded as a category, its opposite category is the category which has a morphism $x \to y$ if and only if $y \leq x$. For example, ω^{op} is the category freely generated by the graph

$$\cdots \to 3 \to 2 \to 1 \to 0.$$

(iii) If G is a group, regarded as a one-object groupoid, the category G^{op} is again a one-object groupoid, and hence a group. This group is called the **opposite group** and is used to define right actions as a special case of left actions; see Example 1.3.6.(v) below.

(iv) The category $\Gamma = \mathsf{Fin}_*^{op}$ was described by Graeme Segal in [**Seg74**] as follows:

> Γ is the category whose objects are all finite sets, and whose morphisms from S to T are the maps $\theta \colon S \to P(T)$ such that $\theta(\alpha)$ and $\theta(\beta)$ are disjoint when $\alpha \neq \beta$. The composite of $\theta \colon S \to P(T)$ and $\phi \colon T \to P(U)$ is $\psi \colon S \to P(U)$, where $\psi(\alpha) = \bigcup_{\beta \in \theta(\alpha)} \phi(\beta)$.

This syntactical duality has a very important consequence for the development of category theory. Any theorem of the form "for all categories C" also necessarily applies to the opposites of these categories. Interpreting the result in the dual context leads to a **dual theorem**, proven by the dual of the original proof, in which the direction of each arrow appearing in the argument is reversed. The result is a two-for-one deal: any proof in category theory simultaneously proves two theorems, the original statement and its dual.[13]

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

Lemma 1.2.3. *The following are equivalent:*

(i) $f \colon x \to y$ *is an isomorphism in* C.

(ii) For all objects $c \in \mathsf{C}$*, post-composition with* f *defines an isomorphism*

$$f_* \colon \mathsf{C}(c, x) \to \mathsf{C}(c, y).$$

(iii) For all objects $c \in \mathsf{C}$*, pre-composition with* f *defines an isomorphism*

$$f^* \colon \mathsf{C}(y, c) \to \mathsf{C}(x, c).$$

[13] More generally, the proof of a statement of the form "for all categories $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_n$" leads to 2^n dual theorems. In practice, however, not all of the dual statements will differ meaningfully from the original.

In language that we will make precise in Chapter 2, Lemma 1.2.3 asserts that isomorphisms in a locally small category are defined *representably* in terms of isomorphisms in the category of sets.

Proof. We will prove the equivalence (i)⇔(ii) and conclude the equivalence (i)⇔(iii) by duality.

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

$$g_*\colon \mathsf{C}(c, y) \to \mathsf{C}(c, x)$$

to f_* in the sense that the composites $g_* f_*$ and $f_* g_*$ are both the identity function: for any $h\colon c \to x$ and $k\colon c \to y$, $g_* f_*(h) = gfh = h$ and $f_* g_*(k) = fgk = k$.

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

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

$$(f^{\mathrm{op}})_*\colon \mathsf{C}^{\mathrm{op}}(c, y) \to \mathsf{C}^{\mathrm{op}}(c, x) \text{ is an isomorphism for all } c \in \mathsf{C}^{\mathrm{op}}. \tag{1.2.4}$$

Interpreting the data of C^{op} in its opposite category C, the statement (1.2.4) expresses the same mathematical content as

$$f^*\colon \mathsf{C}(y, c) \to \mathsf{C}(x, c) \text{ is an isomorphism for all } c \in \mathsf{C}. \tag{1.2.5}$$

That is: $\mathsf{C}^{\mathrm{op}}(c, x) = \mathsf{C}(x, c)$, post-composition in C^{op} translates to pre-composition in the opposite category C. Similarly, $f^{\mathrm{op}}\colon y \to x$ is an isomorphism in C^{op} if and only if $f\colon x \to y$ is an isomorphism in C; the notion of isomorphism, as defined in 1.1.5, is self-dual. A similar translation, as just demonstrated between the statements (1.2.4) and (1.2.5), transforms the proof of (i)⇔(ii) into a proof of (i)⇔(iii). □

A nice exposition of the duality principle in category theory may be found in [**Awo10**, §3.1]. As we become more comfortable with arguing by duality, dual proofs and eventually also dual statements will rarely be described in this much detail.

Example 1.2.6. The real numbers with their usual linear order defines a poset $(\mathbb{R}, \le)$ and thus a category. Consider a subset $A \subset \mathbb{R}$ of objects. If $x \in \mathbb{R}$ satisfies

$$(\forall a \in A, \exists\, a \to x) \text{ and } (\forall y \in \mathbb{R}\, (\forall a \in A, \exists\, a \to y) \text{ implies } (\exists\, x \to y)),$$

then x is the **supremum** of A. Note that we have defined the supremum in a way that refers only to the structure of the category $(\mathbb{R}, \le)$: in plain English, the second clause asserts that "if y is an object so that for any object a in A there exists a morphism from a to y, then there exists a morphism from x to y." In particular, we could also interpret this statement in the opposite category $(\mathbb{R}, \ge)$. The resulting dual statement

$$(\forall a \in A, \exists\, x \to a) \text{ and } (\forall y \in \mathbb{R}\, (\forall a \in A, \exists\, y \to a) \text{ implies } (\exists\, y \to x)),$$

defines the **infimum**.

Exercises.

Exercise 1.2.1. The dual definitions of supremum and infimum given in Example 1.2.6 can be generalized to any poset $(\mathsf{P}, \leq)$, regarded as a category. Prove that the supremum of a subset of objects of P is unique in such a way that the dual proof demonstrates that the infimum of a subset of objects in unique.

1.3. Functoriality

The first component of what might be called the philosophy of category theory is that any mathematical object should be considered together with its accompanying notion of structure-preserving morphism. Categories are themselves mathematical objects, if of a somewhat unfamiliar sort. So what is a morphism between categories?

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

- An object $Fc \in \mathsf{D}$, for each object $c \in \mathsf{C}$.
- A morphism $Ff: Fc \to Fc' \in \mathsf{D}$, for each morphism $f: c \to c' \in \mathsf{C}$. Note we explicitly require that the domain and codomain of Ff equal F applied to the domain or codomain of f.

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

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

Put concisely, a functor consists of a mapping on objects and a mapping on morphisms that preserves all of the structure of a category, namely domains and codomains, composition, and identities.

Examples 1.3.2.

(i) There is an endofunctor[14] $P: \mathsf{Set} \to \mathsf{Set}$ that sends a set A to its powerset $P(A) = \{A' \subset A\}$ and function $f: A \to B$ to the direct-image function $f_*: P(A) \to P(B)$ that sends $A' \subset A$ to $f(A') \subset B$.

(ii) Each of the categories listed in Example 1.1.2 has a **forgetful functor** whose codomain is the category of sets. For example $U: \mathsf{Group} \to \mathsf{Set}$ sends a group to its underlying set and a group homomorphism to its underlying function. The functor $U: \mathsf{Top} \to \mathsf{Set}$ sends a space to its set of points.

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

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

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

(vi) The **fundamental group** defines a functor $\pi_1: \mathsf{Top}_* \to \mathsf{Group}$; a continuous function $f: (X, x) \to (Y, y)$ of based spaces induces a group homomorphism $f_*: \pi_1(X, x) \to \pi_1(Y, y)$ and this assignment is functorial, satisfying the two

[14] A **endofunctor** is a functor whose domain is equal to its codomain.

functoriality axioms described above. A precise expression of the statement that "the fundamental group is a homotopy invariant" is that this functor factors through the functor $\mathsf{Top}_* \to \mathsf{Htpy}_*$ to define a functor $\pi_1 \colon \mathsf{Htpy}_* \to \mathsf{Group}$.

(vii) There is a functor $F \colon \mathsf{Set} \to \mathsf{Group}$ that sends a set X to the **free group** on X. This is the group whose elements are finite *words* whose letters are elements $x \in X$ or their formal inverses x^{-1}, modulo an equivalence relation that equates the words xx^{-1} and $x^{-1}x$ with the empty word. Multiplication is by concatenation, with the empty word serving as the identity.

(viii) The chain rule expresses the functoriality of the derivative. Let Diff_* denote the category whose objects are pointed finite-dimensional real vector spaces $(\mathbb{R}^n, a)$ of varying dimensions and whose morphisms are pointed differentiable functions. The **total derivative** of $f \colon \mathbb{R}^n \to \mathbb{R}^m$, carrying the designated basepoint $a \in \mathbb{R}^n$ to $f(a) \in \mathbb{R}^m$, gives rise to a matrix called the **Jacobian matrix** containing the partial derivatives of f at the point a. This defines the action on morphisms of a functor $D \colon \mathsf{Diff}_* \to \mathsf{Mat}_{\mathbb{R}}$; on objects, D assigns a pointed vector space its dimension. Given $g \colon \mathbb{R}^m \to \mathbb{R}^k$ carrying the designated basepoint $f(a) \in \mathbb{R}^m$ to $gf(a) \in \mathbb{R}^k$, functoriality of D asserts that the product of the Jacobian of f at a with the Jacobian of g at $f(a)$ equals the Jacobian of gf at a. This is the chain rule.[15]

More examples of functors will appear shortly, but first we want to illustrate the utility of knowing that the assignment of a mathematical object of one type to mathematical objects of another type is *functorial*.

Theorem 1.3.3 (Brouwer Fixed Point Theorem). *Any continuous endomorphism of a 2-dimensional disk has a fixed point.*

Proof. This follows more-or-less immediately from the functoriality of the fundamental group $\pi_1 \colon \mathsf{Top}_* \to \mathsf{Group}$. Supposing $f \colon D^2 \to D^2$ is such that $f(x) \neq x$ for all $x \in D^2$, we can define a continuous function $r \colon D^2 \to S^1$ that carries a point $x \in D^2$ to the intersection of the ray from $f(x)$ to x with the boundary circle S^1. Note that the function r fixes the points on the boundary circle $S^1 \subset D^2$. Thus, it is a *retraction* of the inclusion $i \colon S^1 \hookrightarrow D^2$, which is to say, the composite $S^1 \xrightarrow{i} D^2 \xrightarrow{r} S^1$ is the identity.

Pick any base point on the boundary circle S^1 and apply π_1 to obtain a composable pair of group homomorphisms:

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

By the functoriality axioms, we must have

$$\pi_1(r) \cdot \pi_1(i) = \pi_1(ri) = \pi_1(1_{S^1}) = 1_{\pi_1(S^1)}.$$

However, a computation involving covering spaces reveals that $\pi_1(S^1) = \mathbb{Z}$, while $\pi_1(D^2) = 0$, the trivial group. The composite endomorphism $\pi_1(r) \cdot \pi_1(i)$ of $\mathbb{Z}$ must be zero, since it factors through the trivial group. Thus, it cannot equal the identity homomorphism, which carries the generator $1 \in \mathbb{Z}$ to itself $(0 \neq 1)$. This contradiction proves that the retraction r cannot exist, and so f must have a fixed point. □

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

[15] Taking a more sophisticated perspective, we could regard the derivative as the action on morphisms of a functor from the category Man to the category of tangent bundles.

Definition 1.3.4. A **contravariant functor** F from C to D is a functor $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$.[16] Explicitly, this consists of the following data:

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

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

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

Notation 1.3.5. As ysed in the statement of definition 1.3.4, the author believes that the following directional conventions offer the least possibility for confusion. A morphism in the domain of a functor $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$ will always be depicted as an arrow $f\colon c \to c'$ in C, pointing from its domain in C to its codomain in C. Similarly, its image will always be depicted as an arrow $Ff\colon Fc' \to Fc$ in D, pointing from its domain to its codomain. Note that these conventions require that the domain and codomain objects switch their relative places, from left to right, but in examples, for instance in the case where C and D are concrete categories, this placement will be the familiar one. Graphically, the mapping on morphisms given by a contravariant functor will be depicted as follows:

$$\mathsf{C}^{\mathrm{op}} \xrightarrow{F} \mathsf{D}$$

$$\begin{array}{ccc} c & \mapsto & Fc \\ {\scriptstyle f}\big\downarrow & \mapsto & \big\uparrow{\scriptstyle Ff} \\ c' & \mapsto & Fc' \end{array}$$

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

In summary, even in the presence of opposite categories, we always make an effort to draw arrows pointing in the "correct way" and depict composition in the usual order.[17]

Examples 1.3.6.

(i) The contravariant powerset functor $P\colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$ sends a set A to its powerset $P(A)$ and a function $f\colon A \to B$ to the inverse-image function $f^{-1}\colon P(B) \to P(A)$ that sends $B' \subset B$ to $f^{-1}(B') \subset A$.

(ii) There is a functor $(-)^*\colon \mathsf{Vect}_{\Bbbk}^{\mathrm{op}} \to \mathsf{Vect}_{\Bbbk}$ that carries a vector space to its **dual vector space** $V^* = \mathrm{Hom}(V, \Bbbk)$. A vector in V^* is a **linear functional** on V, i.e., a linear map $V \to \Bbbk$. This functor is contravariant, with a linear map

[16] In this text, a contravariant functor F from C to D will always be written as $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$. Some mathematicians omit the "op" and let the context or surrounding verbiage convey the variance. We think this is bad practice, as the co- or contra-variance is an essential part of the data of a functor, which is not necessarily determined by its assignation on objects. More to the point, we find that this notational conventions helps mitigate the consequences of temporary distraction. Seeing $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$ written on a chalkboard immediately conveys that F is a contravariant functor from C to D, even to the most spaced-out observer. A similar ethic will motivate other notational conventions stressed below.

[17] Of course, technically there is no meaning to the phrase "opposite category": every category is the opposite of some other category (its opposite category). But in practice, there is no question which of Set and $\mathsf{Set}^{\mathrm{op}}$ is the "opposite category," and sufficiently many of the other cases can be deduced from this one.

$\phi\colon V \to W$ sent to the linear map $\phi^*\colon W^* \to V^*$ that pre-composes a linear functional $W \xrightarrow{\omega} \Bbbk$ with ϕ to obtain a linear functional $V \xrightarrow{\phi} W \xrightarrow{\omega} \Bbbk$.

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

(iv) Let G be a group, regarded as a one-object category $\mathsf{B}G$. A functor $X\colon \mathsf{B}G \to \mathsf{C}$ specifies an object $X \in \mathsf{C}$ (the unique object in its image) together with an endomorphism $g_*\colon X \to X$ for each $g \in G$. This assignment must satisfy two conditions:
  (a) $h_*g_* = (hg)_*$ for all $g, h \in G$.
  (b) $e_* = 1_X$, where $e \in G$ is the identity element.

In summary, the functor $\mathsf{B}G \to \mathsf{C}$ defines an **action** of the group G on the object $X \in \mathsf{C}$. When $\mathsf{C} = \mathsf{Set}$, the object X endowed with such an action is called a ***G*-set**. When $\mathsf{C} = \mathsf{Vect}_{\Bbbk}$, the object X is called a ***G*-representation**. When $\mathsf{C} = \mathsf{Top}$, the object X is called a ***G*-space**. Note the utility of this categorical language for defining several analogous concepts simultaneously.

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

(vi) For a generic small category C, a functor $\mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ is called a (Set-valued) **presheaf** on C. A typical example is the functor $O(X)^{\mathrm{op}} \to \mathsf{Set}$ whose domain is the poset $O(X)$ of open subsets of a topological space X and whose value at $U \subset X$ is the set of continuous real-valued functions on U. The action on morphisms is by restriction. This presheaf is a *sheaf*, satisfying an axiom that is discussed in §**??**.

The following result, appearing immediately after the definition of functors first appeared in [**EM42b**], is arguably the first lemma in category theory.

Lemma 1.3.7. *Functors preserve isomorphisms.*

Proof. Consider a functor $F\colon \mathsf{C} \to \mathsf{D}$ and an isomorphism $f\colon x \to y$ in C with inverse $g\colon y \to x$. Applying the two functoriality axioms, we calculate that

$$F(g)F(f) = F(gf) = F(1_x) = 1_{Fx}.$$

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

Example 1.3.6.(iv) shows that a representation (or group action, more generally) may be regarded as a functor whose domain is the group in question. Because the elements $g \in G$ are isomorphisms, when regarded as morphisms in the 1-object category $\mathsf{B}G$ that represents the group, their images under any such functor must also be isomorphisms in the target category. In particular, in the case of a G-representation $V\colon \mathsf{B}G \to \mathsf{Vect}_{\Bbbk}$, the linear map $g_*\colon V \to V$ must be an *automorphism* of the vector space V. The point is that the functoriality axioms (i) and (ii) imply automatically that each g_* is an automorphism and that $(g^{-1})_* = (g_*)^{-1}$; the proof is a special case of Lemma 1.3.7. Neither condition need be explicitly required. In summary:

COROLLARY 1.3.8. *When a group G acts functorially on an object X in a category C, its elements g must act by automorphisms $g_*: X \to X$ and, moreoever, $(g_*)^{-1} = (g^{-1})_*$.*

DEFINITION 1.3.9. If C is locally small, then for any object $c \in \mathsf{C}$ we may define a pair of covariant and contravariant **functors represented by** c

$$\mathsf{C}(c, -): \mathsf{C} \to \mathsf{Set} \qquad \mathsf{C}(-, c): \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}.$$

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

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

We leave it to the reader to verify that the assignments just described satisfy the two functoriality axioms. Note that Lemma 1.3.7 specializes in the case of represented functors to give a proof of the implications (i)⇒(ii) and (i)⇒(iii) of Lemma 1.2.3. These functors will play a staring role in Chapter 2, where a number of examples will be given.

The data of the covariant and contravariant functors introduced in Definition 1.3.9 may be encoded in a single **bifunctor**, by which we mean a functor of two variables. Its domain is given by the product of a pair of categories.

DEFINITION 1.3.10. For any categories C and D there is a category $\mathsf{C} \times \mathsf{D}$ which we call their **product** whose

- objects are ordered pairs (c, d), where c is an object of C and d is an object of D,
- morphisms are ordered pairs $(f, g): (c, d) \to (c', d')$, with $f: c \to c' \in \mathsf{C}$ and $g: d \to d' \in \mathsf{D}$,

and in which composition and identities are defined componentwise.

DEFINITION 1.3.11. If C is locally small, then there is a **two-sided represented functor**

$$\mathsf{C}(-, -): \mathsf{C}^{\mathrm{op}} \times \mathsf{C} \to \mathsf{Set}$$

defined in the evident manner. A pair of objects (x, y) is mapped to the hom-set $\mathsf{C}(x, y)$. A pair of morphisms $f: w \to x$ and $h: y \to z$ is sent to the function

$$\mathsf{C}(x, y) \xrightarrow{(f^*, h_*)} \mathsf{C}(w, z)$$

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

At the beginning of this section, we suggested that functors define morphisms between categories. Indeed, categories and functors assemble into a category. Here we are confronted with size issues of an even more significant nature than we were confronted with in Remark 1.1.4. Let Cat denote the category whose objects are small categories and whose morphisms are functors between them. This category is locally small but not small: it contains Set, Group, $\mathsf{Groupoid}$, Monoid, and Poset as proper subcategories. However, none of these categories are *objects* of Cat.

The non-small categories of Example 1.1.2 are objects of CAT, some category of "large" categories and functors between them. We do not want CAT to be so large as to contain itself, so we require each object in CAT to be a locally small category; the category CAT defined in this way is not locally small, and so is thus excluded. There is an inclusion functor $\mathsf{Cat} \to \mathsf{CAT}$ but no obvious functor pointing in the other direction.

The category of categories gives rise to a notion of an **isomorphism of categories**, defined by interpreting Definition 1.1.5 in CAT. Namely, an isomorphism of categories is given by a pair of inverse functors $F : \mathsf{C} \to \mathsf{D}$ and $G : \mathsf{D} \to \mathsf{C}$ so that the composites GF and FG respectively equal the identity functors on C and on D. An isomorphism induces a bijection between the objects of C and objects of D and likewise for the morphisms.

Examples 1.3.12. For instance:

(i) The functor $(-)^{\mathrm{op}} : \mathsf{Cat} \to \mathsf{Cat}$ defines a non-trivial automorphism of the category of categories.
(ii) For any group G the categories $\mathsf{B}G$ and $\mathsf{B}G^{\mathrm{op}}$ are isomorphic via the functor $(-)^{-1}$ that sends each morphism $g \in G$ to its inverse. Similarly any groupoid is isomorphic to the opposite category via the functor that acts as the identity on objects and sends a morphism to its unique inverse morphism.
(iii) For any space X, there is a contravariant isomorphism of poset categories $\mathcal{O}(X) \cong \mathcal{C}(X)^{\mathrm{op}}$ that associates an open subset of X to its closed complement.
(iv) The category $\mathsf{Mat}_{\Bbbk}$ is isomorphic to its opposite via an identity-on-objects functor that carries a matrix to its transpose.

A category, however, is not generally isomorphic to its opposite category. Aside from the finite ordinals, every (non-groupoidal) category mentioned thusfar provides a counterexample.

Example 1.3.13. Let E/F be a finite **Galois extension**: this means that F is a finite-index subfield of E, every element of E satisfies some non-zero polynomial with coefficients in F, and the field of elements of E fixed by the automorphism group $\mathrm{Aut}(E/F)$ of automorphisms of E fixing every element of F is F (and no larger). In this case $G = \mathrm{Aut}(E/F)$ is called the **Galois group** of the Galois extension E/F.

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

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

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

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

There is a functor $(-)_+ \colon \mathsf{Set}^\partial \to \mathsf{Set}_*$, whose codomain is the category of pointed sets, that sends a set X to the pointed set X_+ which is defined to be the disjoint union of X with a freely-added basepoint. By the axiom of regularity, we might define $X_+ := X \cup \{X\}$.[18] A partial function $f \colon X \to Y$ gives rise a pointed function $f_+ \colon X_+ \to Y_+$ that sends every point outside of the domain of definition of f to the formally added basepoint of Y_+. The inverse functor $U \colon \mathsf{Set}_* \to \mathsf{Set}^\partial$ discards the basepoint and sends a based function $f \colon (X, x) \to (Y, y)$ to the partial function $X\backslash\{x\} \to Y\backslash\{y\}$ with the maximal possible domain of definition.

By construction, we see that the composite $U(-)_+$ is the identity endofunctor of the category Set^∂ of sets and partially defined functions. By contrast, the other composite $(U-)_+ \colon \mathsf{Set}_* \to \mathsf{Set}_*$ sends a pointed set (X, x) to $(X\backslash\{x\} \cup \{X\backslash\{x\}\}, X\backslash\{x\})$. Now these sets are isomorphic but they are not identical. Nor is another set-theoretical construction of the "freely added basepoint" likely to define an inverse to the functor $U \colon \mathsf{Set}_* \to \mathsf{Set}^\partial$. It is too restrictive to ask for the categories Set^∂ and Set_* to be isomorphic.

Indeed, there is a better way to decide whether two categories may safely be regarded as "the same." To define it, we must relax the identities $GF = 1_{\mathsf{C}}$ and $FG = 1_{\mathsf{D}}$ between functors $F \colon \mathsf{C} \to \mathsf{D}$ and $G \colon \mathsf{D} \to \mathsf{C}$ that define an isomorphism of categories. To do so, we introduce what in French is called a *morphisme de foncteurs*, the notion that launched the entire subject of category theory: a *natural transformation*.

Exercises.

Exercise 1.3.1. What is a functor between groups, regarded as one-object categories?

Exercise 1.3.2. What is a functor between preorders, regarded as categories?

Exercise 1.3.3. What is the difference between a functor $\mathsf{C}^{\mathrm{op}} \to \mathsf{D}$ and a functor $\mathsf{C} \to \mathsf{D}^{\mathrm{op}}$? What is the difference between a functor $\mathsf{C} \to \mathsf{D}$ and a functor $\mathsf{C}^{\mathrm{op}} \to \mathsf{D}^{\mathrm{op}}$?

Exercise 1.3.4. Define an isomorphism between Segal's category Γ defined in Example 1.2.2.(iv) and $\mathsf{Fin}_*^{\mathrm{op}}$.

1.4. Naturality

Consider the category $\mathsf{Vect}_{\Bbbk}^{\mathrm{fd}}$ of finite dimensional $\Bbbk$-vector spaces. Any object $V \in \mathsf{Vect}_{\Bbbk}^{\mathrm{fd}}$ is isomorphic to its linear dual, the vector space $V^* = \mathrm{Hom}(V, \Bbbk)$ described in Example 1.3.6.(ii), because the dimension of V^* equals the dimension of V. This can be proven through the construction of an explicit **dual basis**: choose a basis $e_1, \ldots, e_n$ for V and then define $e_1^*, \ldots, e_n^* \in V^*$ by

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

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

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

[18] In the axioms of Zermelo-Fraenkel set theory, elements of sets (like everything else in its mathematical universe) are themselves sets. The axiom of regularity prohibits a set from being an element of itself. As $X \notin X$, we are free to add the element X as a disjoint basepoint.

As it turns out, this isomorphism has a simpler description. For any $v \in V$, the "evaluation function" $\mathrm{ev}_v \colon f \mapsto f(v) \colon V^* \to \Bbbk$ defines a linear functional on V^*. It turns out the assignment $v \mapsto \mathrm{ev}_v$ defines a linear isomorphism $V \cong V^{**}$, this time requiring no "unnatural" choice of basis.[19]

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

Definition 1.4.1. Given categories C and D and functors $F, G \colon \mathsf{C} \rightrightarrows \mathsf{D}$,[20] a **natural transformation** $\alpha \colon F \Rightarrow G$ consists of:

- an arrow $\alpha_c \colon Fc \to Gc$ in D for each object $c \in \mathsf{C}$, called the **components** of the natural transformation,

so that, for any morphism $f \colon c \to c'$ in C, the following square of morphisms in D

(1.4.2)
$$\begin{array}{ccc} Fc & \xrightarrow{\alpha_c} & Gc \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Gf} \\ Fc' & \xrightarrow[\alpha_{c'}]{} & Gc' \end{array}$$

commutes, i.e., has a common composite $Fc \to Gc'$ in D. A **natural isomorphism** is a natural transformation in which every component is an isomorphism.

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

$$\mathsf{C} \underset{G}{\overset{F}{\Downarrow \alpha}} \mathsf{D}$$

Examples 1.4.3.

(i) For vector spaces of any dimension, the map $\mathrm{ev} \colon V \to V^{**}$ that sends $v \in V$ to the linear function $\mathrm{ev}_v \colon V^* \to \Bbbk$ defines the components of a natural transformation from the identity endofunctor on $\mathsf{Vect}_\Bbbk$ to the double dual functor. To

[19] In fact, $e_i^{**}(e_j^*) = e_j^*(e_i) = \mathrm{ev}_{e_i}(e_j^*)$, and so the two isomorphisms $V \cong V^{**}$ are the same — it is only our description that has improved.

[20] We use the rather suggestive symbol "$\rightrightarrows$" as an abbreviation for a **parallel pair** of morphisms in a category, e.g., the pair of functors F and G with common domain C and codomain D.

check that the naturality square

$$\begin{array}{ccc} V & \xrightarrow{\text{ev}} & V^{**} \\ {\scriptstyle \phi}\downarrow & & \downarrow{\scriptstyle \phi^{**}} \\ W & \xrightarrow{\text{ev}} & W^{**} \end{array}$$

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

(ii) By contrast, the identity functor and the single dual functor are not naturally isomorphic. One technical obstruction is somewhat beside the point: the identity functor is covariant while the dual functor is contravariant.[21] But there's also an essential failure of naturality. The isomorphisms $V \cong V^*$ that exist when V is finite dimensional require the choice of a basis, which will be preserved by essentially no linear maps, indeed by no non-identity linear endomorphism.[22]

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

(iv) For G a group, Example 1.3.6.(iv) shows that a functor $X\colon \mathsf{B}G \to \mathsf{C}$ corresponds to an object $X \in \mathsf{C}$ equipped with a left action of G. What is a natural transformation between a pair $X, Y\colon \mathsf{B}G \to \mathsf{C}$ of such functors? Its data consists of a morphism $\eta\colon X \to Y$ in C that is G**-equivariant**, meaning that for each $g \in G$ the diagram

$$\begin{array}{ccc} X & \xrightarrow{\eta} & Y \\ {\scriptstyle g_*}\downarrow & & \downarrow{\scriptstyle g_*} \\ X & \xrightarrow[\eta]{} & Y \end{array}$$

commutes.

(v) The open and closed subset functors described in Example 1.3.6.(iii), here regarded as functors $O, C\colon \mathsf{Top}^{\text{op}} \rightrightarrows \mathsf{Set}$, are naturally isomorphic. The components $O(X) \cong C(X)$ of the natural isomorphism are defined by taking an open subset of X to its complement, which is closed. Naturality asserts that the process of forming complements commutes with the operation of taking preimages.

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

[21] A more flexible notion of *extranatural transformation* can accommodate functors with conflicting variance [**ML98a**, IX.4]; see Exercise 1.4.4.

[22] A proof that there exists no extranatural isomorphism between the identity and dual functors on the categories of finite dimensional vector spaces is given in [**EM45**, p 234].

Now given any homomorphism $\phi\colon G \to H$, the diagram

$$\begin{array}{ccc} G & \xrightarrow{\eta_G} & G^{\mathrm{op}} \\ {\scriptstyle\phi}\big\downarrow & & \big\downarrow{\scriptstyle\phi^{\mathrm{op}}} \\ H & \xrightarrow[\eta_H]{} & H^{\mathrm{op}} \end{array}$$

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

(vii) Define an endofunctor of $\mathsf{Vect}_{\Bbbk}$ by $V \mapsto V \otimes V$. There is a natural transformation from the identity functor to this endofunctor whose components are the zero maps, but this is the only such natural transformation: there is no basis-independent way to define a linear map $V \to V \otimes V$. The same result is true for the category of Hilbert spaces and linear operators between them, in which context it is related to the "no cloning theorem" in quantum physics.[23]

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

Proposition 1.4.4. *The isomorphisms $A \cong TA \oplus (A/TA)$ are not natural in $A \in \mathsf{Ab}_{\mathrm{fg}}$.*

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

$$A \to A/TA \to TA \oplus (A/TA) \cong A \tag{1.4.5}$$

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

To derive the contradiction, we first show that every natural endomorphism α of the identity functor on $\mathsf{Ab}_{\mathrm{fg}}$ is multiplication by some $n \in \mathbb{Z}$. Clearly the component of α at $\mathbb{Z}$ has this description for some n, and moreover by inspecting (1.4.5) in the case $A = \mathbb{Z}$ we see that $n \neq 0$. But note that homomorphisms $\mathbb{Z} \xrightarrow{a} A$ correspond bijectively to elements $a \in A$ by choosing a to be the image of $1 \in \mathbb{Z}$. Thus, commutativity of

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\alpha_{\mathbb{Z}} = n\cdot -} & \mathbb{Z} \\ {\scriptstyle a}\big\downarrow & & \big\downarrow{\scriptstyle a} \\ A & \xrightarrow[\alpha_A]{} & A \end{array}$$

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

Finally, consider $A = \mathbb{Z}/2n\mathbb{Z}$. This group is torsion, so any map, such as $\alpha_{\mathbb{Z}/2n\mathbb{Z}}$, which factors through the quotient by its torsion subgroup is zero. But $n \neq 0 \in \mathbb{Z}/2n\mathbb{Z}$, a contradiction. □

Example 1.4.6. The **Riesz representation theorem** can be expressed as a natural isomorphism of functors from the category KHaus of compact Hausdorff spaces and continuous maps to the category Ban of real Banach spaces and continuous linear maps. Let

[23] The states in a quantum mechanical system are modeled by vectors in a Hilbert space and the observables are operators on that space. See [**Bae06**] for more.

$\Sigma\colon \mathsf{KHaus} \to \mathsf{Ban}$ be the functor that carries a compact Hausdorff space X to the Banach space $\Sigma(X)$ of signed Baire measures on X and sends a continuous map $f\colon X \to Y$ to the map $\mu \mapsto \mu \circ f^{-1}\colon \Sigma(X) \to \Sigma(Y)$. Let $C^*\colon \mathsf{KHaus} \to \mathsf{Ban}$ be the functor that carries X to the linear dual $C(X)^*$ of the Banach space $C(X)$ of continuous real-valued functions on X.

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

$$\phi_\mu(g) := \int_X g\, d\mu, \qquad g \in C(X).$$

For each $\mu \in \Sigma(X)$, $f\colon X \to Y$, and $h \in C(Y)$ we have

$$\int_X hf\, d\mu = \int_Y h\, d(\mu \circ f^{-1}),$$

which says that the assignment $\mu \mapsto \phi_\mu$ defines the components of a natural transformation $\eta\colon \Sigma \to C^*$. The Riesz representation theorem asserts that this natural transformation is a natural isomorphism.

Example 1.4.7. Consider morphisms $f\colon w \to x$ and $h\colon y \to z$ in a locally small category C. Post-composition by h and pre-composition by f define functions between hom-sets

$$\begin{array}{ccc} \mathsf{C}(x,y) & \xrightarrow{h\cdot -} & \mathsf{C}(x,z) \\ {\scriptstyle -\cdot f}\downarrow & & \downarrow{\scriptstyle -\cdot f} \\ \mathsf{C}(w,y) & \xrightarrow[h\cdot -]{} & \mathsf{C}(w,z) \end{array} \tag{1.4.8}$$

In Definition 1.3.11 and elsewhere, $h\cdot -$ was denoted by h_* and $-\cdot f$ was denoted by f^*, but we find this less-concise notation to be move evocative here. Associativity of composition implies that this diagram commutes: for any $g\colon x \to y$, the common image is $hgf\colon w \to z$.

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

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

Example 1.4.9. For sets A and B, let $A \times B$ denote their cartesian product, let $A + B$ denote their disjoint union, and let A^B denote the set of functions from B to A. Then we have the following natural isomorphisms

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

In the first instance, the isomorphism defines the components of a natural transformation between a pair of functors $\mathsf{Set} \times \mathsf{Set} \times \mathsf{Set} \to \mathsf{Set}$. For the others, the variance in the variables appearing as "exponents" is contravariant. This is because the assignment $(B, A) \mapsto A^B$ defines a functor $\mathsf{Set}^{\mathrm{op}} \times \mathsf{Set} \to \mathsf{Set}$, a special case of the two-sided represented functor of Definition 1.3.11.

The displayed natural isomorphisms restrict to the category $\mathsf{Fin}_{\mathrm{iso}}$ of finite sets and bijections. Our interest in this category is on account of the cardinality functor $|-|\colon \mathsf{Fin}_{\mathrm{iso}} \to$

$\mathbb{N}$, whose codomain is the discrete category of natural numbers.[24] Writing $a = |A|$, $b = |B|$, and $c = |C|$, the cardinality functor carries these natural isomorphisms to the equations

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

through a process called **decategorification**. Reversing directions, $\mathsf{Fin}_{\text{iso}}$ is a **categorification** of the natural numbers, which reveals that the familiar laws of arithmetic follow from more fundamental natural isomorphisms between various constructions on sets.

Exercises.

Exercise 1.4.1. Suppose $\alpha\colon F \Rightarrow G$ is a natural isomorphism. Show that the inverses of the component morphisms define the components of a natural isomorphism $\alpha^{-1}\colon G \Rightarrow F$.

Exercise 1.4.2. What is a natural transformation between a parallel pair of functors between groups, regarded as one-object categories?

Exercise 1.4.3. What is a natural transformation between a parallel pair of functors between preorders, regarded as categories?

Exercise 1.4.4. Given a pair of functors $F\colon \mathsf{A} \times \mathsf{B} \times \mathsf{B}^{\text{op}} \to \mathsf{D}$ and $G\colon \mathsf{A} \times \mathsf{C} \times \mathsf{C}^{\text{op}} \to \mathsf{D}$ a family of morphisms

$$\alpha_{a,b,c}\colon F(a,b,b) \to G(a,c,c)$$

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

$$\begin{array}{ccc} F(a,b,b) & \xrightarrow{\alpha_{a,b,c}} & G(a,c,c) \\ {\scriptstyle F(f,1_b,1_b)}\downarrow & & \downarrow{\scriptstyle G(f,1_c,1_c)} \\ F(a',b,b) & \xrightarrow[\alpha_{a',b,c}]{} & G(a',c,c) \end{array} \qquad \begin{array}{ccc} F(a,b,b') & \xrightarrow{F(1_a,1_b,g)} & F(a,b,b) \\ {\scriptstyle F(1_a,g,1_{b'})}\downarrow & & \downarrow{\scriptstyle \alpha_{a,b,c}} \\ F(a,b',b') & \xrightarrow[\alpha_{a,b',c}]{} & G(a,c,c) \end{array} \qquad \begin{array}{ccc} F(a,b,b) & \xrightarrow{\alpha_{a,b,c'}} & G(a,c',c') \\ {\scriptstyle \alpha_{a,b,c}}\downarrow & & \downarrow{\scriptstyle G(1_a,1_{c'},h)} \\ G(a,c,c) & \xrightarrow[G(1_a,h,1_c)]{} & G(a,c',c) \end{array}$$

The left-hand square asserts that the components $\alpha_{-,b,c}\colon F(-,b,b) \Rightarrow G(-,c,c)$ define a natural transformation in a for each $b \in \mathsf{B}$ and $c \in \mathsf{C}$. The remaining squares assert that the components $\alpha_{a,-,c}\colon F(a,-,-) \Rightarrow G(a,c,c)$ and $\alpha_{a,b,-}\colon F(a,b,b) \Rightarrow G(a,-,-)$ define transformations that are respectively extranatural in b and in c. Explain why the functors F and G must have a common target category for this definition to make sense.

1.5. Equivalence of categories

Natural transformations bear close analogy with the notion of homotopy from topology with one important difference: natural transformations are not generally invertible.[25] As in Example 1.1.3.(iv), let $\mathbb{1}$ denote the discrete category with a single object and let $\mathbb{2}$ denote the category with two objects $0, 1 \in \mathbb{2}$ and a single non-identity arrow $0 \to 1$. There are two evident functors $i_0, i_1\colon \mathbb{1} \to \mathbb{2}$ whose subscripts designate the objects in their image. A natural transformation $\alpha\colon F \Rightarrow G$ between functors $F, G\colon \mathsf{C} \rightrightarrows \mathsf{D}$ is precisely a functor $H\colon \mathsf{C} \times \mathbb{2} \to \mathsf{D}$ such that H restricts along i_0 and i_1 to the functors F and G, i.e., so

[24] Mathematical invariants often take the form of a functor from a groupoid to a discrete category.

[25] A natural transformation is invertible if and only if each of its constituent arrows is an isomorphism, in which case the pointwise inverses assemble into a natural transformation by Exercise 1.4.1.

that

(1.5.1)
$$\begin{array}{ccccc} \mathsf{C} & \xrightarrow{i_0} & \mathsf{C}\times\mathbb{2} & \xleftarrow{i_1} & \mathsf{C} \\ & {\scriptstyle F}\searrow & \downarrow{\scriptstyle H} & \swarrow{\scriptstyle G} & \\ & & \mathsf{D} & & \end{array}$$

commutes. Here i_0 denotes the functor defined on objects by $c \mapsto (c, 0)$; it may be regarded as the product of the identity functor on C with the functor i_0.

For instance, if $\mathsf{C} = \mathbb{2}$, each functor $F, G\colon \mathbb{2} \to \mathsf{D}$ picks out an arrow of D, which we also denote by F and G. The directed graph underlying the category $\mathbb{2} \times \mathbb{2}$ looks like

(1.5.2)
$$\begin{array}{ccc} \bullet & \longrightarrow & \bullet \\ \downarrow & \searrow & \downarrow \\ \bullet & \longrightarrow & \bullet \end{array}$$

together with four identity arrows not depicted here; the diagonal serves as the common composite of the edges of the square. The functor H necessarily maps the top and bottom arrows of (1.5.2) to F and G, respectively. The vertical arrows define the components α_0 and α_1 of the natural transformation, and the diagonal arrow witnesses that the square analogous to (1.4.2) commutes.

If, in (1.5.1), the category $\mathbb{2}$ were replaced by the category $\mathbb{I}$ with two objects and a single arrow in each hom-set, necessarily an isomorphism, then "homotopies" with this interval would be precisely natural *isomorphisms*. The category $\mathbb{2}$ defines the **walking arrow** or **free-living arrow**, while $\mathbb{I}$ defines the **walking isomorphism** or **free-living isomorphism**, in a sense that will be explained in 2.1.2.(ix) and (x). An equivalence of categories is precisely a "homotopy equivalence" where the notion of homotopy is defined using the category $\mathbb{I}$.

Definition 1.5.3. An **equivalence of categories** consists of functors $F\colon \mathsf{C} \leftrightarrows \mathsf{D}\colon G$ together with natural isomorphisms $\eta\colon 1_{\mathsf{C}} \cong GF$, $\epsilon\colon FG \cong 1_{\mathsf{D}}$.[26] Categories C and D are **equivalent**, written $\mathsf{C} \simeq \mathsf{D}$, if there exists an equivalence between them.

Example 1.5.4. The functors $(-)_+\colon \mathsf{Set}^{\partial} \to \mathsf{Set}_*$ and $U\colon \mathsf{Set}_* \to \mathsf{Set}^{\partial}$ introduced in §1.3 define an equivalence of categories between the category of pointed sets and the category of sets and partial functions. The composite $U(-)_+$ is the identity on Set^{∂}, so one of the required natural isomorphisms is the identity. There is a natural isomorphism $\eta\colon 1_{\mathsf{Set}_*} \Rightarrow (U-)_+$ whose components

$$\eta_{(X,x)}\colon (X, x) \to (X\backslash\{x\} \cup \{X\backslash\{x\}\}, X\backslash\{x\})$$

are defined to be the based functions that act as the identity on $X\backslash\{x\}$.

Consider the categories $\mathsf{Mat}_{\mathbb{k}}$ and $\mathsf{Vect}^{\mathrm{fd}}_{\mathbb{k}}$ of $\mathbb{k}$-matrices and finite dimensional $\mathbb{k}$-vector spaces together with an intermediate category $\mathsf{Vect}^{\mathrm{basis}}_{\mathbb{k}}$ whose objects are finite dimensional vector spaces with chosen basis and whose morphisms are arbitrary (not necessarily basis-preserving) linear maps. These categories are related by the displayed sequence of functors:

$$\mathsf{Mat}_{\mathbb{k}} \underset{H}{\overset{\mathbb{k}^{(-)}}{\rightleftarrows}} \mathsf{Vect}^{\mathrm{basis}}_{\mathbb{k}} \underset{C}{\overset{U}{\rightleftarrows}} \mathsf{Vect}^{\mathrm{fd}}_{\mathbb{k}}$$

[26] The notion of equivalence of categories was introduced by Grothendieck in the form of what we would now call an *adjoint equivalence*; this definition will appear in Proposition 4.3.3. This explains the directions we have adopted for the natural isomorphisms η and ϵ, which are otherwise immaterial (cf Exercise 1.4.1).

Here $U: \mathsf{Vect}_{\Bbbk}^{\text{basis}} \to \mathsf{Vect}_{\Bbbk}^{\text{fd}}$ is the forgetful functor. The functor $\Bbbk^{(-)}: \mathsf{Mat}_{\Bbbk} \to \mathsf{Vect}_{\Bbbk}^{\text{basis}}$ sends n to the vector space $\Bbbk^n$, equipped with the standard basis. An $m \times n$-matrix, interpreted with respect to the standard bases on $\Bbbk^n$ and $\Bbbk^m$ defines a linear map $\Bbbk^n \to \Bbbk^m$ and this assignment is functorial. The functor H carries a vector space to its dimension and a linear map $\phi: V \to W$ to the matrix expressing the action of ϕ on the chosen basis of V using the chosen basis of W. The functor C is defined by choosing a basis for each vector space.

These functors display equivalences of categories $\mathsf{Mat}_{\Bbbk} \simeq \mathsf{Vect}_{\Bbbk}^{\text{basis}} \simeq \mathsf{Vect}_{\Bbbk}^{\text{fd}}$. The composite equivalence $\mathsf{Mat}_{\Bbbk} \simeq \mathsf{Vect}_{\Bbbk}^{\text{fd}}$, which exists by Exercise 1.5.6, expresses an equivalence between concrete and abstract presentations of linear algebra. A direct proof of these equivalences, by defining suitable natural isomorphisms, is not difficult, but we prefer to give an indirect proof via a useful general theorem characterizing those functors forming part of an equivalence of categories. To state this result, we need a few definitions.

Definition 1.5.5. A functor $F: \mathsf{C} \to \mathsf{D}$ is

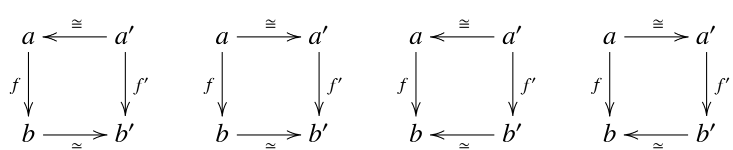
- **full** if for each $x, y \in \mathsf{C}$, the map $\mathsf{C}(x, y) \to \mathsf{D}(Fx, Fy)$ is surjective;
- **faithful** if for each $x, y \in \mathsf{C}$, the map $\mathsf{C}(x, y) \to \mathsf{D}(Fx, Fy)$ is injective;
- and **essentially surjective on objects** if for every object $d \in \mathsf{D}$ there is some $c \in \mathsf{C}$ such that d is isomorphic to Fc.

Remark 1.5.6. Fullness and faithfulness are *local* conditions; a *global* condition, by contrast, applies "everywhere." A faithful functor need not be injective on morphisms; neither must a full functor be surjective on morphisms. A full and faithful functor may be referred to as **fully faithful**, for short. A faithful functor that is injective on objects is called an **embedding**; in this case, faithfulness implies that the functor is (globally) injective on arrows. We say that a fully faithful injective-on-objects functor defines a **full embedding** of the domain category into the codomain category. The domain then defines a **full subcategory** of the codomain.

Theorem 1.5.7. *A functor defining an equivalence of categories is full, faithful, and essentially surjective on objects. Assuming the axiom of choice, any functor with these properties defines an equivalence of categories.*

The proof of Theorem 1.5.7 makes repeated use of the following elementary lemma.

Lemma 1.5.8. *Any morphism $f: a \to b$ and fixed isomorphisms $a \cong a'$ and $b \cong b'$ determine a unique morphism $f': a' \to b'$ so that any of, or equivalently, all of the following four diagrams commute.*



Proof. The left-hand diagram defines f'. The commutativity of the remaining diagrams is left as Exercise 1.5.1. □

Proof of Theorem 1.5.7. First suppose that $F: \mathsf{C} \to \mathsf{D}$, $G: \mathsf{D} \to \mathsf{C}$, $\eta: 1_{\mathsf{C}} \cong GF$, and $\epsilon: FG \cong 1_{\mathsf{D}}$ define an equivalence of categories. For any $d \in \mathsf{D}$, the component of the natural isomorphism $\epsilon_d: FGd \cong d$ demonstrates that F is essentially surjective. Consider a parallel pair $f, g: c \rightrightarrows c'$ in C. If $Ff = Fg$, then both f and g define an arrow $c \to c'$

making the diagram

$$\begin{array}{ccc} c & \xrightarrow[\cong]{\eta_c} & GFc \\ {\scriptstyle f \text{ or } g}\downarrow & & \downarrow{\scriptstyle GFf=GFg} \\ c' & \xrightarrow[\eta_{c'}]{\cong} & GFc' \end{array}$$

that expresses the naturality of η commute. Lemma 1.5.8 implies that there is a unique arrow $c \to c'$ with this property, whence $f = g$. Thus, F is faithful, and by symmetry, so is G. Given $k\colon Fc \to Fc'$, by Lemma 1.5.8 Gk and the isomorphisms η_c and $\eta_{c'}$ define a unique $h\colon c \to c'$ for which both Gk and GFh make the diagram

$$\begin{array}{ccc} c & \xrightarrow[\cong]{\eta_c} & GFc \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle Gk \text{ or } GFh} \\ c' & \xrightarrow[\eta_{c'}]{\cong} & GFc' \end{array}$$

commute. By Lemma 1.5.8 again, $GFh = Gk$, whence $Fh = k$ by faithfulness of G. Thus, F is full, faithful, and essentially surjective.

For the converse, suppose now that $F\colon \mathsf{C} \to \mathsf{D}$ is full, faithful, and essentially surjective on objects. Using essential surjectivity and the axiom of choice, we may define, for each $d \in \mathsf{D}$, an object $Gd \in \mathsf{C}$ and an isomorphism $\epsilon_d\colon FGd \cong d$. For each $\ell\colon d \to d'$, Lemma 1.5.8 defines a unique morphism making the square

$$\begin{array}{ccc} FGd & \xrightarrow[\cong]{\epsilon_d} & d \\ \vdots & & \downarrow{\scriptstyle \ell} \\ FGd' & \xrightarrow[\epsilon_{d'}]{\cong} & d' \end{array}$$

commute. Since F is fully faithful, there is a unique morphism $Gd \to Gd'$ with this image under F, which we define to be $G\ell$. This definition is arranged so that the chosen isomorphisms assemble into the components of a natural transformation $\epsilon\colon FG \Rightarrow 1_{\mathsf{D}}$. It remains to prove that the assignment of arrows $\ell \mapsto G\ell$ is functorial and to define the natural isomorphism $\eta\colon 1_{\mathsf{C}} \Rightarrow GF$.

Functoriality of G is another consequence of Lemma 1.5.8 and faithfulness of F. The morphisms $FG1_d$ and $F1_{Gd}$ both make

$$\begin{array}{ccc} FGd & \xrightarrow[\cong]{\epsilon_d} & d \\ {\scriptstyle FG1_d \text{ or } F1_{Gd}}\downarrow & & \downarrow{\scriptstyle 1_d} \\ FGd & \xrightarrow[\epsilon_d]{\cong} & d \end{array}$$

commute, whence $G1_d = 1_{Gd}$. Similarly, given $\ell'\colon d' \to d''$, both $F(G\ell' \cdot G\ell)$ and $FG(\ell'\ell)$ make

$$\begin{array}{ccc} FGd & \xrightarrow[\cong]{\epsilon_d} & d \\ {\scriptstyle F(G\ell' \cdot G\ell) \text{ or } FG(\ell'\ell)}\downarrow & & \downarrow{\scriptstyle \ell'\ell} \\ FGd'' & \xrightarrow[\epsilon_{d''}]{\cong} & d'' \end{array}$$

commute, whence $G\ell' \cdot G\ell = G(\ell'\ell)$.

Finally, by full and faithfulness of F, we may define the isomorphisms $\eta_c \colon c \to GFc$ by specifying isomorphisms $F\eta_c \colon Fc \to FGFc$; see Exercise 1.5.8. Define $F\eta_c$ to be ϵ_{Fc}^{-1}. For any $f \colon c \to c'$, the outer rectangle

$$\begin{array}{ccccc} Fc & \xrightarrow{F\eta_c} & FGFc & \xrightarrow{\epsilon_{Fc}} & Fc \\ {\scriptstyle Ff}\downarrow & & {\scriptstyle FGFf}\downarrow & & \downarrow{\scriptstyle Ff} \\ Fc' & \xrightarrow[F\eta_{c'}]{} & FGFc' & \xrightarrow[\epsilon_{Fc'}]{} & Fc' \end{array}$$

commutes, both composites being Ff. The right-hand rectangle commutes, by naturality of ϵ. Because $\epsilon_{Fc'}$ is an isomorphism, this implies that the left-hand square commutes; see Lemma 1.6.25. Faithfulness of F tells us that $\eta_{c'} \cdot f = GFf \cdot \eta_c$, i.e., that η is a natural transformation. □

Applying Theorem 1.5.7, it is easy to see that the functors

$$\mathsf{Mat}_{\Bbbk} \leftrightarrows \mathsf{Vect}_{\Bbbk}^{\text{basis}} \rightleftarrows \mathsf{Vect}_{\Bbbk}^{\text{fd}}$$

define an equivalence of categories. For instance, the morphisms in the category $\mathsf{Vect}_{\Bbbk}^{\text{basis}}$ are defined so that U is fully faithful.

A category is **connected** if any pair of objects can be connected by a finite zig-zag of morphisms.

PROPOSITION 1.5.9. *Any connected groupoid is equivalent, as a category, to the automorphism group of any of its objects.*

PROOF. Choose any object g of a connected groupoid G and let $G = \mathsf{G}(g, g)$ denote its automorphism group. The inclusion $\mathsf{B}G \hookrightarrow \mathsf{G}$ mapping the unique object of $\mathsf{B}G$ to $g \in \mathsf{G}$ is full and faithful, by definition, and essentially surjective, since G was assumed to be connected. Apply Theorem 1.5.7. □

As a special case, we obtain the following result:

COROLLARY 1.5.10. *In a path connected space X, any choice of basepoint $x \in X$ yields an isomorphic fundamental group $\pi_1(X, x)$.*

PROOF. Recall from Example 1.1.8.(ii) that any space X has a fundamental groupoid $\Pi_1(X)$ whose objects are points in X and whose morphisms are basepoint-preserving homotopy classes of paths in X. Picking any point x, the group of automorphisms of the object $x \in \Pi_1(X)$ is exactly the fundamental group $\pi_1(X, x)$. Proposition 1.5.9 implies that every automorphism group is equivalent, as a category, to $\Pi_1(X)$; thus, by Exercise 1.5.6, any pair are equivalent to each other. An equivalence between 1-object categories is an isomorphism. Exercise 1.4.2 reveals that an isomorphism of groups, regarded as 1-object categories, is exactly an isomorphism of groups in the usual sense (a bijective homomorphism). Thus, all of the fundamental groups defined by choosing a basepoint in a path-connected space are isomorphic. □

REMARK 1.5.11. Frequently, one functor of an equivalence of categories can be defined canonically, while the inverse equivalence requires the axiom of choice. In the case of the equivalence between the fundamental group and fundamental groupoid of a path-connected space, we can say more precisely that one direction is natural, while the other is not. Write

$\mathsf{Top}_*^{\mathrm{pc}}$ for the category of path-connected based topological spaces. We regard the fundamental group π_1 and fundamental groupoid Π_1 as a parallel pair of functors:

$$\pi_1 \colon \mathsf{Top}_*^{\mathrm{pc}} \xrightarrow{\pi_1} \mathsf{Group} \hookrightarrow \mathsf{Cat} \qquad \text{and} \qquad \Pi_1 \colon \mathsf{Top}_*^{\mathrm{pc}} \xrightarrow{U} \mathsf{Top} \xrightarrow{\Pi_1} \mathsf{Groupoid} \hookrightarrow \mathsf{Cat}.$$

The inclusion of the fundamental group into the fundamental groupoid defines a natural transformation $\pi_1 \Rightarrow \Pi_1$ such that each component $\pi_1(X, x) \to \Pi_1(X)$, itself a functor, is furthermore an equivalence of categories. The definition of the inverse equivalence $\Pi_1(X) \to \pi_1(X, x)$, requires the choice, for each point $p \in X$, of a path connecting p to the basepoint x. These (path homotopy classes of) chosen paths need not be preserved by maps in $\mathsf{Top}_*^{\mathrm{pc}}$. Thus the inverse equivalences $\Pi_1(X) \to \pi_1(X, x)$ do not assemble into a natural transformation.

The group of automorphisms of any object in a connected groupoid, considered in Proposition 1.5.9, is one example of the **skeleton of a category**.

Definition 1.5.12. A category C is **skeletal** if it contains just one object in each isomorphism class. The **skeleton** skC of a category C is the unique (up to isomorphism) skeletal category that is equivalent to C.

The category skC may be constructed by choosing one object in each isomorphism class in C and defining skC to be the full subcategory on this collection of objects. Immediately, the inclusion $\mathsf{skC} \hookrightarrow \mathsf{C}$ defines an equivalence of categories. This construction, however, fails to define a functor $\mathsf{sk}(-) \colon \mathsf{CAT} \to \mathsf{CAT}$.

Note than an equivalence between skeletal categories is necessarily an isomorphism of categories. As Exercise 1.5.6 demonstrates, equivalence of categories is an equivalence relation. Thus, two categories are equivalent if and only if their skeletons are isomorphic. For this reason, we feel free to speak of *the* skeleton of a category, even though its construction is not canonical.

Examples 1.5.13.

(i) The skeleton of a connected groupoid is the group of automorphisms of any of its objects (see Proposition 1.5.9).
(ii) The skeleton of the category defined by a preorder, as described in Example 1.1.3.(iii), is a poset.
(iii) The skeleton of the category $\mathsf{Vect}_{\Bbbk}^{\mathrm{fd}}$ is the category $\mathsf{Mat}_{\Bbbk}$.
(iv) The skeleton of the category $\mathsf{Fin}_{\mathrm{iso}}$ is the category whose objects are positive integers and with $\mathrm{Hom}(n, n) = \Sigma_n$, the group of permutations of n elements. The hom-sets between distinct natural numbers are all empty.
(v) Let $X \colon \mathsf{B}G \to \mathsf{Set}$ be a left G-set. Its **translation groupoid** $\mathsf{T}_G X$ has elements of X as objects. A morphism $g \colon x \to y$ is an element $g \in G$ so that $g \cdot x = y$. The objects in the skeleton $\mathsf{skT}_G X$ are the connected components in the translation groupoid. These are precisely the **orbits** of the group action, which partition X in precisely this manner. Consider $x \in X$ as a representative of its orbit O_x. Because the translation groupoid is equivalent to its skeleton, we must have $\mathrm{Hom}_{\mathsf{skT}_G X}(O_x, O_x) \cong \mathrm{Hom}_{\mathsf{T}_G X}(x, x)$, the set of automorphisms of x. This group consists of precisely those $g \in G$ so that $g \cdot x = x$. In other words, the group $\mathrm{Hom}_{\mathsf{T}_G X}(x, x)$ is the **stabilizer** G_x of x with respect to the G-action. Note that this argument implies that any pair of elements in the same orbit must have isomorphic stabilizers. As is always the case for a skeletal groupoid, there are no morphisms between distinct objects. In summary, the skeleton of the translation

groupoid, as a category, is the disjoint union of the stabilizer groups, indexed by the orbits of the action of G on X.

The set of morphisms in the translation category with domain x is isomorphic to G. This set may be expressed as a disjoint union of hom-sets $\mathrm{Hom}_{\mathsf{T}_G X}(x, y)$, where y ranges over the orbit O_x. Each of these hom-sets is isomorphic to $\mathrm{Hom}_{\mathsf{T}_G X}(x, x) = G_x$. In particular, $|G| = |O_x| \cdot |G_x|$, proving the **orbit-stabilizer theorem**.

A guiding principle in category theory is that categorically-defined concepts should be equivalence invariant. Some category theorists go so far as to call a definition "evil" if it is not invariant under equivalence of categories. The only evil definitions we have introduced thus far are smallness and discreteness. A category is **essentially small** if it is equivalent to a small category, or, equivalently, if its skeleton is a small category. A category is **essentially discrete** if it is equivalent to a discrete category.

The following constructions and definitions are equivalence invariant:

- If a category is locally small, any category equivalent to it is again locally small.
- If a category is a groupoid, any category equivalent to it is again a groupoid.
- If $\mathsf{C} \simeq \mathsf{D}$ then $\mathsf{C}^{\mathrm{op}} \simeq \mathsf{D}^{\mathrm{op}}$.
- The product of a pair of categories is equivalent to the product of any pair of equivalent categories.
- An arrow in C is an isomorphism if and only if its image under an equivalence $\mathsf{C} \to \mathsf{D}$ is an isomorphism.

We conclude with a final remark. By Theorem 1.5.7, a full and faithful functor $F\colon \mathsf{C} \to \mathsf{D}$ defines an equivalence onto its **essential image**, the full subcategory of objects isomorphic to Fc for some $c \in \mathsf{C}$. Fully faithful functors have a useful property stated as Exercise 1.5.8: if F is full and faithful and Fc and Fc' are isomorphic in D, then c and c' are isomorphic in C. We will introduce what are easily the most important fully faithful functors in category theory in Chapter 2: the covariant and contravariant Yoneda embeddings.

Exercises.

Exercise 1.5.1. Prove Lemma 1.5.8.

Exercise 1.5.2. Show that any category that is equivalent to a locally small category is locally small.

Exercise 1.5.3. Let G be a connected groupoid and let G be the group of automorphisms at any of its objects. The inclusion $G \hookrightarrow \mathsf{G}$ defines an equivalence of categories. Construct an inverse equivalence $\mathsf{G} \to G$.

Exercise 1.5.4. Characterize the categories that are equivalent to discrete categories. A category that is connected and essentially discrete is called **chaotic**.

Exercise 1.5.5. Prove that the composite of a pair of full, faithful, or essentially surjective functors again has the same properties.

Exercise 1.5.6. Prove that if $\mathsf{C} \simeq \mathsf{D}$ and $\mathsf{D} \simeq \mathsf{E}$ then $\mathsf{C} \simeq \mathsf{E}$. Conclude that equivalence of categories is an equivalence relation.

Exercise 1.5.7. Consider the functors $\mathsf{Ab} \to \mathsf{Group}$ (inclusion), $\mathsf{Ring} \to \mathsf{Ab}$ (forgetting the multiplication), $(-)^{\times}\colon \mathsf{Ring} \to \mathsf{Group}$ (taking the group of units), $\mathsf{Ring} \to \mathsf{Rng}$ (dropping the multiplicative unit), $\mathsf{Field} \to \mathsf{Ring}$ (inclusion), $\mathsf{Mod}_R \to \mathsf{Ab}$ (forgetful). Determine which functors are full, which are faithful, and which are essentially surjective. Do any define an equivalence of categories?

Exercise 1.5.8. In the presence of a full and faithful functor $F: \mathsf{C} \to \mathsf{D}$ prove that

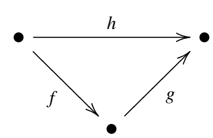
(i) Objects $x, y \in \mathsf{C}$ are isomorphic if and only if Fx and Fy are isomorphic in D.
(ii) A morphism $f: x \to y \in \mathsf{C}$ is an isomorphism if and only if $Ff: Fx \to Fy \in \mathsf{D}$ is an isomorphism.

1.6. The art of the diagram chase

> The diagrams incorporate a large amount of information. Their use provides extensive savings in space and in mental effort. In the case of many theorems, the setting up of the correct diagram is the major part of the proof. We therefore urge that the reader stop at the end of each theorem and attempt to construct for himself the relevant diagram before examining the one which is given in the text. Once this is done, the subsequent demonstration can be followed more readily; in fact, the reader can usually supply it himself.
>
> Samuel Eilenberg and Norman Steenrod, *Foundations of Algebraic Topology*, xi.

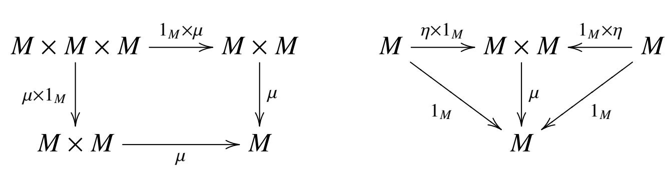
Speaking loosely, a **diagram** in a category consists of a collection of morphisms, usually depicted as a directed graph. The diagram **commutes** if any two paths of composable arrows in this directed graph with common source and target have the same composite; a more precise definition will be given in a moment. For example, a commutative triangle

(1.6.1)



asserts that the hypotenuse h equals the composite gf of the two legs. Commutative diagrams in a category can be used to define more complicated mathematical objects. For example:

Definition 1.6.2. A **monoid** is an object $M \in \mathsf{Set}$ together with morphisms $\mu: M \times M \to M$ and $\eta: 1 \to M$ so that the following diagrams commute:



The morphism $\mu: M \times M \to M$ defines a binary "multiplication" operation on M. The morphism $\eta: 1 \to M$, whose domain is a singleton set, identifies an element $\eta \in M$. The three axioms demand that multiplication is associative and that multiplication on the left or right by the element η acts as the identity. The advantage of the commutative diagrams approach to this definition is that it readily generalizes to other categories. For example:

Definition 1.6.3. A **topological monoid** is an object $M \in \mathsf{Top}$ together with morphisms $\mu\colon M \times M \to M$ and $\eta\colon 1 \to M$ so that the following diagrams commute:

$$\begin{array}{ccc} M \times M \times M & \xrightarrow{1_M \times \mu} & M \times M \\ {\scriptstyle \mu \times 1_M}\downarrow & & \downarrow{\scriptstyle \mu} \\ M \times M & \xrightarrow[\mu]{} & M \end{array} \qquad \begin{array}{ccccc} M & \xrightarrow{\eta \times 1_M} & M \times M & \xleftarrow{1_M \times \eta} & M \\ & {\scriptstyle 1_M}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_M} & \\ & & M & & \end{array}$$

A **unital ring**[27] is an object $R \in \mathsf{Ab}$ together with morphisms $\mu\colon R \otimes_{\mathbb{Z}} R \to R$ and $\eta\colon \mathbb{Z} \to R$ so that the following diagrams commute:

$$\begin{array}{ccc} R \otimes_{\mathbb{Z}} R \otimes_{\mathbb{Z}} R & \xrightarrow{1_R \otimes_{\mathbb{Z}} \mu} & R \otimes_{\mathbb{Z}} R \\ {\scriptstyle \mu \otimes_{\mathbb{Z}} 1_R}\downarrow & & \downarrow{\scriptstyle \mu} \\ R \otimes_{\mathbb{Z}} R & \xrightarrow[\mu]{} & R \end{array} \qquad \begin{array}{ccccc} R & \xrightarrow{\eta \otimes_{\mathbb{Z}} 1_R} & R \otimes_{\mathbb{Z}} R & \xleftarrow{1_R \otimes_{\mathbb{Z}} \eta} & R \\ & {\scriptstyle 1_R}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_R} & \\ & & R & & \end{array}$$

A $\Bbbk$-**algebra** is an object $R \in \mathsf{Vect}_{\Bbbk}$ together with morphisms $\mu\colon R \otimes_{\Bbbk} R \to R$ and $\eta\colon \Bbbk \to R$ so that the following diagrams commute:

$$\begin{array}{ccc} R \otimes_{\Bbbk} R \otimes_{\Bbbk} R & \xrightarrow{1_R \otimes_{\Bbbk} \mu} & R \otimes_{\Bbbk} R \\ {\scriptstyle \mu \otimes_{\Bbbk} 1_R}\downarrow & & \downarrow{\scriptstyle \mu} \\ R \otimes_{\Bbbk} R & \xrightarrow[\mu]{} & R \end{array} \qquad \begin{array}{ccccc} R & \xrightarrow{\eta \otimes_{\Bbbk} 1_R} & R \otimes_{\Bbbk} R & \xleftarrow{1_R \otimes_{\Bbbk} \eta} & R \\ & {\scriptstyle 1_R}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_R} & \\ & & R & & \end{array}$$

There are evident formal similarities in each of these four definitions; they are all special cases of a general notion of a monoid in a *monoidal category*, as defined in §E.2. The morphisms $\eta\colon 1 \to M$ in the case of topological monoids, $\eta\colon \mathbb{Z} \to R$ in the case of unital rings, and $\eta\colon \Bbbk \to R$ in the case of $\Bbbk$-algebras do no more and no less than specify an element of M or R to serve as the multiplicative unit. We will introduce language to describe the role played in each case by the topological space 1, the abelian group $\mathbb{Z}$, and the vector space $\Bbbk$ in Chapter 2.

In the case of a topological monoid, the condition that $\mu\colon M \times M \to M$ is a morphism in Top demands that the multiplication function is continuous; an example is the circle $S^1 \subset \mathbb{C}$ with addition of angles. For unital rings, the morphism $\mu\colon R \otimes_{\mathbb{Z}} R \to R$ represents a bilinear homomorphism of abelian groups from $R \times R$ to R; in particular multiplication distributes over addition in R. The role of the tensor product in the definition of a $\Bbbk$-algebra is similar.

Definition 1.6.4. A **diagram** in a category C is a functor $F\colon \mathsf{J} \to \mathsf{C}$ whose domain, the **indexing category**, is a small category.

A diagram is typically depicted by drawing the objects and morphisms in its image, with the domain category left implicit. Nonetheless, the indexing category J plays an important role. Functoriality requires that any composition relation that holds in J must hold in the image of the diagram. To say that the diagram is **commutative** is generally to assert that the indexing category J is a preorder, so that any two paths of composable arrows have a common composite. Functoriality of the diagram then implies that these

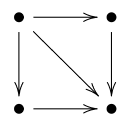
[27] A not-necessarily unital ring may be defined by ignoring the morphism η and the pair of commutative triangles. We decline to introduce this definition, however, because Bjorn Poonen makes a persuasive case why all rings should have a 1 [**Poo14**].

composition relations must also hold in C. As an immediate consequence of the form of our definition of a commutative diagram, we have the following result.

Lemma 1.6.5. *Functors preserve commutative diagrams.*

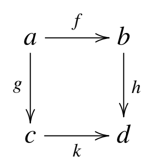
Proof. A diagram in C is given by a functor $F\colon \mathsf{J} \to \mathsf{C}$, whose domain is a small category. Given any functor $G\colon \mathsf{C} \to \mathsf{D}$, the composite $GF\colon \mathsf{J} \to \mathsf{D}$ defines the image of the diagram in D. ◻

Example 1.6.6. Consider $\mathbb{2} \times \mathbb{2}$, the category with four objects and the displayed non-identity morphisms

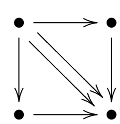


In $\mathbb{2} \times \mathbb{2}$, the diagonal morphism is the composite of both the top and right morphisms and the left and bottom morphisms; in particular, these composites are equal. A diagram indexed by $\mathbb{2}\times\mathbb{2}$, typically drawn without the diagonal composite, is a **commutative square**.

Remark 1.6.7. In practice, one thinks of the indexing category as a directed graph, defining the **shape** of the diagram, together with specified commutativity relations. For example, to define a functor with domain $\mathbb{2} \times \mathbb{2}$ it suffices to specify the images of the four objects together with four morphisms

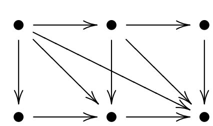


subject to the relation that $hf = kg$. When indexing categories are represented in this way, the commutativity relations become an essential part of the data. They distinguish between the category $\mathbb{2} \times \mathbb{2}$ that indexes a commutative square and the category



that indexes a not-necessarily commutative square; here the two diagonals represent distinct composites of the two paths along the edges of the square.

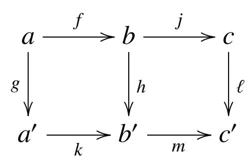
Example 1.6.8. Consider $\mathbb{2}\times\mathbb{3}$, the category with six objects and the displayed non-identity morphisms



The long diagonal asserts that the outer composite triples coincide. The short diagonals assert, respectively, that the left-hand and right-hand squares commute. The inner parallelogram also commutes. A diagram indexed by $\mathbb{2} \times \mathbb{3}$, typically drawn without any of the diagonals, is a **commutative rectangle**.

Two commutative squares define a commutative rectangle: a collection of morphisms with the indicated sources and targets

(1.6.9)

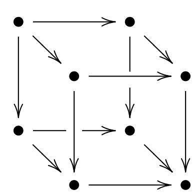


define a $\mathbb{2} \times \mathbb{3}$-shaped diagram provided that $hf = kg$ and $\ell j = mh$. This is a special case of the following more general result, which describes the induced relations in the "algebra of composition" encoded by the arrows in a category.

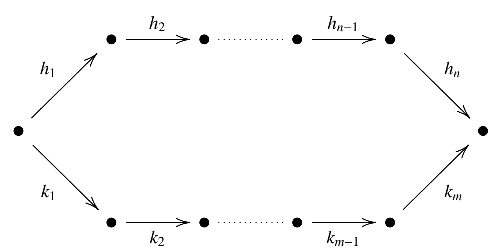
Lemma 1.6.10. *Suppose $f_1, \ldots, f_n$ is a composable sequence ("path") of morphisms in a category. If the composite $f_k f_{k-1} \cdots f_{i+1} f_i$ equals $g_m \cdots g_1$, for another composable sequence of morphisms $g_1, \ldots, g_m$, then $f_n \cdots f_1 = f_n \cdots f_{k+1} g_m \cdots g_1 f_{i-1} \cdots f_1$.*

Proof. Composition is well-defined: if the composites $g_m \cdots g_1$ and $f_k f_{k-1} \cdots f_{i+1} f_i$ define the same arrow, then the results of pre- or post-composing with other sequences of arrows must also be the same. □

This very simple result underlies most proofs by "diagram chasing." When a diagram is depicted by a simple (meaning there is at most one edge between any two vertices) acyclic directed graph, the most common convention is to include commutativity relations that assert that any two paths in the diagram with a common source and target commute. For example, the category $\mathbb{2} \times \mathbb{2} \times \mathbb{2}$ indexes the **commutative cube**, which is typically depicted as follows:



In such cases, Lemma 1.6.10 and transitivity of equality implies that commutativity of the entire diagram may be checked by establishing commutativity of each minimal subdiagram. Here, a minimal subdiagram corresponds to a composition relation $h_n \cdots h_1 = k_m \cdots k_1$ that cannot be factored into a relation between shorter paths of composable morphisms. The graph corresponding to a minimal relation is a "directed polygon"



a commutative triangle, as in (1.6.1), being the simplest case. This sort of argument is called "equational reasoning" in [**Sim11**, 2.1], which provides an excellent short introduction to diagram chasing.

The following results have simple proofs by diagram chasing.

LEMMA 1.6.11. *Consider a commutative square $\beta\alpha = \delta\gamma$ in which each of the morphisms is an isomorphism. Then $\alpha^{-1}\beta^{-1} = \gamma^{-1}\delta^{-1}$.*

PROOF. Post-compose the composition relation $\beta\alpha = \delta\gamma$ with $\alpha^{-1}\beta^{-1}$ and pre-compose with $\gamma^{-1}\delta^{-1}$. □

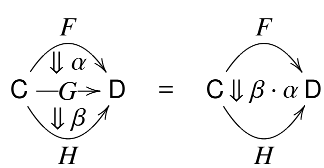
A number of important facts about natural transformations are also proven by diagram chasing. A natural transformation in French is a *morphisme de foncteurs*. Indeed, for any fixed pair of categories C and D, there is a category D^{C} whose objects are functors $\mathsf{C} \to \mathsf{D}$ and whose morphisms are natural transformations of such.[28] Given a functor $F\colon \mathsf{C} \to \mathsf{D}$, the natural transformation $1_F\colon F \Rightarrow F$ is defined with components $(1_F)_c = 1_{Fc}$. The following lemma describes composition of morphisms in D^{C}.

LEMMA 1.6.12. *Suppose $\alpha\colon F \Rightarrow G$ and $\beta\colon G \Rightarrow H$ are natural transformations between parallel functors $F, G, H\colon \mathsf{C} \to \mathsf{D}$. Then there is a natural transformation $\beta\alpha\colon F \Rightarrow H$ whose components $(\beta \cdot \alpha)_c = \beta_c \cdot \alpha_c$ are defined to be the composites of the components of α and β.*

PROOF. Naturality of α and β implies that for any $f\colon c \to c'$ in the domain category each square, and thus also the composite rectangle, commutes:

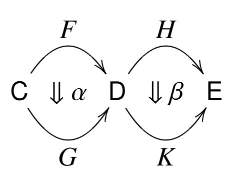
$$\begin{array}{ccccc} Fc & \xrightarrow{\alpha_c} & Gc & \xrightarrow{\beta_c} & Hc \\ {\scriptstyle Ff}\downarrow & & {\scriptstyle Gf}\downarrow & & \downarrow{\scriptstyle Hf} \\ Fc' & \xrightarrow[\alpha_{c'}]{} & Gc' & \xrightarrow[\beta_{c'}]{} & Hc' \end{array}$$
□

The composition operation defined in Lemma 1.6.12 is called **vertical composition**. Drawing the parallel functors horizontally, a composable pair of natural transformations in the category D^{C} fits into a pasting diagram



As the terminology suggests, there is also a **horizontal composition** operation, defined by the following lemma.

LEMMA 1.6.13. *Given a pair of natural transformations*



[28]Care should be taken with size when discussing **functor categories**. If C and D are small, then D^{C} is again a small category, but if C and D are locally small then D^{C} need not be; this is only guaranteed if D is locally small and C is small. In summary, the formation of functor categories defines a bifunctor $\mathsf{Cat}^{\mathrm{op}} \times \mathsf{Cat} \to \mathsf{Cat}$ or $\mathsf{Cat}^{\mathrm{op}} \times \mathsf{CAT} \to \mathsf{CAT}$, but the category of functors between two non-small categories may be even larger than these categories are.

there is a natural transformation $\beta * \alpha \colon HF \Rightarrow KG$ whose component at $c \in \mathsf{C}$ is defined as the composite of the following commutative square

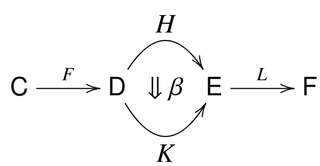
$$\begin{array}{ccc} HFc & \xrightarrow{\beta_{Fc}} & KFc \\ {\scriptstyle H\alpha_c}\downarrow & & \downarrow{\scriptstyle K\alpha_c} \\ HGc & \xrightarrow[\beta_{Gc}]{} & KGc \end{array} \tag{1.6.14}$$

Proof. The square (1.6.14) commutes by naturality of $\beta \colon H \Rightarrow K$ applied to the morphism $\alpha_c \colon Fc \to Gc$ in D. To prove that the components $(\beta * \alpha)_c \colon HFc \to KGc$ so-defined are natural, we must show that $KGf \cdot (\beta * \alpha)_c = (\beta * \alpha)_{c'} \cdot HFf$ for any $f \colon c \to c'$ in C. This relation holds on account of the commutative rectangle

$$\begin{array}{ccccc} HFc & \xrightarrow{H\alpha_c} & HGc & \xrightarrow{\beta_{Gc}} & KGc \\ {\scriptstyle HFf}\downarrow & & {\scriptstyle HGf}\downarrow & & \downarrow{\scriptstyle KGf} \\ HFc' & \xrightarrow[H\alpha_{c'}]{} & HGc' & \xrightarrow[\beta_{Gc'}]{} & KGc' \end{array}$$

The right-hand square commutes by naturality of β. The left-hand square commutes by naturality of α and Lemma 1.6.5, which states that functors, in this case the functor H, preserve commutative diagrams. □

Remark 1.6.15. The natural transformations $H\alpha \colon HF \Rightarrow HG$, $K\alpha \colon KF \Rightarrow KG$, $\beta F \colon HF \Rightarrow KF$, and $\beta G \colon HG \Rightarrow KG$ appearing in Lemma 1.6.13 are defined by **whiskering** the natural transformations α and β with the functors H and K or F and G, respectively. The terminology is on account of the following graphical depiction of the whiskered composite



$L\beta F \colon LHF \Rightarrow LKF$ of the natural transformation β with the functors F and L. As demonstrated by Lemma 1.6.13, we are also interested in the case where either L or F is an identity. See Exercises 1.6.2 and 1.6.3.

In certain special cases, commutativity of diagrams can be automatic. For instance, any parallel sequences of composable morphisms in a preorder must have a common composite precisely because any hom-set in a preorder has at most one element!

Definition 1.6.16. An object $x \in \mathsf{C}$ is **initial** if for every $c \in \mathsf{C}$ there is a unique morphism $x \to c$. Dually, an object $x \in \mathsf{C}$ is **terminal** if for every $c \in \mathsf{C}$ there is a unique morphism $c \to x$.

Lemma 1.6.17. *Let $f_1, \ldots, f_n$ and $g_1, \ldots, g_m$ be composable sequences of morphisms so that the domain of f_1 equals the domain of g_1 and the codomain of f_n equals the codomain of g_m. If this common codomain is a terminal object, or if this common domain is an initial object, then $f_n \cdots f_1 = g_m \cdots g_1$.*

Proof. The two dual statements are immediate consequences of the uniqueness part of Definition 1.6.16. □

Examples 1.6.18. Many of the categories we have met have initial and terminal objects.

(i) The empty set is an initial object in Set and any singleton set is terminal.
(ii) In Top, the empty and singleton spaces are respectively initial and terminal.
(iii) In Set_*, any singleton set is both initial and terminal.
(iv) In Mod_R, the zero module is both initial and terminal. Similarly, the trivial group is both initial and terminal in Group.
(v) The zero ring is a terminal object in Ring. To identify an initial object, we must clarify what sort of rings are meant. Henceforth, let Ring denote the category of unital rings and ring homomorphisms that preserve the multiplicative identity. This is a non-full subcategory of the larger category Rng of rings which do not necessarily have a multiplicative identity and homomorphisms which need not preserve one if it happens to exist. The integers define an initial object in Ring but not in Rng, in which the zero ring is initial.
(vi) The category Field of fields has neither initial nor terminal objects. Indeed, there are no homomorphisms between fields of different characteristic.
(vii) The empty category defines an initial object in Cat, and the category $\mathbb{1}$ is terminal.
(viii) An initial object in a preorder is a global minimal element, and a terminal object is a global maximal element.

In certain cases, one can prove that a diagram commutes by appealing to "elements" of the objects. For instance, this is possible in any concrete category.

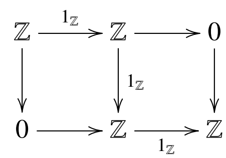
Definition 1.6.19. A **concrete category** is a category C equipped with a faithful functor $U\colon \mathsf{C} \to \mathsf{Set}$.

The functor U typically carries an object of C to its "underlying set." The faithfulness condition asserts that any parallel pair of morphisms $f, g\colon c \rightrightarrows c'$ that induce the same function $Uf = Ug$ between the underlying sets must be equal in C. The idea is that the question of whether a map between the underlying sets of objects in a concrete category is a map in the category is a condition (e.g., continuity). By contrast, the functor $U\colon \mathsf{C} \to \mathsf{Set}$ is not faithful if the maps in C have extra structure that is not visible at the level of the underlying sets (e.g., homotopy classes of maps [**Fre04**]).

Example 1.6.20. Each category listed in Example 1.1.2 is concrete, although care must be taken in the case of Graph. The most obvious forgetful functors, which send a graph to its set of vertices or its set of edges, are not faithful. However, the functor $U\colon \mathsf{Graph} \to \mathsf{Set}$ that sends a graph to the union of its set of vertices and edges is faithful.

Consider a faithful functor $U\colon \mathsf{C} \to \mathsf{D}$. Then any diagram in C whose image commutes in D also commutes in C; the point is that the necessary relations between composable sequences of morphisms are reflected from D to C by the faithful functor. In particular, to prove that a diagram in a concrete category commutes, it suffices to prove commutativity of the induced diagram of underlying sets. This amounts to showing that certain composite functions between underlying sets are the same, and this can be checked by considering the elements of those sets.

We close with a word of warning. We have seen that commutativity of a pair of adjacent squares as in (1.6.9) implies commutativity of the exterior rectangle, but the converse need not hold, as illustrated by the following diagram in Ab, in which the outer rectangle commutes but neither square does:



Neither does commutativity of one of the two squares, plus the outer rectangle, imply commutativity of the other in general. The issue is that a composition relation of the form

$$gh_n \cdots h_1 f = gk_m \cdots k_1 f$$

need not imply that $h_n \cdots h_1 = k_m \cdots k_1$ unless f and g have a special property, which we now introduce.

Definition 1.6.21. A morphism $f : x \to y$ in a category is

(i) a **monomorphism** if for any parallel morphisms $h, k : w \rightrightarrows x$, $fh = fk$ implies that $h = k$.
(ii) an **epimorphism** if for any parallel morphisms $h, k : y \rightrightarrows z$, $hf = kf$ implies that $h = k$.

In adjectival form, a monomorphism is **monic** and an epimorphism is **epic**. In common shorthand a monomorphism is a **mono** and an epimorphism is an **epi**.

Example 1.6.22. Suppose $f : X \to Y$ is a monomorphism in the category of sets. Then, in particular, given any two maps $x, x' : 1 \rightrightarrows X$, whose domain is the singleton set, if $fx = fx'$ then $x = x'$. Thus, monomorphisms are injective functions. Conversely, any injective function can easily be seen to be a monomorphism.

Similarly, a map $f : X \to Y$ in the category of sets is an epimorphism if and only if it is surjective. Given functions $h, k : Y \rightrightarrows Z$, the equation $hf = kf$ says exactly that h is equal to k on the image of f.

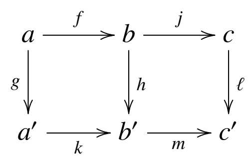
Thus, monomorphisms and epimorphisms should be regarded as categorical analogs of the notions of injective and surjective functions. As will be demonstrated in Exercise 1.6.6, a morphism $f : x \to y$ in a concrete category C is a monomorphism if its induced function $Uf : Ux \to Uy$ between the underlying sets of x and y is injective. Dually, f is an epimorphism if Uf is surjective. However, a concrete category may have more monomorphisms and epimorphisms than just those maps that have injective or surjective underlying functions (see Exercise 1.6.7).

Example 1.6.23. Suppose that $x \xrightarrow{s} y \xrightarrow{r} x$ are morphisms so that $rs = 1_x$. The map s is a **section** or **right inverse** to r, while the map r defines a **retraction** or **left inverse** to s. In this case, s is always a monomorphism and r is always an epimorphism. To acknowledge the presence of these one-sided inverses, s is said to be a **split monomorphism** and r is said to be a **split epimorphism**.

A functor may or may not preserve monomorphisms or epimorphisms, but it necessarily preserves split monomorphisms and split epimorphisms. One thinks of the retraction or section as an "equational witness" for the mono or the epi.

Example 1.6.24. By the previous example, an isomorphism is necessarily both monic and epic, but the converse need not hold in general. For example, the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is both monic and epic in the category Rng or in Ring, but this map is not an isomorphism in either category.

Lemma 1.6.25. *Consider morphisms with the indicated sources and targets*



and suppose that the outer rectangle commutes. This data defines a commutative rectangle if either:

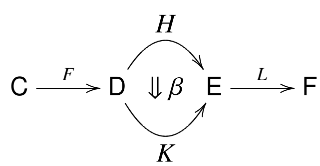
(i) *the left-hand square commutes and m is a monomorphism.*
(ii) *the right-hand square commutes and f is an epimorphism.*

Proof. The statements are dual. Assuming (i), we have $mkg = \ell jf = mhf$ by commutativity of the outer rectangle and right-hand square. Since m is a monomorphism, it follows that $kg = hf$ and thus that the rectangle commutes. □

Exercises.

Exercise 1.6.1. Show that any map from a terminal object in a category to an initial one is an isomorphism.

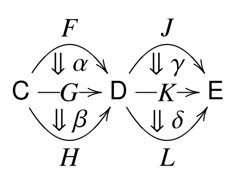
Exercise 1.6.2. Given a natural transformation $\beta\colon H \Rightarrow K$ and functors F and L as displayed in



define a natural transformation $L\beta F\colon LHF \Rightarrow LKF$ by $(L\beta F)_c = L\beta_{Fc}$. This is the **whiskered composite** of β with L and F. Prove that $L\beta F$ is natural.

Exercise 1.6.3. Redefine the horizontal composition of natural transformations introduced in Lemma 1.6.13 using vertical composition and whiskering.

Exercise 1.6.4. Given functors and natural transformations



prove that $(\delta\gamma) * (\beta\alpha) = (\delta * \beta)(\gamma * \alpha)$. That is, prove that the natural transformation $JF \Rightarrow LH$ defined by first composing vertically and then composing horizontally equals the natural transformation defined by first composing horizontally and then composing vertically. This is the rule of **middle four interchange**.

Exercise 1.6.5. Prove that the monomorphisms in any category define a subcategory of that category. Apply duality to prove that the epimorphisms also define a subcategory.

Exercise 1.6.6. Show that any faithful functor reflects monomorphisms. That is, if $F\colon \mathsf{C} \to \mathsf{D}$ is faithful, prove that if Ff is a monomorphism in D, then f is a monomorphism in C. Argue by duality that faithful functors also reflect epimorphisms. Conclude that in any concrete category, any morphism that defines an injection of underlying sets is a monomorphism and any morphism that defines a surjection of underlying sets is an epimorphism.

Exercise 1.6.7. Find a concrete category that contains a monomorphism whose underlying function is not injective. Find a concrete category that contains an epimorphism whose underlying function is not surjective.

Exercise 1.6.8. For any group G, we may define other groups:

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

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

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

Exercise 1.6.9. Prove that a bifunctor $F\colon \mathsf{C} \times \mathsf{D} \to \mathsf{E}$ is uniquely determined by:

(i) A functor $F(c, -)\colon \mathsf{D} \to \mathsf{E}$ for each $c \in \mathsf{C}$.
(ii) A natural transformation $F(f, -)\colon F(c, -) \Rightarrow F(c', -)$ for each $f\colon c \to c'$ in C, defined functorially in C.

In other words, prove that there is a bijection between functors $\mathsf{C} \times \mathsf{D} \to \mathsf{E}$ and functors $\mathsf{C} \to \mathsf{E}^{\mathsf{D}}$.

CHAPTER 2

Representability and the Yoneda lemma

Traditional approaches to the foundations of mathematics are based on set theory. In particular, many mathematical objects are defined to be sets with additional structure. For this reason, we are particularly interested in functors taking values in the category of sets. This chapter is devoted to their study.

2.1. Representable functors

A paradigmatic example is the bifunctor $\mathsf{C}(-,-)\colon \mathsf{C}^{\mathrm{op}} \times \mathsf{C} \to \mathsf{Set}$ defined for any locally small category C, the two-sided represented functor introduced in Definition 1.3.11. It sends a pair of objects $x, y \in \mathsf{C}$ to the set $\mathsf{C}(x,y)$ of morphisms in C with domain x and codomain y. A pair of morphisms $f\colon x' \to x$ and $h\colon y \to y'$ define a function

$$\mathsf{C}(x,y) \xrightarrow{h \circ - \circ f} \mathsf{C}(x',y')$$

that maps a morphism $g\colon x \to y$ to the composite morphism $hgf\colon x' \to y'$. Fixing an object in either the domain or the codomain variable, this bifunctor restricts to define the covariant and contravariant functors represented by the fixed objects

$$\mathsf{C}(x,-)\colon \mathsf{C} \to \mathsf{Set} \qquad \mathsf{C}(-,y)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$$

that were introduced in Definition 1.3.9.

Definition 2.1.1. A covariant or contravariant functor F from a locally small category C to Set is **representable** if there is an object $c \in \mathsf{C}$ and a natural isomorphism between F and the functor of appropriate variance[1] represented by c. A **representation** for a covariant functor F is then a choice of object $c \in \mathsf{C}$ together with a specified natural isomorphism $\mathsf{C}(c,-) \cong F$; similarly for a contravariant functor. In this case, we say that the functor F is **represented by** the object c.

We will have more to say about the distinction between a functor being representable and a choice of representation soon, where we will also answer the question of how unique is the choice of representation for a representable functor.

One reason for our particular interest in representable functors is because many examples occur "in nature."

Examples 2.1.2. The following covariant functors are representable.

(i) The identity functor $1_{\mathsf{Set}}\colon \mathsf{Set} \to \mathsf{Set}$ is represented by the singleton set. That is, for any set X, there is a natural isomorphism $\mathsf{Set}(1,X) \cong X$ that defines a bijection between elements $x \in X$ and functions $x\colon 1 \to X$ carrying the singleton

[1] Some authors use the term "corepresentable" for covariant representable functors, reserving "representable" for the contravariant case. We find this distinction unnecessary since the variance is always evident from the definition of the functor F.

element to x. Naturality says that for any $f\colon X \to Y$, the diagram

$$\begin{array}{ccc} \mathsf{Set}(1,X) & \xrightarrow{\cong} & X \\ {\scriptstyle f_*}\downarrow & & \downarrow{\scriptstyle f} \\ \mathsf{Set}(1,Y) & \xrightarrow[\cong]{} & Y \end{array}$$

commutes, i.e., that the composite function $1 \xrightarrow{x} X \xrightarrow{f} Y$ corresponds to the element $f(x) \in Y$, as is evidently the case.

(ii) The forgetful functor $U\colon \mathsf{Group} \to \mathsf{Set}$ is represented by the group $\mathbb{Z}$. That is, for any group G, there is a natural isomorphism $\mathsf{Group}(\mathbb{Z}, G) \cong UG$ that associates, to every element $g \in UG$, the unique homomorphism $\mathbb{Z} \to G$ that maps the integer 1 to g. This defines a bijection because every homomorphism $\mathbb{Z} \to G$ is determined by the image of the generator 1; that is to say, $\mathbb{Z}$ is the **free group on a single generator**. This bijection is natural because the composite group homomorphism $\mathbb{Z} \xrightarrow{g} G \xrightarrow{\phi} H$ carries the integer 1 to $\phi(g) \in H$.

(iii) For any unital ring R, the forgetful functor $U\colon \mathsf{Mod}_R \to \mathsf{Set}$ is represented by the R-module R. That is, there is a natural bijection between R-module homomorphisms $R \to M$ and elements of the underlying set of M, in which $m \in UM$ is associated to the unique R-module homomorphism that carries the multiplicative identity to m. This explains the appearance of the abelian group $\mathbb{Z}$ and the vector space $\Bbbk$ in Definition 1.6.3, where maps with these domains were used to specify elements in the codomains.

(iv) The functor $U\colon \mathsf{Ring} \to \mathsf{Set}$ is represented by the unital ring $\mathbb{Z}[x]$, the polynomial ring in one variable with integer coefficients. A unital ring homomorphism $\mathbb{Z}[x] \to R$ is uniquely determined by the image of x; put another way, $\mathbb{Z}[x]$ is the **free unital ring on a single generator**.

(v) The functor $U(-)^n\colon \mathsf{Group} \to \mathsf{Set}$ that sends a group G to the set of n-tuples of elements of G is represented by the free group on n generators. Similarly, the functor $U(-)^n\colon \mathsf{Ab} \to \mathsf{Set}$ is represented by the free abelian group on n generators.

(vi) The functor $(-)^\times\colon \mathsf{Ring} \to \mathsf{Set}$ that sends a unital ring to its set of units is represented by the ring $\mathbb{Z}[x, x^{-1}]$ of Laurent polynomials in one variable. That is to say, a ring homomorphism $\mathbb{Z}[x, x^{-1}] \to R$ may be defined by sending x to any unit of R and is completely determined by this assignment, and moreover there are no ring homomorphisms that carry x to a non-unit.

(vii) The forgetful functor $U\colon \mathsf{Top} \to \mathsf{Set}$ is represented by the singleton space: there is a natural bijection between elements of a topological space and continuous functions from the point.

(viii) The functor $\mathrm{ob}\colon \mathsf{Cat} \to \mathsf{Set}$ that takes a small category to its set of objects is represented by the terminal category $\mathbb{1}$: a functor $\mathbb{1} \to \mathsf{C}$ is no more and no less than a choice of object in C.

(ix) The functor $\mathrm{mor}\colon \mathsf{Cat} \to \mathsf{Set}$ that takes a small category to its set of morphisms is represented by the category $\mathbb{2}$: a functor $\mathbb{2} \to \mathsf{C}$ is no more and no less than a choice of morphism in C.

(x) The functor $\mathrm{iso}\colon \mathsf{Cat} \to \mathsf{Set}$ that takes a small category to its set of isomorphisms (pointing in a chosen direction) is represented by the category $\mathbb{I}$, with two objects and exactly one morphism in each hom-set.

(xi) The functor comp: $\mathsf{Cat} \to \mathsf{Set}$ that takes a small category to the set of composable pairs of morphisms in it is represented by the category $\mathbb{3}$. Generalizing, the ordinal $\mathbb{n} + \mathbb{1} = 0 \to 1 \to \cdots \to n$ represents the functor that takes a small category to the set of paths of n composable morphisms in it.

(xii) The forgetful functor $U \colon \mathsf{Set}_* \to \mathsf{Set}$ is represented by the two element based set: based functions out of this set correspond naturally and bijectively to elements of the target based set.

(xiii) The functors $\mathsf{Top} \to \mathsf{Set}$ that carry a topological space to its set of paths or its set of loops are representable by the unit interval I and the circle S^1 by definition. A **path** in X is a continuous function $I \to X$ while a **loop** in X is a continuous function $S^1 \to X$.

EXAMPLES 2.1.3. The following contravariant functors are representable.

(i) The contravariant powerset functor $P \colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$ is represented by the set $\Omega = \{\top, \bot\}$ with two elements. The natural isomorphism $\mathsf{Set}(A, \Omega) \cong P(A)$ is defined by the bijection that associates a function $A \to \Omega$ with the subset that is the preimage of $\top$; reversing perspectives, a subset $A' \subset A$ is identified with its **classifying function** $\chi_{A'} \colon A \to \Omega$, which sends exactly the elements of A' to the element $\top$. Naturality says that for any function $f \colon A \to B$, the diagram

$$\begin{array}{ccc} \mathsf{Set}(B,\Omega) & \xrightarrow{\cong} & P(B) \\ {\scriptstyle f^*}\downarrow & & \downarrow{\scriptstyle f^{-1}} \\ \mathsf{Set}(A,\Omega) & \xrightarrow[\cong]{} & P(A) \end{array}$$

commutes. That is, given a function $\chi_{B'} \colon B \to \Omega$ classifying the subset $B' \subset B$, the composite function $A \xrightarrow{f} B \xrightarrow{\chi_{B'}} \Omega$ classifies the subset $f^{-1}(B') \subset A$, which is easily seen to be the case.

(ii) The functor $O \colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Set}$ that sends a space to its set of open subsets is represented by the **Sierpinski space** S. This is the topological space with two points, one closed and one open. The natural bijection $\mathsf{Top}(X, S) \cong O(X)$ associates a continuous function $X \to S$ to the preimage of the open point. This bijection is natural because a composite function $Y \to X \to S$ classifies the preimage of the open subset of X under the function $Y \to X$.

(iii) The Sierpinski space also represents the functor $C \colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Set}$ that sends a space to its set of closed subsets. Composing the natural isomorphisms $O \cong \mathsf{Top}(-, S) \cong C$ we see that the closed set and open set functors are naturally isomorphic. Unpacking the definitions, we see that the composite natural isomorphism carries an open subset to its complement, which is closed. This recovers the natural isomorphism described in Example 1.4.3.(v).

(iv) The functor $\mathrm{Hom}(- \times A, B) \colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$ that sends a set X to the set of functions $X \times A \to B$ is represented by the set B^A of functions from A to B. That is, there is a natural bijection between functions $X \times A \to B$ and functions $X \to B^A$. This natural isomorphism is referred to as **currying** in computer science; by fixing a variable in a two-variable function, one obtains a family of functions in a single variable.

(v) The functor $U(-)^* \colon \mathsf{Vect}_{\mathbb{k}}^{\mathrm{op}} \to \mathsf{Set}$ that sends a vector space to the set of vectors in its dual space is represented by the vector space $\mathbb{k}$, i.e., linear maps $V \to \mathbb{k}$ are, by definition, precisely the vectors in the dual space V^*.

(vi) Fix a topological space Y, a set X, and a function $f\colon X \to UY$. Define a functor $T_f\colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Set}$ by taking $T_f(A)$ to be the set of functions $UA \to X$ so that the composite function $UA \to X \to UY$ is a continuous function between the spaces A and Y. This functor is represented by assigning X the coarsest topology so that f is continuous. It has a subbasis given by preimages of open sets in (a subbasis of) Y.

The representations considered in Examples 2.1.2 and 2.1.3 describe **universal properties** of the representing objects in a sense that we will soon make precise. Representability of certain functors can also be used to define categorical properties of the domain category, a strategy we will use repeatedly. Here is a first example.

Example 2.1.4. For any category C, not necessarily locally small, we can define a constant functor $*\colon \mathsf{C} \to \mathsf{Set}$ that sends every object $c \in \mathsf{C}$ to the singleton set. What does it mean for this functor to be representable? In particular, representability implies that there is an object $i \in \mathsf{C}$ so that for any $c \in \mathsf{C}$ the set $\mathsf{C}(i, c)$ contains exactly one element. In other words, the functor $*\colon \mathsf{C} \to \mathsf{Set}$ is representable if and only if C admits an initial object i. Dually, $\mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ is representable if and only if C admits a terminal object.

Exercises.

Exercise 2.1.1. For each of the three functors

$$\mathbb{1} \;\substack{\xrightarrow{\;0\;} \\ \xleftarrow{\;!\;} \\ \xrightarrow[\;1\;]{}}\; \mathbb{2}$$

between the categories $\mathbb{1}$ and $\mathbb{2}$, describe the corresponding natural transformations between the covariant functors $\mathsf{Cat} \to \mathsf{Set}$ represented by these categories.

Exercise 2.1.2. A functor F defines a **subfunctor** of G if there is a natural transformation $\alpha\colon F \Rightarrow G$ whose components are monomorphisms. In the case of $G\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ a subfunctor is given by a collection of subsets $Fc \subset Gc$ so that each $Gf\colon Gc \to Gc'$ restricts to define a function $Ff\colon Fc \to Fc'$. Characterize those subsets that assemble into a subfunctor of the representable functor $\mathsf{C}(-, c)$.

Exercise 2.1.3. Suppose $F\colon \mathsf{C} \to \mathsf{Set}$ is equivalent to $G\colon \mathsf{D} \to \mathsf{Set}$ in the sense that there is an equivalence of categories $H\colon \mathsf{C} \to \mathsf{D}$ so that GH and F are naturally isomorphic.

(i) If F is representable, then is G representable?
(ii) If G is representable, then is F representable?

2.2. The Yoneda lemma

> Yoneda enjoyed relating the story of the origins of this lemma, as follows. He had guided Samuel Eilenberg during Eilenberg's visit to Japan, and in this process learned homological algebra. Soon Yoneda spent a year in France (apparently in 1954 and 1955). There he met Saunders Mac Lane. Mac Lane, then visiting Paris, was anxious to learn from Yoneda, and commenced an interview with Yoneda in a café at Gare du Nord. The interview was continued on Yoneda's train until its departure. In its course, Mac Lane learned about the lemma and subsequently baptized it.
>
> Saunders Mac Lane, "The Yoneda lemma"

To gain a better understanding of the representable functors considered above and of questions of representability, such as posed in Example 2.1.4, we would like to answer the following question: given a functor $F\colon \mathsf{C} \to \mathsf{Set}$, what data is needed to define a natural isomorphism $\mathsf{C}(c,-) \cong F$? More generally, what data is needed to define a natural transformation $\mathsf{C}(c,-) \Rightarrow F$?

As a first example, consider diagrams whose shape is given by the ordinal category ω. A functor $F\colon \omega \to \mathsf{Set}$ is given by a family of sets $(F_n)_{n\in\mathbb{N}}$ together with functions $f_{n,n+1}\colon F_n \to F_{n+1}$. The functor $\omega(k,-)\colon \omega \to \mathsf{Set}$ represented by the object $k \in \omega$ corresponds to a family whose first k sets, indexed by objects $n < k$, are empty and whose remaining sets, indexed by $n \geq k$ are singletons. A natural transformation $\alpha\colon \omega(k,-) \Rightarrow F$, as always, is given by components $\alpha_n\colon \omega(k,n) \to F_n$ satisfying a naturality condition. Using the fact that the sets $\omega(k,n)$ are either empty or singletons, we display the data in a diagram

$$\begin{array}{ccccccccccccccc}
\emptyset & \longrightarrow & \emptyset & \longrightarrow \cdots \longrightarrow & \emptyset & \longrightarrow & * & \longrightarrow & * & \longrightarrow \cdots \longrightarrow & * & \longrightarrow \cdots \\
\downarrow{\scriptstyle \alpha_0} & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_{k-1}} & & \downarrow{\scriptstyle \alpha_k} & & \downarrow{\scriptstyle \alpha_{k+1}} & & \downarrow{\scriptstyle \alpha_n} & \\
F_0 & \xrightarrow[f_{0,1}]{} & F_1 & \longrightarrow \cdots \longrightarrow & F_{k-1} & \xrightarrow[f_{k-1,k}]{} & F_k & \xrightarrow[f_{k,k+1}]{} & F_{k+1} & \longrightarrow \cdots \longrightarrow & F_n & \longrightarrow \cdots
\end{array}$$

Evidently, the components α_n contain no information for $n < k$. For $n \geq k$, the components determine elements $\alpha_n \in F_n$. The naturality condition demands that $\alpha_{n+1} = f_{n,n+1}(\alpha_n)$. By an inductive argument, we see that the natural transformation $\alpha\colon \omega(k,-) \Rightarrow F$ is completely and uniquely determined by the choice of the first element $\alpha_k \in F_k$, which we think of as the image of the morphism 1_k under the k-th component map $\alpha_k\colon \omega(k,k) \to F_k$. Moreover, any element of F_k may be chosen as this image; naturality places no further restrictions.

As a second example, we consider diagrams whose indexing categories are groups.

Example 2.2.1. A group G, when regarded as a category, has a single object. Thus, there is a unique covariant represented functor $G \to \mathsf{Set}$ and a unique contravariant represented functor $G^{\mathrm{op}} \to \mathsf{Set}$. Example 1.3.6.(iv) characterized the functors from G to Set; in the covariant case, they correspond to left G-sets and in the contravariant case, they correspond to right G-sets. Recalling that the elements of G define the set of automorphisms of the unique object in the category, we see that the covariant represented functor is the G-set G, with its action by left multiplication, while the contravariant represented functor is again G, but with its right action by right multiplication. The same remarks hold for monoids.

Consider a G-set $X\colon G \to \mathsf{Set}$. Example 1.4.3.(iv) observed that a natural transformation $\phi\colon G \Rightarrow X$ is exactly a G-equivariant map $\phi\colon G \to X$. Here $\phi\colon G \to X$ is the unique component of the natural transformation; equivariance of this map expresses the naturality condition. To define ϕ we must specify elements $\phi(g) \in X$ for each $g \in G$. Equivariance demands that $\phi(g \cdot h) = g \cdot \phi(h)$. Taking h to be the identity element, we see that $\phi(g) = g \cdot \phi(e)$. In other words, the choice of $\phi(e) \in X$ forces us to define $\phi(g)$ to be $g \cdot \phi(e)$. Moreover, any choice of $\phi(e) \in X$ is permitted, because the left action of G on G is free.[2] Thus, we have proven:

Proposition 2.2.2. *G-equivariant maps $G \to X$ correspond bijectively to elements of X, the image of the identity $e \in G$.*

[2]The action of a group G on a set X is **free** if every stabilizer group is trivial. In the context of the left action of G on itself, if there were distinct elements k, h so that $g = k \cdot e = h \cdot e$, we might be forced to make contradictory definitions of $\phi(g)$.

For diagrams indexed by the poset ω and by a group G, we have seen that natural transformations whose domain is a represented functor are determined by the choice of a single element, which lives in the set defined by evaluating the codomain functor at the representing object. In each case, this element is the image of the identity morphism at the representing object.[3] This is no coincidence. Indeed, the same is true for diagrams of any shape by what is arguably the most important result in category theory: the Yoneda lemma.[4]

Theorem 2.2.3 (Yoneda lemma). *For any functor* $F\colon \mathsf{C} \to \mathsf{Set}$*, whose domain* C *is locally small, there is a bijection*

$$\mathrm{Hom}(\mathsf{C}(c,-),F) \cong Fc$$

that identifies a natural transformation $\alpha\colon \mathsf{C}(c,-) \Rightarrow F$ *with the element* $\alpha_c(1_c) \in Fc$*. Moreover this correspondence is natural in both* c *and* F*.*

Proof of the bijection. There is clearly a function $\Phi\colon \mathrm{Hom}(\mathsf{C}(c,-),F) \to Fc$ that maps a natural transformation $\alpha\colon \mathsf{C}(c,-) \Rightarrow F$ to the image of 1_c under the component function $\alpha_c\colon \mathsf{C}(c,c) \to Fc$, i.e., $\Phi(\alpha) := \alpha_c(1_c)$. Our first aim is to define an inverse $\Psi\colon Fc \to \mathrm{Hom}(\mathsf{C}(c,-),F)$ that constructs a natural transformation $\Psi(x)\colon \mathsf{C}(c,-) \Rightarrow F$ from any $x \in Fc$. To this end, we must define components $\Psi(x)_d\colon \mathsf{C}(c,d) \to Fd$ so that naturality squares, such as the following for $f\colon c \to d$ in C, commute:

$$\begin{array}{ccc} \mathsf{C}(c,c) & \xrightarrow{\Psi(x)_c} & Fc \\ {\scriptstyle f_*}\downarrow & & \downarrow{\scriptstyle Ff} \\ \mathsf{C}(c,d) & \xrightarrow[\Psi(x)_d]{} & Fd \end{array}$$

The image of the identity element $1_c \in \mathsf{C}(c,c)$ under the left-bottom composite is $\Psi(x)_d(f) \in Fd$, the value of the component $\Psi(x)_d$ at the element $f \in \mathsf{C}(c,d)$. The image under the top-right composite is $Ff(\Psi(x)_c(1_c))$. In order for Ψ to define an inverse for Φ we should let $\Psi(x)_c(1_c) = x$. Thus, naturality forces us to define

$$\Psi(x)_d(f) := Ff(x).$$

This condition completely defines the components of $\Psi(x)$.

It remains to verify that $\Psi(x)$ is natural. To that end, consider a generic morphism $g\colon d \to e$ in C (one whose domain is not necessarily the distinguished object c). We must show that

$$\begin{array}{ccc} \mathsf{C}(c,d) & \xrightarrow{\Psi(x)_d} & Fd \\ {\scriptstyle g_*}\downarrow & & \downarrow{\scriptstyle Fg} \\ \mathsf{C}(c,e) & \xrightarrow[\Psi(x)_e]{} & Fe \end{array}$$

commutes. The image of $f\colon c \to d$ along the left-bottom composite is $\Psi(x)_e(gf) := F(gf)(x)$. The image along the top-right composite is $Fg(\Psi(x)_d(f)) := Fg(Ff(x))$, which by functoriality of F, is the same element. This completes the definition of the function $\Psi\colon Fc \to \mathrm{Hom}(\mathsf{C}(c,-),F)$.

[3] Recall, the identity element in a group corresponds to the identity morphism in the one-object category.

[4] For the history of the Yoneda lemma, which first appeared in print in a paper of Grothendieck [**Gro60**], see [**ML98b**], quoted at the beginning of this chapter. I find Mac Lane's use of third person to be extremely curious.

By construction $\Phi\Psi(x) = \Psi(x)_c(1_c) = x$, so Ψ is a right inverse to Φ. We wish to show that $\Psi\Phi(\alpha) = \alpha$, i.e., that the natural transformation $\Psi(\alpha_c(1_c))$ is α. It suffices to show that these natural transformations have the same components. By definition

$$\Psi(\alpha_c(1_c))_d(f) = Ff(\alpha_c(1_c)).$$

By naturality of α, the square

$$\begin{array}{ccc} \mathsf{C}(c,c) & \xrightarrow{\alpha_c} & Fc \\ {\scriptstyle f_*}\downarrow & & \downarrow{\scriptstyle Ff} \\ \mathsf{C}(c,d) & \xrightarrow[\alpha_d]{} & Fd \end{array} \tag{2.2.4}$$

commutes, from which we see that $Ff(\alpha_c(1_c)) = \alpha_d(f)$. Thus, $\Psi(\alpha_c(1_c))_d = \alpha_d$, proving that Ψ is also a left inverse to Φ. Thus, evaluation of a natural transformation at the identity of the representing object defines an isomorphism

$$\mathrm{Hom}(\mathsf{C}(c,-),F) \xrightarrow{\cong} Fc,$$

as claimed. □

Proof of naturality. The naturality in the statement of the Yoneda lemma amounts to the following pair of assertions. Firstly, given a natural transformation $\beta\colon F \Rightarrow G$, the element of Gc representing the composite natural transformation $\beta\alpha\colon \mathsf{C}(c,-) \Rightarrow F \Rightarrow G$ for any $\alpha \in \mathrm{Hom}(\mathsf{C}(c,-),F)$ is the image under $\beta_c\colon Fc \to Gc$ of the element of Fc representing α, i.e., the diagram

$$\begin{array}{ccc} \mathrm{Hom}(\mathsf{C}(c,-),F) & \xrightarrow{\Phi^F} & Fc \\ {\scriptstyle \beta_*}\downarrow & & \downarrow{\scriptstyle \beta_c} \\ \mathrm{Hom}(\mathsf{C}(c,-),G) & \xrightarrow[\Phi^G]{} & Gc \end{array}$$

commutes. By definition $\Phi(\beta\alpha) = (\beta\alpha)_c(1_c)$, which is $\beta_c(\alpha_c(1_c))$ by the definition of vertical composition of natural transformations, and this is $\beta_c(\Phi(\alpha))$.

Secondly, given a morphism $f\colon c \to d$ in C, the element of Fd representing the composite natural transformation $\alpha f^*\colon \mathsf{C}(d,-) \Rightarrow \mathsf{C}(c,-) \Rightarrow F$ is the image under $Ff\colon Fc \to Fd$ of the element of Fc representing α, i.e., the diagram

$$\begin{array}{ccc} \mathrm{Hom}(\mathsf{C}(c,-),F) & \xrightarrow{\Phi^c} & Fc \\ {\scriptstyle (f^*)_*}\downarrow & & \downarrow{\scriptstyle Ff} \\ \mathrm{Hom}(\mathsf{C}(d,-),F) & \xrightarrow[\Phi^d]{} & Fd \end{array}$$

commutes. Here, the image of α along the top-right is $Ff(\alpha_c(1_c))$, and the image along the left-bottom is $(\alpha f^*)_d(1_d)$. By the definition of vertical composition, the d-th component of the composite natural transformation αf^* is the function

$$\begin{array}{ccccc} \mathsf{C}(d,d) & \xrightarrow{f^*} & \mathsf{C}(c,d) & \xrightarrow{\alpha_d} & Fd \\ 1_d & \mapsto & f & \mapsto & \alpha_d(f) \end{array}$$

As demonstrated by the commutative square (2.2.4), $\alpha_d(f) = Ff(\alpha_c(1_c))$, which is what we had hoped to show. □

REMARK 2.2.5. Were it not for size issues, we could express Theorem 2.2.3 more concisely as saying that the maps Φ define the components of a natural transformation of two functors that we now introduce. The pair c and F in the statement of the Yoneda lemma define an object in the product category $\mathsf{C} \times \mathsf{Set}^{\mathsf{C}}$; recall $\mathsf{Set}^{\mathsf{C}}$ is the category of functors $\mathsf{C} \to \mathsf{Set}$ and natural transformations between them. There is a functor $\mathrm{ev}\colon \mathsf{C} \times \mathsf{Set}^{\mathsf{C}} \to \mathsf{Set}$ that maps (c, F) to the set Fc, i.e., the set obtained when evaluating the functor F at the object c. This functor defines the codomain of the natural isomorphism Φ.

The definition of the domain of Φ makes use of a functor

$$\mathsf{C}^{\mathrm{op}} \xrightarrow{\ y\ } \mathsf{Set}^{\mathsf{C}}$$

$$\begin{array}{ccc} c & \mapsto & \mathsf{C}(c,-) \\ f\big\downarrow & \mapsto & \big\uparrow f^* \\ d & \mapsto & \mathsf{C}(d,-) \end{array}$$

This is also the functor obtained by applying Exercise 1.6.9 to Definition 1.3.11. Using Exercise 1.3.3 to regard y as a functor $\mathsf{C} \to (\mathsf{Set}^{\mathsf{C}})^{\mathrm{op}}$, the domain of Φ is defined to be the composite

$$\mathsf{C} \times \mathsf{Set}^{\mathsf{C}} \xrightarrow{y \times 1_{\mathsf{Set}^{\mathsf{C}}}} (\mathsf{Set}^{\mathsf{C}})^{\mathrm{op}} \times \mathsf{Set}^{\mathsf{C}} \xrightarrow{\ \mathrm{Hom}\ } \mathsf{SET}$$

$$(c, F) \quad \mapsto \quad (\mathsf{C}(c,-), F) \quad \mapsto \quad \mathrm{Hom}(\mathsf{C}(c,-), F)$$

Here is where the size issues arise. If C is small, then $\mathsf{Set}^{\mathsf{C}}$ is locally small, and we have a Set-valued hom functor, as in Definition 1.3.11. However, if C is only locally small, $\mathsf{Set}^{\mathsf{C}}$ need not be locally small; we write SET to indicate that the collection of natural transformations between a pair of functors $F, G\colon \mathsf{C} \to \mathsf{Set}$ might not be a set.

For the composite functor $\mathrm{Hom}(y(-),-)\colon \mathsf{C} \times \mathsf{Set}^{\mathsf{C}} \to \mathsf{Set}$ there is no problem, however, by the proof just given. So the Yoneda lemma asserts that evaluating at the identity morphism on the representing object defines a natural isomorphism

$$\mathsf{C} \times \mathsf{Set}^{\mathsf{C}} \underset{\mathrm{ev}}{\overset{\mathrm{Hom}(y(-),-)}{\rightrightarrows}} \mathsf{Set} \qquad \cong\!\Downarrow\!\Phi$$

The Yoneda lemma completely characterizes natural transformations between representable functors. By Theorem 2.2.3, natural transformations

$$\alpha\colon \mathsf{C}(c,-) \Rightarrow \mathsf{C}(d,-)$$

correspond to elements of $\mathsf{C}(d, c)$, i.e., to morphisms $f\colon d \to c$ in C. Given $f\colon d \to c$, the Yoneda lemma defines the components of the resulting natural transformation $\Psi(f)$ as follows:

$$\mathsf{C}(c, e) \xrightarrow{\Psi(f)_e} \mathsf{C}(d, e)$$

$$g\colon c \to e \quad \mapsto \quad gf\colon d \to e$$

Here $gf \in \mathsf{C}(d, e)$ is the image of the element $f \in \mathsf{C}(d, c)$ under the function $g_*\colon \mathsf{C}(d, c) \to \mathsf{C}(d, e)$. In other words, $\Psi(f)$ is the natural transformation $\mathsf{C}(c,-) \Rightarrow \mathsf{C}(d,-)$ defined by pre-composition by f. It's clear that distinct morphisms induce distinct natural transformations: the images of the identity morphism at c will necessarily differ. In this way, we have arrived at the following important corollary to the Yoneda lemma.

Corollary 2.2.6 (Yoneda embedding). *The functors*

$$
\begin{array}{ccc}
\mathsf{C} & \overset{y}{\hookrightarrow} & \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \\
c & \mapsto & \mathsf{C}(-,c) \\
f\downarrow & \mapsto & \downarrow f_* \\
d & \mapsto & \mathsf{C}(-,d)
\end{array}
\qquad\qquad
\begin{array}{ccc}
\mathsf{C}^{\mathrm{op}} & \overset{y}{\hookrightarrow} & \mathsf{Set}^{\mathsf{C}} \\
c & \mapsto & \mathsf{C}(c,-) \\
f\downarrow & \mapsto & \uparrow f^* \\
d & \mapsto & \mathsf{C}(d,-)
\end{array}
$$

define full and faithful embeddings.

In other words, C is isomorphic to the full subcategory of $\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ spanned by the contravariant represented functors, and C^{op} is isomorphic to the full subcategory of $\mathsf{Set}^{\mathsf{C}}$ spanned by the covariant represented functors. The functors of Corollary 2.2.6 define the **covariant** and **contravariant Yoneda embeddings**.

Fullness of the Yoneda embedding allows us to prove:

Corollary 2.2.7. *Every row operation on matrices with k rows is defined by left multiplication by some $k \times k$ matrix.*

Proof. Recall $\mathsf{Mat}_{\mathbb{k}}$ is the category whose objects are non-zero natural numbers and in which a morphism $n \to m$ is a $m \times n$ matrix, with m rows and n columns. The elements in the image of the represented functor $\mathrm{Hom}(-,k)$ are matrices with k rows. The row operations of elementary linear algebra, for instance, replacing the ith row with the sum of the ith and jth row, are easily seen to define natural endomorphisms of $\mathrm{Hom}(-,k)$; naturality here follows from linearity of (right) matrix multiplication. Thus, by Corollary 2.2.6, every row operation must be definable by left multiplication by a suitable $k \times k$ matrix. Moreover, Theorem 2.2.3 allows us to identify this matrix: it's the result of applying the row operation in question to the $k \times k$ identity matrix. □

Example 2.2.8. In the case of a group G regarded as a one-object category, the image of the covariant Yoneda embedding is the right G-set G, with G acting by right multiplication. Corollary 2.2.6 tells us that the only G-equivariant endomorphisms of the right G-set G are those maps defined by left multiplication with a fixed element of G. In particular, any G-equivariant endomorphism of G must be an automorphism, a fact that might otherwise come as a surprise.

In this way, the Yoneda embedding defines an isomorphism between G and the automorphism group of the right G-set G, an object in $\mathsf{Set}^{G^{\mathrm{op}}}$. Composing with the forgetful functor $\mathsf{Set}^{G^{\mathrm{op}}} \to \mathsf{Set}$, we obtain an isomorphism between G and a subgroup of the automorphism of group of the set G, which in this context is typically denoted by $\mathrm{Sym}(G)$. This result, that any abstract group may be realized as a subgroup of a permutation group, is known as **Cayley's theorem**.

Exercises.

Exercise 2.2.1. As discussed in Section 2.2, diagrams of shape ω are determined by a countably infinite family of objects and a countable infinite sequence of morphisms. Describe the Yoneda embedding $y\colon \omega \hookrightarrow \mathsf{Set}^{\omega^{\mathrm{op}}}$ in this manner (as a family of ω^{op}-indexed functors and natural transformations). Prove, without appealing to the Yoneda lemma, that y is full and faithful.

Exercise 2.2.2. There is a natural automorphism ι of the contravariant power-set functor $P\colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$ whose component functions $\iota_A\colon P(A) \to P(A)$ send a subset $A' \subset A$ to its complement. The Yoneda lemma tells us that ι is induced by an endomorphism of the representing object. What is it? Does this function induce a natural automorphism of the covariant power-set functor?

Exercise 2.2.3. Do there exist any non-identity natural endomorphisms of the category of spaces? I.e., does there exist any family of continuous maps $X \to X$, defined for all spaces X and not all of which are identities, that are natural in all maps in the category Top?

2.3. Universal properties

The Yoneda lemma, of Theorem 2.2.3 and Corollary 2.2.6, allows us to define objects in some category by means of their universal properties in a sense we now make precise. We will make use of the following strengthening of Lemma 1.2.3, which shows that objects in a category are isomorphic if and only if they are **representably isomorphic**.

Lemma 2.3.1. *The following are equivalent:*

(i) $f\colon x \to y$ *is an isomorphism in* C.
(ii) $f_*\colon \mathsf{C}(-, x) \Rightarrow \mathsf{C}(-, y)$ *is a natural isomorphism.*
(iii) $f^*\colon \mathsf{C}(y, -) \Rightarrow \mathsf{C}(x, -)$ *is a natural isomorphism.*

Proof. The implications (i)⇒(ii) and (i)⇒(iii) are instances of Lemma 1.3.7, following from the functoriality of the Yoneda embeddings. The implications (ii)⇒(i) and (iii)⇒(i) follow from Corollary 2.2.6 and the fact that full and faithful functors reflect isomorphisms. □

Corollary 2.3.2. *If* x *and* y *represent the same functor* $F\colon \mathsf{C} \to \mathsf{Set}$*, then* x *and* y *are isomorphic in* C.

More precisely, the representing natural isomorphisms $\mathsf{C}(x, -) \cong F \cong \mathsf{C}(y, -)$ induce a canonical isomorphism $x \cong y$ in C. For this reason, we often refer to *the* representing object of a representable functor. Category theorists often use the singular "the" in contexts where the object in question is well-defined up to canonical isomorphism.

Corollary 2.3.3. *Any two terminal objects in* C *are (uniquely) isomorphic. Dually, initial objects are uniquely isomorphic.*

Proof. Terminal objects represent the functor $\mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ that is constant at the singleton set. □

In common mathematical practice, a definition by means of a universal property takes the following form.

Definition 2.3.4. A representable functor $F\colon \mathsf{C} \to \mathsf{Set}$ or $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ expresses a **universal property** of its representing object.

Put another way, a universal property of an object x is a description of the covariant representable functor $\mathrm{Hom}(x, -)$ or of the contravariant representable functor $\mathrm{Hom}(-, x)$. Corollary 2.3.2 demonstrates that a representable functor determines a representing object up to canonical isomorphism.

The Yoneda lemma also allows us to be more precise about what data is entailed by a representation.

Example 2.3.5. A functor $E\colon G \to \mathsf{Set}$, i.e., a left G-set E, is representable if and only if there is an isomorphism $G \cong E$ of left G-sets. This implies that the action of G on E is **free** (every stabilizer group is trivial) and **transitive** (the orbit of any point is the entire set), and that E is non-empty. Conversely, any non-empty free and transitive left G-set is representable; a proof will be given in Example 2.4.6. By the Yoneda lemma, the data of a representation for E, i.e., a specific isomorphism $G \cong E$, is determined by the choice of an element $e \in E$, serving as the image of the identity $e \in G$. In other words, a representable left G-set E is a group that has forgotten its identity element. This is called a G-**torsor**.

For example, a torsor for the group $\mathbb{R}^n$ under addition is the set called n-dimensional **affine space**, commonly denoted by $\mathbb{A}^n$ (or $\mathbb{A}^n_{\mathbb{R}}$ to emphasize the ground field). We think of $\mathbb{A}^n$ as being the collection of points in n-dimensional real space, but without a chosen origin. Thinking of points in $\mathbb{R}^n$ as vectors, the free and transitive action of $\mathbb{R}^n$ on $\mathbb{A}^n$ is easy to describe: a point in $\mathbb{A}^n$ is transported along the vector in question. This action is free, because every non-zero vector acts on every point non-trivially, and transitive, because any two points in affine space can be "subtracted" to define a unique vector in $\mathbb{R}^n$ that carries the one to the other. A choice of an origin point o in $\mathbb{A}^n$ determines coordinates for all of the other points, hence an isomorphism $\mathbb{R}^n \cong \mathbb{A}^n$. The coordinates for another point $p \in \mathbb{A}^n$ are defined to be the unique vector $\vec{x} \in \mathbb{R}^n$ so that $\vec{x} \cdot o = p$. This construction is a special case of the construction of the natural transformation in the proof of the Yoneda lemma.

Example 2.3.6. Fix $\mathbb{k}$-vector spaces V and W and consider the functor

$$\mathrm{Bilin}(V, W; -)\colon \mathsf{Vect}_{\mathbb{k}} \to \mathsf{Set}$$

that sends a vector space U to the set of $\mathbb{k}$-bilinear maps $V \times W \to U$. A **bilinear map** $f\colon V \times W \to U$ is a function of two variables so that for all $v \in V$, $f(v, -)\colon W \to U$ is a linear map and for all $w \in W$, $f(-, w)\colon V \to U$ is a linear map. Equivalently, f may be defined to be a linear map $V \to \mathrm{Hom}(W, U)$ or $W \to \mathrm{Hom}(V, U)$, where the codomains are vector spaces of linear maps.

A representation for the functor $\mathrm{Bilin}(V, W; -)$ defines a vector space $V \otimes_{\mathbb{k}} W$, the **tensor product** of V and W. That is, the tensor product is defined by an isomorphism

$$\mathsf{Vect}_{\mathbb{k}}(V \otimes_{\mathbb{k}} W, U) \cong \mathrm{Bilin}(V, W; U), \tag{2.3.7}$$

between the set of linear maps $V \otimes_{\mathbb{k}} W \to U$ and the set of bilinear maps $V \times W \to U$ that is natural in U.[5]

Theorem 2.2.3 tells us that the natural isomorphism (2.3.7) is determined by an element of $\mathrm{Bilin}(V, W; V \otimes_{\mathbb{k}} W)$, i.e., by a bilinear map $\otimes\colon V \times W \to V \otimes_{\mathbb{k}} W$. One says that $V \otimes_{\mathbb{k}} W$ is the **universal vector space equipped with a bilinear map** from $V \times W$. The Yoneda lemma allows us to unpack what this means. Via the natural isomorphism (2.3.7), any bilinear map $f\colon V \times W \to U$ is associated to a linear map $f'\colon V \otimes_{\mathbb{k}} W \to U$. Consider the naturality square induced by f'.

$$\begin{array}{ccc}
\mathsf{Vect}_{\mathbb{k}}(V \otimes_{\mathbb{k}} W, V \otimes_{\mathbb{k}} W) & \xrightarrow{\cong} & \mathrm{Bilin}(V, W; V \otimes_{\mathbb{k}} W) \\
{\scriptstyle f'_*}\downarrow & & \downarrow{\scriptstyle f'_*} \\
\mathsf{Vect}_{\mathbb{k}}(V \otimes_{\mathbb{k}} W, U) & \xrightarrow[\cong]{} & \mathrm{Bilin}(V, W; U)
\end{array}$$

Tracing $1_{V \otimes_{\mathbb{k}} W}$ around the commutative square, we see that the bilinear map f factors uniquely through the bilinear map $\otimes$ along the linear map f'.

$$\begin{array}{ccc}
V \times W & \xrightarrow{\otimes} & V \otimes_{\mathbb{k}} W \\
 & {\scriptstyle f}\searrow & \downarrow{\scriptstyle \exists!\, f'} \\
 & & U
\end{array}$$

[5]This isomorphism is also natural in V and W.

This universal property tells us that the bilinear map $\otimes\colon V \times W \to V \otimes_{\Bbbk} W$ is initial in a category that we shall introduce below.

Indeed, the defining universal property of the tensor product gives a recipe for its construction. Supposing the vector space $V \otimes_{\Bbbk} W$ exists, consider its quotient by the vector space spanned by the image of the bilinear map[6] $- \otimes -$. By definition the quotient map $V \otimes_{\Bbbk} W \to V \otimes_{\Bbbk} W/\langle v \otimes w\rangle$ precomposes with $- \otimes -$ to yield the zero bilinear map. But the zero map $V \otimes_{\Bbbk} W \to V \otimes_{\Bbbk} W/\langle v \otimes w\rangle$ also has this property, so by the universal property of $V \otimes W$, these linear maps must agree. Because the quotient map is surjective, this implies that $V \otimes_{\Bbbk} W$ is isomorphic to the span of the vectors $v \otimes w$ for all $v \in V$ and $w \in W$ modulo the bilinearity relations satisfied by $- \otimes -$. This is of course the usual constructive definition.

The universal property of the tensor product of a pair of $\Bbbk$-vector spaces, more than simply characterizing the vector space up to isomorphisms, allows one to prove useful properties about it without ever appealing to a specific basis.

Lemma 2.3.8. *For any $\Bbbk$-vector spaces V and W, $V \otimes_{\Bbbk} W \cong W \otimes_{\Bbbk} V$.*

Proof. The functors $\mathrm{Bilin}(V, W; -)$ and $\mathrm{Bilin}(W, V; -)$ are naturally isomorphic: the natural isomorphism $\mathrm{Bilin}(V, W; U) \cong \mathrm{Bilin}(W, V; U)$ sends a bilinear map $f\colon V \times W \to U$ to the bilinear map $f^\sharp\colon W \times V \to U$ defined by $f^\sharp(w, v) := f(v, w)$. Composing natural isomorphisms, we see that the represented functors

$$\mathsf{Vect}_{\Bbbk}(V \otimes_{\Bbbk} W, -) \cong \mathrm{Bilin}(V, W; -) \cong \mathrm{Bilin}(W, V; -) \cong \mathsf{Vect}_{\Bbbk}(W \otimes_{\Bbbk} V, -) \tag{2.3.9}$$

are naturally isomorphic. By Lemma 2.3.1, a corollary of the Yoneda lemma, this natural isomorphism must arise from an isomorphism $V \otimes_{\Bbbk} W \cong W \otimes_{\Bbbk} V$ in $\mathsf{Vect}_{\Bbbk}$ between the representing objects. □

Moreover, the Yoneda lemma provides an explicit isomorphism $V \otimes_{\Bbbk} W \cong W \otimes_{\Bbbk} V$. The image of the identity linear transformation under the composite isomorphism

$$\mathsf{Vect}_{\Bbbk}(V \otimes_{\Bbbk} W, V \otimes_{\Bbbk} W) \cong \mathrm{Bilin}(V, W; V \otimes_{\Bbbk} W) \cong \mathrm{Bilin}(W, V; V \otimes_{\Bbbk} W) \cong \mathsf{Vect}_{\Bbbk}(W \otimes_{\Bbbk} V, V \otimes_{\Bbbk} W)$$

defines an isomorphism $\phi\colon W \otimes_{\Bbbk} V \xrightarrow{\cong} V \otimes_{\Bbbk} W$ so that the natural isomorphism (2.3.9) is defined by precomposing with ϕ. In Example 2.3.6, we saw that the identity linear map is sent under the bijection $\mathsf{Vect}_{\Bbbk}(V \otimes_{\Bbbk} W, V \otimes_{\Bbbk} W) \cong \mathrm{Bilin}(V, W; V \otimes_{\Bbbk} W)$ to the universal bilinear map $\otimes\colon V \times W \to V \otimes_{\Bbbk} W$. This in turn is sent to the bilinear map $W \times V \xrightarrow{(w,v) \mapsto (v,w)} V \times W \xrightarrow{\otimes} V \otimes_{\Bbbk} W$. Appealing to the universal property of $\otimes\colon W \times V \to W \otimes_{\Bbbk} V$, which defines the natural isomorphism $\mathrm{Bilin}(W, V; V \otimes_{\Bbbk} W) \cong \mathsf{Vect}_{\Bbbk}(W \otimes_{\Bbbk} V, V \otimes_{\Bbbk} W)$, this composite bilinear map is sent to the unique linear map ϕ that makes the diagram

$$\begin{array}{ccc} W \times V & \xrightarrow{\otimes} & W \otimes_{\Bbbk} V \\ {\scriptstyle (w,v)\mapsto(v,w)}\downarrow & & \downarrow{\scriptstyle \exists!\, \phi} \\ V \times W & \xrightarrow[\otimes]{} & V \otimes_{\Bbbk} W \end{array}$$

commute. This ϕ is the linear isomorphism we were seeking.

[6]The image of a bilinear map is not itself a sub-vector space, so closing under span is necessary.

Exercises.

Exercise 2.3.1. Use the defining universal property of the tensor product to prove that

(i) $\Bbbk \otimes_{\Bbbk} V \cong V$ for any $\Bbbk$-vector space V

(ii) $U \otimes_{\Bbbk} (V \otimes_{\Bbbk} W) \cong (U \otimes_{\Bbbk} V) \otimes_{\Bbbk} W$ for any $\Bbbk$-vector spaces U, V, W.

Exercise 2.3.2. The set B^A of functions from a set A to a set B represents the contravariant functor $\mathsf{Set}(- \times A, B) \colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$. In analogy with the universal bilinear map $\otimes$ of Example 2.3.6, describe the universal property of the element

$$\mathrm{ev} \colon B^A \times A \to B$$

in $\mathsf{Set}(B^A \times A, B) \cong \mathsf{Set}(B^A, B^A)$ that classifies the natural isomorphism.

2.4. The category of elements

The definition of a universal property given in 2.3.4 is somewhat unsatisfying. A functor $F \colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ encodes a universal property of a representing object $c \in \mathsf{C}$. By the Yoneda lemma, the data of the universal property is given entirely by the functor F and an element $x \in Fc$ inducing a natural isomorphism $\mathsf{C}(-, c) \cong F$. But how do we know whether a given functor F is representable? And if this is the case, how do we know whether an element $x \in Fc$ determines a natural isomorphism $\mathsf{C}(-, c) \cong F$, rather than simply a natural transformation $\mathsf{C}(-, c) \Rightarrow F$?

To answer these questions, we introduce the category of elements of a Set-valued functor.

Definition 2.4.1. The **category of elements** of a covariant functor $F \colon \mathsf{C} \to \mathsf{Set}$, denoted $\int F$, has

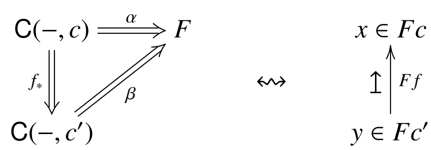
- pairs (c, x) where $c \in \mathsf{C}$ and $x \in Fc$ as objects
- a morphism $(c, x) \to (c', x')$ is a morphism $f \colon c \to c'$ in C so that $Ff(x) = x'$.

The category of elements has an evident forgetful functor $\int F \to \mathsf{C}$.

For a contravariant functor $F \colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$, the category of elements $\int F$ is defined to be the opposite of the category of elements of F, regarded as covariant functor of C^{op}. The convention is that, for both covariant and contravariant functors on C, the category of elements should have a canonical forgetful functor $\int F \to \mathsf{C}$. Explicitly, the category of elements of $F \colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ has

- pairs (c, x) where $c \in \mathsf{C}$ and $x \in Fc$ as objects
- a morphism $(c, x) \to (c', x')$ is a morphism $f \colon c \to c'$ in C so that $Ff(x') = x$.

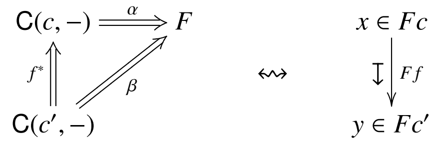
The Yoneda lemma supplies an alternative definition of the category of elements of F. In the contravariant case, an object in $\int F$ is a natural transformation $\alpha \colon \mathsf{C}(-, c) \Rightarrow F$ whose domain is a represented functor. A morphism from α to $\beta \colon \mathsf{C}(-, c') \Rightarrow F$ is a natural transformation $\mathsf{C}(-, c) \Rightarrow \mathsf{C}(-, c')$, i.e., by Corollary 2.2.6, a morphism $f \colon c \to c'$, so that the triangle of natural transformations



commutes.

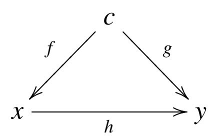
Similarly, in the covariant case, an object in $\int F$ is a natural transformation $\alpha \colon \mathsf{C}(c, -) \Rightarrow F$, and a morphism from α to $\beta \colon \mathsf{C}(c', -) \Rightarrow F$ is a natural transformation $\mathsf{C}(c', -) \to$

$\mathsf{C}(c,-)$, i.e., a morphism $c \to c'$, so that



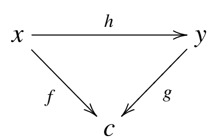
commutes. In this case, the direction of the morphism from α to β is less intuitive. The principle that the category of elements should always have a forgetful functor $\int F \to \mathsf{C}$ will ensure that the directions are chosen correctly.

Example 2.4.2. Objects in the category of elements of $\mathsf{C}(c,-)$ are morphisms $f\colon c \to x$ in C, i.e., the objects of $\int \mathsf{C}(c,-)$ are morphisms with domain c. A morphism from $f\colon c \to x$ to $g\colon c \to y$ is a morphism $h\colon x \to y$ so that $g = hf$; we say that h is a morphism under c.



This category has another name: it is the **slice category** c/C **under** the object $c \in \mathsf{C}$. The forgetful functor $c/\mathsf{C} \to \mathsf{C}$ sends a morphism $f\colon c \to x$ to its codomain and a commutative triangle to the leg opposite the object c.

Dually, $\int \mathsf{C}(-,c)$ is the **slice category** C/c **over** the object $c \in \mathsf{C}$. Objects are morphisms $f\colon x \to c$ with codomain c, and a morphism from f to $g\colon y \to c$ is a morphism $h\colon x \to y$ so that $gh = f$; h is a morphism over c.



The forgetful functor $\mathsf{C}/c \to \mathsf{C}$ projects onto the domain.

Note that $\mathsf{C}(-,c) \cong \mathsf{C}^{\mathrm{op}}(c,-)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$. And indeed $c/(\mathsf{C}^{\mathrm{op}})$, the category of elements of $\mathsf{C}(-,c)$ regarded as a covariant functor of C^{op}, is isomorphic to $(\mathsf{C}/c)^{\mathrm{op}}$, the opposite of the category of elements of the contravariant functor $\mathsf{C}(-,c)$.

Example 2.4.3. For a concrete category C, the objects in the category of elements for the forgetful functor $U\colon \mathsf{C} \to \mathsf{Set}$ are objects $c \in \mathsf{C}$ together with an element $x \in Uc$ in their underlying sets. Morphisms are maps in C whose underlying functions preserve the chosen elements. Extending previous notation introduced in certain special cases, we identify $\int U$ with C_*, the category of **based objects** in C and maps preserving the specified elements.

Our interest in the category of elements stems from the following result.

Proposition 2.4.4. *A covariant* Set*-valued functor is representable if and only if its category of elements has an initial object. Dually, a contravariant* Set*-valued functor is representable if and only if its category of elements has a terminal object.*

It's easy to see that this condition is necessary. A natural isomorphism $\mathsf{C}(c,-) \cong F$ induces an isomorphism of categories $\int F \cong \int \mathsf{C}(c,-) \cong c/\mathsf{C}$, and the latter has an initial object: the identity $1_c \in c/\mathsf{C}$. Thus, if F is representable, then $\int F$ has an initial object. The surprise is that it is also sufficient.

Proof. Let $F\colon \mathsf{C} \to \mathsf{Set}$ be a functor and suppose $(c, x) \in \int F$ is initial. We will show that the natural transformation $\Psi(x)\colon \mathsf{C}(c, -) \Rightarrow F$ defined by the Yoneda lemma is a natural isomorphism. For any $y \in Fd$, initiality says there exists a unique morphism $(c, x) \to (d, y)$, i.e., a unique morphism $f\colon c \to d$ in C so that $Ff(x) = y$. This says exactly that the component $\Psi(x)_d\colon \mathsf{C}(c, d) \to Fd$ is an isomorphism: in the proof of the Yoneda lemma, $\Psi(x)_d(f)$ was defined to be $Ff(x)$. Existence of the morphism $(c, x) \to (d, y)$ says that $\Psi(x)_d$ is surjective and uniqueness tells us that it is injective.

Reversing this argument, a natural isomorphism $\alpha\colon \mathsf{C}(c, -) \cong F$ defines an object $\alpha_c(1_c) \in Fc$ in the category of elements. The bijection $\alpha_d\colon \mathsf{C}(c, d) \xrightarrow{\cong} Fd$ says that for each object $(d, y) \in \int F$, there is a unique morphism $f\colon (c, \alpha_c(1_c)) \to (d, y)$. Thus, the element $(c, \alpha_c(1_c))$ defining the natural isomorphism α is initial in the category of elements. □

Recall that a representation for a functor $F\colon \mathsf{C} \to \mathsf{Set}$ consists of an object $c \in \mathsf{C}$ together with a natural isomorphism $\mathsf{C}(c, -) \cong F$. I.e., a representation for F is an initial object in $\int F$. Representations are not strictly unique: if c' is any object isomorphic to c, then a representation $\mathsf{C}(c, -) \cong F$ induces a representation $\mathsf{C}(c', -) \cong \mathsf{C}(c, -) \cong F$. However, they are unique in an appropriate category theoretic sense.

Proposition 2.4.5. *For any functor $F\colon \mathsf{C} \to \mathsf{Set}$, the full subcategory of $\int F$ spanned by its representations is either empty or a contractible groupoid.*

A **contractible groupoid** is a category that is equivalent to the terminal category $\mathbb{1}$.

Proof. If F is not representable then it has no representations and this category is empty. Otherwise, Proposition 2.4.4 implies that any representation defines an initial object in $\int F$. In any category, the subcategory spanned by the initial objects is either empty or is a contractible groupoid: there exists a unique (iso)morphism between any two initial objects. □

Example 2.4.6. The category of elements of a G-set $X\colon G \to \mathsf{Set}$ is the translation groupoid introduced in Example 1.5.13.(v). Assuming X is non-empty, this groupoid is contractible if and only if for each pair of elements $x, y \in X$, there is a unique $g \in G$ so that $g \cdot x = y$. Existence tells us that the action of G on X is transitive. Uniqueness, in the case $x = y$, tells us that it is free. Returning to Example 2.3.5, we conclude that X is representable if and only if it defines a G-torsor.

In Definition 2.3.4, we said that a universal property for an object $c \in \mathsf{C}$ was given either by a contravariant functor F together with a representation $\mathsf{C}(-, c) \cong F$ or by a covariant functor F together with a representation $\mathsf{C}(c, -) \cong F$. The representations define a natural characterization of the maps into (in the contravariant case) or out of (in the covariant case) the object c. Lemma 2.3.1 implies that a universal property characterizes the object $c \in \mathsf{C}$ up to isomorphism. More precisely, there is a unique isomorphism between c and another object c' representing F that commutes with the chosen representations. This can be proven directly from the Yoneda lemma or extracted from Proposition 2.4.5.

In practice, in such contexts one says that "c is the universal object in C with a x," where x is the element of Fc classifying the natural isomorphism that defines the representation. The phrase **universal object** typically means that whatever data is being described is either initial or terminal in the appropriate category. This category frequently turns out to be the category of elements for the corresponding representable functor.

Example 2.4.7. For example, a set X with an endomorphism $f\colon X \to X$ and a distinguished element x_0 is called a **discrete dynamical system**. This data allows one to consider the discrete-time evolution of the initial element x_0, a sequence defined by $x_{n+1} = f(x_n)$. The principle of mathematical recursion makes clear that the natural numbers $\mathbb{N}$, the successor function $\mathbb{N} \to \mathbb{N}$, and the element $0 \in \mathbb{N}$ define the universal discrete dynamical system: there is a unique function $r\colon \mathbb{N} \to X$ so that $r(n) = x_n$, i.e., so that $r(0) = x_0$ and so that the diagram

(2.4.8)
$$\begin{array}{ccc} \mathbb{N} & \xrightarrow{s} & \mathbb{N} \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle r} \\ X & \xrightarrow[f]{} & X \end{array}$$

commutes.

The category of discrete dynamical systems is the category of elements for the functor $U\colon \mathsf{End} \to \mathsf{Set}$ whose domain is the category of sets equipped with an endomorphism and whose maps are functions so that the diagram analogous to (2.4.8) commutes. This functor is represented by the object $(\mathbb{N}, s\colon \mathbb{N} \to \mathbb{N})$ and the representation is defined by the element $0 \in \mathbb{N} = U(\mathbb{N}, s)$. Proposition 2.4.4 then implies the universal property just observed: that $(\mathbb{N}, s\colon \mathbb{N} \to \mathbb{N}, 0 \in \mathbb{N})$ is the universal discrete dynamical system.

Examples 2.4.9.

(i) The objects in the category of elements of the contravariant powerset functor are pairs $(A' \subset A)$. A morphism $f\colon (A' \subset A) \to (B' \subset B)$ is a function $f\colon A \to B$ so that $f^{-1}(B') = A'$. The terminal object, corresponding to the representation described in Example 2.1.3.(i), is the set Ω of two elements $\{\top, \bot\}$ with the distinguished singleton subset $\{\top\} \subset \Omega$. This is the universal set equipped with a subset.[7] For any $(A' \subset A)$, there is a unique function $h\colon A \to \Omega$ so that $h^{-1}(\top) = A'$. Note that even if we begin with a skeletal category of sets, $(\{\top\} \subset \Omega)$ is not the unique terminal object in $\int P$. It is isomorphic to $(\{\bot\} \subset \Omega)$.

(ii) An object in the category of elements of the forgetful functor $U\colon \mathsf{Vect}_{\Bbbk} \to \mathsf{Set}$ is a vector v in some $\Bbbk$-vector space V. A morphism $(V, v) \to (W, w)$ is a linear map $T\colon V \to W$ that carries v to w. The element $(\Bbbk, 1)$, which represents U, defines an initial object but it's not the only one. Any non-zero scalar $c \in \Bbbk$ defines a linear isomorphism $c \cdot -\colon \Bbbk \to \Bbbk$. Thus, the pairs $(\Bbbk, c)$ are also initial in the category of elements. More generally, any 1-dimensional vector space V and non-zero vector v defines an initial object. Fixing such a pair, linear maps $V \to W$ are in bijection with vectors $w \in W$, taken to be the image of v.

(iii) Objects in the category of elements of $\mathrm{Bilin}(V, W; -)$ are bilinear maps $f\colon V \times W \to U$, for some $\Bbbk$-vector space U. Morphisms are linear maps $T\colon U \to U'$ so that the diagram of functions

$$\begin{array}{ccc} V \times W & \xrightarrow{f} & U \\ & {\scriptstyle f'}\searrow & \downarrow{\scriptstyle T} \\ & & U' \end{array}$$

commutes, i.e., so that the bilinear map f' is the composite of the bilinear map f and the linear map T. The universal property of the universal bilinear map $\otimes\colon V \times W \to V \otimes_{\Bbbk} W$ described in Example 2.3.6 says exactly that $\otimes$ is initial in this category.

[7] In the topos of sets, $\{\top\} \subset \Omega$ is called the **subobject classifier**.

(iv) An object in the category of elements of the functor $U(-)^n : \mathsf{Group} \to \mathsf{Set}$ consists of a group G together with an n-tuple of elements $g_1, \ldots, g_n \in G$. A morphism $(g_1, \ldots, g_n \in G) \to (h_1, \ldots, h_n \in H)$ is a homomorphism $\phi : G \to H$ so that $\phi(g_i) = h_i$ for all i. The universal group with n elements, which we will denote by F_n, is some group with specified elements $x_1, \ldots, x_n \in F_n$ so that for any $(g_1, \ldots, g_n \in G)$ there is a unique group homomorphism $\phi : F_n \to G$ so that $\phi(x_i) = g_i$. The free group on n generators $x_1, \ldots, x_n$ has this property. Every element in F_n is generated by the x_i, so choices of n elements $\phi(x_i) \in G$ determines the entire map $\phi : F_n \to G$. Moreover, there are no relations between the x_i in F_n so any choices are permitted.

(v) An object in the category of elements of the functor $U(-)^* : \mathsf{Vect}_{\Bbbk}^{\mathrm{op}} \to \mathsf{Set}$ is a vector space with a dual vector. That is, an object is simply a linear map $f : V \to \Bbbk$. A morphism from $f : V \to \Bbbk$ to $g : W \to \Bbbk$ is a linear map $T : V \to W$ so that $f = gT$. Immediately from this definition it is clear that the category of elements of $U(-)^*$ is the slice category $\mathsf{Vect}_{\Bbbk}/\Bbbk$. Thus the identity on $\Bbbk$ is terminal, i.e., $1 : \Bbbk \to \Bbbk$ is the universal dual vector.

(vi) An object in the category of elements of $U : \mathsf{Ring} \to \mathsf{Set}$ is a unital ring R with an element $r \in R$. Maps are ring homomorphisms preserving the chosen elements. The initial object is $x \in \mathbb{Z}[x]$. In the category $\int U$, $x \in \mathbb{Z}[x]$ has no non-identity endomorphisms: initial objects never do. But in the category of rings, the universal property tells us that maps $\mathbb{Z}[x] \to \mathbb{Z}[x]$ are classified by polynomials with integer coefficients, i.e., by elements $p(x) \in \mathbb{Z}[x]$. By Corollary 2.2.6, all natural endomorphisms of $U : \mathsf{Ring} \to \mathsf{Set}$ must have components $R \to R$ defined in this manner: i.e., the component of any natural endomorphism of rings is a map $p(x) : R \to R$ defined by $r \mapsto p(r)$, where p is an integer polynomial.

Definitions via universal properties will be an important theme in the coming chapters. To whet the reader's appetite, let us consider a basic example. Fixing two objects A, B in a locally small category C, we define a functor

$$\mathsf{C}(-, A) \times \mathsf{C}(-, B) : \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$$

that carries an object X to the set $\mathsf{C}(X, A) \times \mathsf{C}(X, B)$ whose elements are pairs of maps $a : X \to A$ and $b : X \to B$ in C. What would it mean for this functor to be representable?

Exercises.

EXERCISE 2.4.1. Explain how duality can be used to convert the proof that a covariant functor is representable if and only if its category of elements has an initial object into a proof that a contravariant functor is representable if and only if its category of elements has a terminal object.

EXERCISE 2.4.2. Given functors $F : \mathsf{D} \to \mathsf{C}$ and $G : \mathsf{E} \to \mathsf{C}$, the **comma category** $F \downarrow G$ has as objects, triples $(d \in \mathsf{D}, e \in \mathsf{E}, f : Fd \to Ge \in \mathsf{C})$, and as morphisms $(d, e, f) \to (d', e', f')$, a pair of morphisms $(h : d \to d', k : e \to e')$ so that the square

$$\begin{array}{ccc} Fd & \xrightarrow{f} & Ge \\ {\scriptstyle Fh}\downarrow & & \downarrow{\scriptstyle Gk} \\ Fd' & \xrightarrow[f']{} & Ge' \end{array}$$

commutes in C. For example, given an object $c\colon \mathbb{1} \to \mathsf{C}$, we have $c \downarrow 1_{\mathsf{C}} \cong c/\mathsf{C}$ and $1_{\mathsf{C}} \downarrow c \cong \mathsf{C}/c$.

Show that for $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$, the category of elements $\int F$ is isomorphic to the comma category $y \downarrow F$ defined relative to the Yoneda embedding $y\colon \mathsf{C} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ and the object $F\colon \mathbb{1} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$.

Exercise 2.4.3. Given $F\colon \mathsf{C} \to \mathsf{Set}$, show that $\int F$ is isomorphic to the comma category $* \downarrow F$ of the singleton set $*\colon \mathbb{1} \to \mathsf{Set}$ over the functor $F\colon \mathsf{C} \to \mathsf{Set}$.

Exercise 2.4.4. Characterize the terminal objects of C/c.

Exercise 2.4.5. Fixing two objects A, B in a locally small category C, we define a functor

$$\mathsf{C}(-, A) \times \mathsf{C}(-, B)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$$

that carries an object X to the set $\mathsf{C}(X, A) \times \mathsf{C}(X, B)$ whose elements are pairs of map $a\colon X \to A$ and $b\colon X \to B$ in C. What would it mean for this functor to be representable?

Exercise 2.4.6. For a locally small category C, regard the two-sided represented functor $\mathrm{Hom}(-,-)\colon \mathsf{C}^{\mathrm{op}} \times \mathsf{C} \to \mathsf{Set}$ as a covariant functor of its domain. What is the category of elements of Hom?

CHAPTER 3

Limits and Colimits

> ... whenever new abstract objects are constructed in a specified way out of given ones, it is advisable to regard the construction of the corresponding induced mappings on these new objects as an integral part of their definition.
>
> Eilenberg and Mac Lane, "General theory of natural equivalences"

From a very simple topological space, the real line $\mathbb{R}$ with its usual metric, one can build a wide variety of new topological spaces. Taking **products** of $\mathbb{R}$ with itself, one defines the Euclidean spaces $\mathbb{R}^n$, in both finite and infinite dimensions. The space $\mathbb{R}^n$ has interesting **subspaces** including the n-disk D^n and the $(n-1)$-sphere S^{n-1} bounding it. From the sphere S^n, one can define real projective space $\mathbb{R}P^n$ as a **quotient**. And from spheres and disks one can build torii, the Möbius band, the Klein bottle, and indeed any **cell complex** through a sequence of **gluing** constructions, in which disks are attached to an existing space along their boundary spheres. In each case, the newly constructed object is a particular set equipped with a specific topology. Surprisingly, all of these topologies can be defined in a uniform way, via a universal property that characterizes the newly constructed space either as a **limit** or a **colimit** of a particular diagram in the category of topological spaces.

Limits and colimits can be defined in any category. Special cases include constructions of the infimum and supremum, free products, cartesian products, direct sums, kernels, cokernels, fiber products, amalgamated free products, inverse limits of sequences, and unions, among many others. In this chapter, we begin by introducing the abstract notions of limit and colimit and then turn our focus to practical results, which describe how more complicated limits and colimits can be built out of simpler ones. This theme will continue in the next two chapters, where we meet cases in which limits or colimits in one category can be constructed from limits or colimits in another.

3.1. Limits and colimits as universal cones

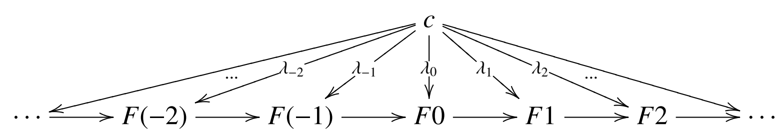
Recall that a **diagram** of **shape** J in a category C is a functor $F\colon \mathsf{J} \to \mathsf{C}$ whose domain is a small category. The **limit** of F, if it exists, is most simply described as the universal cone over the diagram F, while the **colimit** is the universal cone under F.

To define a cone, we need one preliminary notion.

Definition 3.1.1. For any object $c \in \mathsf{C}$ and any small category J there is a **constant functor** $c\colon \mathsf{J} \to \mathsf{C}$ that sends every object of J to $c \in \mathsf{C}$ and every morphism in J to the identity morphism 1_c. Indeed, there is an embedding $\Delta\colon \mathsf{C} \to \mathsf{C}^{\mathsf{J}}$ that sends an object c to the constant functor at c and a morphism $f\colon c \to c'$ to the **constant natural transformation**, in which each component is defined to be the morphism f.

Definition 3.1.2. A **cone over** a diagram $F\colon \mathsf{J} \to \mathsf{C}$ with **summit** or **apex** c, an object in C, is a natural transformation $\lambda\colon c \Rightarrow F$ whose domain is the constant functor at c. The components $(\lambda_j\colon c \to Fj)_{j\in\mathsf{J}}$ are called the **legs** of the cone. A **cone under** F with **nadir** c is a natural transformation $\lambda\colon F \Rightarrow c$.

Cones under a diagram are also called **cocones**: a cone under $F\colon \mathsf{J} \to \mathsf{C}$ is precisely a cone over $F\colon \mathsf{J}^{\mathrm{op}} \to \mathsf{C}^{\mathrm{op}}$, so a cocone is the dual notion of a cone. However, this author finds the terminology of "under" and "over" to be more evocative. For example, if F is a diagram of shape $(\mathbb{Z}, \leq)$, the poset category, then a cone over F with summit c is given by a family of morphisms $\lambda_n\colon c \to Fn$ so that for each $n \leq m$ the triangle given by λ_n, λ_m, and $Fn \to Fm$ commutes



The limit of F, if it exists, is the universal cone over F. As described in Chapter 2 there are two ways to make this precise. We define the limit as a representation for a particular contravariant functor or as a terminal object in its category of elements.

Definition 3.1.3 (limits and colimits I). For any diagram $F\colon \mathsf{J} \to \mathsf{C}$, there is a functor

$$\mathrm{Cone}(-, F)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$$

that sends c to the set of cones over F with summit c. We leave a description of its action on morphisms to Exercise 3.1.1. This functor is isomorphic to $\mathrm{Hom}(\Delta(-), F)$, the restriction of the hom functor for the category C^{J} along the constant functor embedding defined in 3.1.1. A **limit** of F is a representation for $\mathrm{Cone}(-, F)$. By the Yoneda lemma, a limit consists of an object $\lim F \in \mathsf{C}$ together with a cone $\lambda\colon \lim F \Rightarrow F$, called the **limit cone**, defining the natural isomorphism

$$\mathsf{C}(-, \lim F) \cong \mathrm{Cone}(-, F).$$

Dually, there is a functor

$$\mathrm{Cone}(F, -)\colon \mathsf{C} \to \mathsf{Set}$$

that sends c to the set of cones under F with nadir c. This functor is isomorphic to $\mathrm{Hom}(F, \Delta(-))$. A **colimit** of F is a representation for $\mathrm{Cone}(F, -)$. By the Yoneda lemma, a colimit consists of an object $\mathrm{colim}\, F \in \mathsf{C}$ together with a cone $\lambda\colon F \Rightarrow \mathrm{colim}\, F$, called the **colimit cone**, defining the natural isomorphism

$$\mathsf{C}(\mathrm{colim}\, F, -) \cong \mathrm{Cone}(F, -).$$

Applying Proposition 2.4.4, limits and colimits may also be defined to be, respectively, terminal and initial objects in the appropriate categories of elements.

Definition 3.1.4 (limits and colimits II). For any diagram $F\colon \mathsf{J} \to \mathsf{C}$, a **limit** is a terminal object in the category of cones over F, i.e., in the category $\int \mathrm{Cone}(-, F)$; again this data consists an object, called the limit, together with a specified **limit cone**. An object in the category of cones over F is a cone over F, with any nadir. A morphism from a cone $\lambda\colon c \Rightarrow F$ to a cone $\mu\colon d \Rightarrow F$ is a morphism $f\colon c \to d$ in C so that for each $j \in \mathsf{J}$, $\mu_j f = \lambda_j$. In other words, a morphism of cones is a map between the summits so that each leg of the domain cone factors through the corresponding leg of the codomain cone along this map. The forgetful functor $\int \mathrm{Cone}(-, F) \to \mathsf{C}$ takes a cone to its summit.

Dually, a **colimit** is an initial object in the category of cones under F, i.e., in the category $\int \mathrm{Cone}(F, -)$; once more, this data is comprised of the colimit object together with a specified **colimit cone**. An object in the category of cones under F is a cone under F, with any summit. A morphism from a cone $\lambda\colon F \Rightarrow c$ to a cone $\mu\colon F \Rightarrow d$ is a morphism $f\colon c \to d$ so that for each $j \in \mathsf{J}$, $\mu_j = f\lambda_j$. In other words, a morphism of cones is a map between the nadirs so that each leg of the codomain cone factors through the corresponding leg of the domain cone along this map. The forgetful functor $\int \mathrm{Cone}(F, -) \to \mathsf{C}$ again takes a cone to its nadir.

The data of a diagram together with a limit cone over it will be called a **limit diagram** and the data of a diagram together with a colimit cone under it will be called a **colimit diagram**. Limit and colimit diagrams can also be characterized abstractly.

Definition 3.1.5. A **limit diagram** is a commutative diagram indexed by a category with an initial object such that the diagram has the following property: the images of the maps whose domain is the initial object and whose codomain is a distinct object define a limit cone over the diagram obtained by restricted to the full subcategory on the remaining objects.

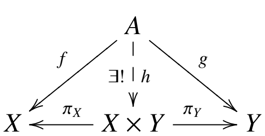
There are special names for limits and colimits with certain diagram shapes.

Example 3.1.6. A **product** is a limit of a diagram indexed by a discrete category, with only identity morphisms. A diagram in C indexed by a discrete category (a set) J is simply a collection of objects $F_j \in \mathsf{C}$ indexed by the elements $j \in J$. A cone over this diagram is a J-indexed family of morphisms $\lambda_j\colon c \to F_j$, subject to no further conditions. The limit is typically denoted by $\prod_j F_j$ and the legs of the limit cone are maps $\pi_k\colon \prod_j F_j \to F_k$ called projections.

For instance, the product of spaces X and Y is a space $X \times Y$ equipped with continuous projection functions

$$X \xleftarrow{\pi_X} X \times Y \xrightarrow{\pi_Y} Y$$

satisfying the following universal property: for any other space A with continuous maps $f\colon A \to X$ and $g\colon A \to Y$, there is a unique continuous function $h\colon A \to X \times Y$ so that the diagram



commutes. Taking A to be the set $X \times Y$ equipped with various topologies, we see that $X \times Y$ is defined to be the coarsest topology on the cartesian product of the underlying sets of X and Y so that the projection functions π_X and π_Y are continuous.[1]

Example 3.1.7. A **terminal object** is often regarded as a trivial special case of a product, where the indexing category is empty. A cone over an empty diagram is just an object in the target category, and any morphism defines a map of cones. The category of elements is just the category itself and so the limit, a terminal object in this category, is just a terminal object in the sense of Definition 1.6.16.

Example 3.1.8. An **equalizer** is a limit of a diagram indexed by the **parallel pair**, the category $\bullet \rightrightarrows \bullet$ with two objects and two parallel non-identity morphisms. A diagram of this shape is simply a parallel pair of morphisms $f, g\colon A \rightrightarrows B$ in the target category C.

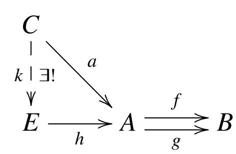
[1] Here **coarsest** means "having the fewest open and closed sets" while **finest** means "having the most open and closed sets."

The data of a cone over this diagram with summit C is comprised of a pair of morphisms $a\colon C \to A$ and $b\colon C \to B$ so that $fa = b$ and $ga = b$; these two equations correspond to the naturality conditions for each of the two non-identity morphisms in the indexing category. Together, they assert that $fa = ga$; the morphism b is necessarily equal to this common composite. Thus, a cone over a parallel pair $f, g\colon A \rightrightarrows B$ is given by a single morphism $a\colon C \to A$ so that $fa = ga$.

The equalizer $h\colon E \to A$ is the universal arrow with this property. In this case, the limit diagram

$$E \xrightarrow{h} A \overset{f}{\underset{g}{\rightrightarrows}} B$$

is often called an **equalizer diagram**. The universal property asserts that given any $a\colon C \to A$ that equalizes the pair f and g, there exists a unique factorization $k\colon C \to E$ of a through h.



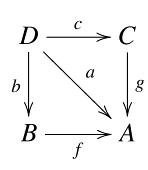
For instance, the equalizer of a group homomorphism $\phi\colon G \to H$ and the trivial homomorphism $e\colon G \to H$, sending every element of G to the identity in H, is the kernel of ϕ, and the leg of the limit cone is the inclusion $\ker\phi \hookrightarrow G$. Indeed, the map from an equalizer into the domain of the parallel pair that it equalizes is always a monomorphism. See Exercise 3.1.3.

EXAMPLE 3.1.9. A **pullback** is a limit of a diagram indexed by the poset category

$$\bullet \to \bullet \leftarrow \bullet$$

comprised of two non-identity morphisms with common codomain. Writing f and g for the image of a diagram of this shape in a category C, a cone with summit D consists of a triple of morphisms, one for each object in the indexing category, so that both triangles in the diagram

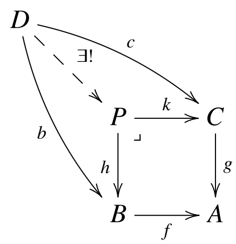
(3.1.10)



commute; the two triangles again represent the two naturality conditions corresponding to the morphisms in the indexing category. The leg a asserts that gc and fb have a common composite. Thus, the data of a cone over $B \xrightarrow{f} A \xleftarrow{g} C$ may be described more simply as a pair of morphisms $B \xleftarrow{b} D \xrightarrow{c} C$ defining a commutative square.

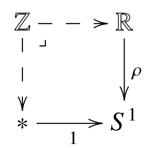
The pullback is a commutative square $fh = gk$ with the following universal property: given any commutative square (3.1.10), there is a unique factorization of its legs through the summit of the pullback cone:

(3.1.11)



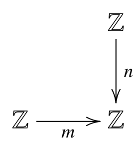
The symbol "⌟" indicates that the square $gk = fh$ is a pullback, i.e., is a limit diagram, and not simply commutative. Pullbacks are also called **fiber products** and are frequently denoted by $B \times_A C$. We will explore the precise relationship between pullbacks and products soon.

When the map f in (3.1.11) represents an "element" of the object A, such as when its domain represents an "underlying set" functor, the pullback P defines the **fiber** of the map $g: C \to A$ over the element $f \in A$. For instance, consider the continuous quotient map $\rho: \mathbb{R} \to S^1$ that sends a real number t to the point $e^{2\pi i t}$ of S^1, thought of as the unit circle in the complex plane. The pullback of the diagram



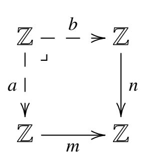
is the **fiber** of ρ over the point $1 \in S^1$. This is the discrete subspace $\mathbb{Z} \subset \mathbb{R}$.

For another example, recall that maps $\mathbb{Z} \to \mathbb{Z}$ of abelian groups correspond to elements $n \in \mathbb{Z}$, because $\mathbb{Z}$ represents the forgetful functor $U: \mathsf{Ab} \to \mathsf{Set}$. Elements of the pullback of



are pairs of integers x and y so that $nx = my$. From this description, assuming m and n are not both zero, it follows that the pullback is isomorphic to the abelian group $\mathbb{Z}$ and the legs of the pullback cone

(3.1.12)

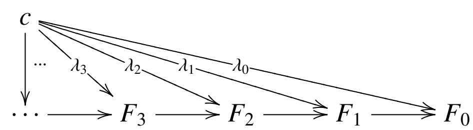


are defined to be the unique integers so that $ma = nb$ is the least common multiple of n and m.

Example 3.1.13. The limit of a diagram indexed by the category ω^{op} is called an **inverse limit** of a tower or a sequence of morphisms. On account of this example, the term "inverse limit" is sometimes used to mean a limit of an arbitrary shape. A diagram indexed by ω^{op} consists of a sequence of objects and morphisms

$$\cdots \longrightarrow F_3 \longrightarrow F_2 \longrightarrow F_1 \longrightarrow F_0$$

together with composites and identities, which are not displayed. A cone over this diagram is an extension of this data to a diagram of shape $(\omega + 1)^{\mathrm{op}}$. Explicitly, a cone consists of a new object "all the way to the left" together with morphisms making every triangle commute:



The inverse limit is the terminal cone.

For instance, the **p-adic integers** are defined to be the inverse limit of the following diagram of rings:

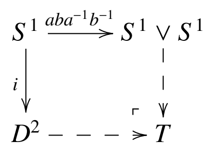
$$\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n \cdots \twoheadrightarrow \mathbb{Z}/p^4 \twoheadrightarrow \mathbb{Z}/p^3 \twoheadrightarrow \mathbb{Z}/p^2 \twoheadrightarrow \mathbb{Z}/p$$

where the maps in the diagram are the canonical quotient maps.

Examples 3.1.6, 3.1.7, 3.1.8, 3.1.9, and 3.1.13 dualize to define **coproducts**, **initial objects**, **coequalizers**, **pushouts**, and **direct limits** of sequences as colimits of diagrams indexed by discrete categories, the empty category, the parallel pair, the poset category $\bullet \leftarrow \bullet \rightarrow \bullet$, and the ordinal ω, respectively.

For instance, the coproduct of the based space S^1 with itself in the category Top_* defines a space $S^1 \vee S^1$ called the figure eight, or the wedge of two circles. This space can also be defined to be the pushout of the diagram $S^1 \leftarrow * \rightarrow S^1$ of unbased spaces. A further pushout diagram in Top

(3.1.14)



defines the torus $T \cong S^1 \times S^1$. Here i is the inclusion of the circle as the boundary of the disk and the map $aba^{-1}b^{-1}$ is the loop in $S^1 \vee S^1$ that wraps once around one circle, then once around the other, then again around the first but in the reversed orientation, and then again around the second but in the reversed orientation. Dualizing the convention introduced in Example 3.1.9, the symbol "$\ulcorner$" indicates that the commutative square (3.1.14) is a pushout, i.e., is a colimit diagram.

The coequalizer of a group homomorphism $\phi \colon G \to H$ paired with the trivial homomorphism $e \colon G \to H$ defines the **cokernel**, the quotient of H by the normal subgroup generated by the image of ϕ. The colimit of a sequence of sets and inclusions

$$X_0 \hookrightarrow X_1 \hookrightarrow X_2 \hookrightarrow \cdots$$

can be understood to be their union $\cup_n X_n$.

The examples mentioned above serve to illustrate the pervasiveness of limits and colimits among interesting mathematical objects, but how does one prove that a particular diagram has a limit or colimit and construct one if it exists? There are two general strategies. The first is to guess and check. A conjectured limit or colimit cone, such as displayed in (3.1.12), can often be shown to have the appropriate universal property via an elementary argument.

A second strategy is to construct a limit of a more complicated diagram out of limits of simpler diagrams, whose constructions are easier to understand. In the next section, we begin to explore this viewpoint in the category of sets.

Exercises.

Exercise 3.1.1. For a fixed diagram $F \in \mathsf{C}^{\mathsf{J}}$, describe the actions of the cone functors $\mathrm{Cone}(-, F)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ and $\mathrm{Cone}(F, -)\colon \mathsf{C} \to \mathsf{Set}$ on morphisms in C.

Exercise 3.1.2. Prove that the category of cones over $F \in \mathsf{C}^{\mathsf{J}}$ is isomorphic to the comma category $\Delta \downarrow F$ formed from the constant functor $\Delta\colon \mathsf{C} \to \mathsf{C}^{\mathsf{J}}$ and the functor $F\colon \mathbb{1} \to \mathsf{C}^{\mathsf{J}}$. Argue by duality the category of cones under F is the comma category $F \downarrow \Delta$.

Exercise 3.1.3. Prove that if

$$E \xrightarrow{\;h\;} A \underset{g}{\overset{f}{\rightrightarrows}} B$$

is an equalizer diagram then h is a monomorphism.

Exercise 3.1.4. Prove that if

$$\begin{array}{ccc} P & \overset{k}{\rightarrowtail} & C \\ {\scriptstyle h}\downarrow & \lrcorner & \downarrow{\scriptstyle g} \\ B & \underset{f}{\rightarrowtail} & A \end{array}$$

is a pullback square and f is a monomorphism then k is a monomorphism.

Exercise 3.1.5. Consider a commutative rectangle

$$\begin{array}{ccccc} \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet \\ \downarrow & & \downarrow & \lrcorner & \downarrow \\ \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet \end{array}$$

whose right-hand square is a pullback. Show that the left-hand square is a pullback if and only if the composite rectangle is a pullback.

Exercise 3.1.6. Show that if J has an initial object then the limit of any functor indexed by J is the value of that functor at an initial object.

3.2. Limits in the category of sets

Consider a diagram $F\colon \mathsf{J} \to \mathsf{Set}$. A limit is a representation

$$\mathrm{Hom}(X, \lim F) \cong \mathrm{Cone}(X, F)$$

of the functor that sends a set X to the set of cones over F with summit X. Specializing to a fixed singleton set 1, we see that

$$\lim F \cong \mathrm{Hom}(1, \lim F) \cong \mathrm{Cone}(1, F), \tag{3.2.1}$$

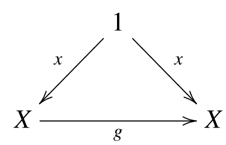
because 1 represents the identity functor $\mathsf{Set} \to \mathsf{Set}$. We can take (3.2.1) as a *definition* of the limit object, for any diagram F valued in the category of sets: the limit is the set of cones over F with summit 1. The legs of the limit cone are functions $\lambda_j\colon \lim F \to Fj$ for each $j \in \mathsf{J}$. Here the value of λ_j at a cone $\mu\colon 1 \Rightarrow F$ is the element $\mu_j\colon 1 \to Fj$ of Fj. Checking that this data has the necessary universal property enables one to prove:

Theorem 3.2.2. *The category* Set *has all limits indexed by small categories. Explicitly, the limit of* $F\colon \mathsf{J} \to \mathsf{Set}$ *is the set of cones over* F *with summit* 1.

Proof. Guided by (3.2.1), let $\lim F$ denote the set of cones over F with summit 1. It is easy to verify that the functions $\lambda_j\colon \lim F \to Fj$ define a cone over F with summit $\lim F$. Consider a cone $\zeta\colon X \Rightarrow F$. We must show that ζ factors uniquely through $\lambda\colon \lim F \Rightarrow F$ along a function $k\colon X \to \lim F$. For each element $x \in X$, thought of as a function $x\colon 1 \to X$, we have a cone $\zeta x\colon 1 \Rightarrow F$. Define $k(x) \in \lim F = \mathrm{Cone}(1, F)$ to be this cone ζx. By the definition of the limit cone $\lambda_j k(x) = (\zeta x)_j = \zeta_j x$, so the cone ζ indeed factors along k through the cone λ. This calculation also reveals that our definition $k(x) = \zeta x$ was necessary, verifying that $\lim F$ has the required universal property. □

It is necessary in 3.2.2 that we restrict to diagrams indexed by small categories. If C is locally small and J is small then the functor category C^{J} is again locally small. Specializing to $\mathsf{C} = \mathsf{Set}$ this proves that the collection of cones over F with summit 1 forms a mere set and not a proper class. A category admitting all limits indexed by small categories is called **complete**.

EXAMPLE 3.2.3. The limit of a left G-set $X\colon G \to \mathsf{Set}$ is the set of cones with summit 1. A map $x\colon 1 \to X$ defines a cone over X if and only if the diagram



commutes for all $g \in G$, i.e., if and only if $gx = x$ for all g in G. In this way we see that $\lim X = X^G$, the set of G-**fixed points** of X.

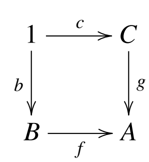
EXAMPLE 3.2.4. The product of sets A_j indexed by the elements $j \in J$ of some set is the set of cones over this collection with summit 1. Via the correspondence between $a\colon 1 \to A_j$ and $a \in A_j$, we see a cone over a discrete diagram of sets with summit 1 is just a tuple of elements of the sets, which is exactly how we define the cartesian product.

EXAMPLE 3.2.5. The terminal object in Set is just the set of cones over the empty diagram with summit 1. There is a unique such cone, which proves that the terminal object is again a singleton set.

EXAMPLE 3.2.6. Given a parallel pair of functions $f, g\colon X \rightrightarrows Y$, their *equalizer* is the set E defined by

$$E := \{x \in X \mid f(x) = g(x)\}.$$

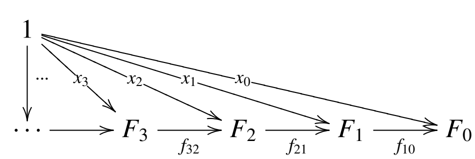
EXAMPLE 3.2.7. Elements of the pullback of functions $B \xrightarrow{f} A \xleftarrow{g} C$ are cones over this diagram with summit 1, i.e., are commutative squares



The data of a cone of this form consists of a pair of elements $b \in B$ and $c \in C$ so that $f(b) = g(c)$. Thus

$$B \times_A C = \{(b, c) \in B \times C \mid f(b) = g(c)\}$$

EXAMPLE 3.2.8. Elements of the limit of a diagram $F\colon \omega^{\mathrm{op}} \to \mathsf{Set}$ are cones



are tuples of elements $(x_n \in F_n)_{n\in\omega}$ making each triangle commute. Thus we see that

$$\lim F = \{(x_n)_n \in \prod_n F_n \mid f_{n,n-1}x_n = x_{n-1}\}.$$

Comparing Example 3.2.7 with Examples 3.2.4 and 3.2.6 it looks like the pullback may be constructed as an equalizer of a diagram involving a product. Namely, the set $B \times_A C$ is the equalizer of the diagram

$$B \times_A C \longrightarrow B \times C \underset{(b,c)\mapsto g(c)}{\overset{(b,c)\mapsto f(b)}{\rightrightarrows}} A$$

Similarly, Example 3.2.8 defines the limit of a tower to be a subset of a product. These are special cases of a much more general result.

Theorem 3.2.9. *Any limit in* Set *may be expressed as an equalizer of a map between products.*

Proof. Elements of the limit of $F\colon \mathsf{J} \to \mathsf{Set}$ are cones with summit 1 over F. The data of such a cone consists of an element λ_j in each set Fj, indexed by the objects $j \in \mathsf{J}$. The conditions that make this family of elements into a cone over F are indexed by the (non-identity) morphisms in J: for each morphism f we demand that

(3.2.10)
$$\begin{array}{ccc} & 1 & \\ {\scriptstyle\lambda_{\mathrm{dom}\,f}}\swarrow & & \searrow{\scriptstyle\lambda_{\mathrm{cod}\,f}} \\ F(\mathrm{dom}\,f) & \xrightarrow[Ff]{} & F(\mathrm{cod}\,f) \end{array} \qquad \text{i.e.,} \qquad Ff(\lambda_{\mathrm{dom}\,f}) = \lambda_{\mathrm{cod}\,f}.$$

The "data" will define the domain of the equalizer diagram and the "conditions" will define the codomain, as in Example 3.2.6.

Explicitly, $\lim F$ is the equalizer of the diagram

$$\lim F \rightarrowtail \prod_{j\in \mathrm{ob}\,\mathsf{J}} Fj \underset{d}{\overset{c}{\rightrightarrows}} \prod_{f\in \mathrm{mor}\,\mathsf{J}} F(\mathrm{cod}\,f)$$

in which the morphisms c and d remain to be defined. The idea is that an element $(\lambda_j)_j \in \prod_j Fj$, thought of as the legs of a cone with summit 1 over F, is sent by c to the element $(\lambda_{\mathrm{cod}\,f})_f \in \prod_f F(\mathrm{cod}\,f)$ and is sent by d to the element $(Ff(\lambda_{\mathrm{dom}\,f}))_f \in \prod_f F(\mathrm{cod}\,f)$. The equalizer is the subset of $\prod_j Fj$ for which these elements are equal, which is precisely the set of legs with the necessary cone compatibility conditions. This, together with our explicit descriptions of products and equalizers in Set given above, completes the proof that $\lim F$ is the equalizer of c and d. □

Remark 3.2.11. The maps in the equalizer diagram of Theorem 3.2.9 can also be defined categorically, using the universal property of the product $\prod_{f\in\mathrm{mor}\,\mathsf{J}} F(\mathrm{cod}\,f)$. To define c and d it is necessary and sufficient to define each component function, by which we mean the composite with the projection π_f. The components of c are themselves projections, as displayed in the top triangle:

$$\begin{array}{ccccc}
 & & & & F(\operatorname{cod} f) \\
 & & \nearrow^{\pi_{\operatorname{cod} f}} & & \uparrow{\scriptstyle \pi_f} \\
\lim F \rightarrowtail & \prod_{j \in \operatorname{ob} \mathsf{J}} Fj & \overset{c}{\underset{d}{\rightrightarrows}} & \prod_{f \in \operatorname{mor} \mathsf{J}} & F(\operatorname{cod} f) \\
 & \downarrow{\scriptstyle \pi_{\operatorname{dom} f}} & & & \downarrow{\scriptstyle \pi_f} \\
 & F(\operatorname{dom} f) & \xrightarrow[Ff]{} & & F(\operatorname{cod} f)
\end{array}$$

That is, the component of the map c at the indexing element $f \in \operatorname{mor} \mathsf{J}$ is the projection from the product $\prod_j Fj$ onto the component indexed by the object $\operatorname{cod} f \in \mathsf{J}$.

The component at f of the map d, displayed in the bottom square, is defined by projecting from the product $\prod_j Fj$ onto the component indexed by the object $\operatorname{dom} f \in \mathsf{J}$ and then composing with the map $Ff \colon F(\operatorname{dom} f) \to F(\operatorname{cod} f)$. Considering the action of these functions on elements, one can verify that the categorical descriptions agree with the explicit ones described above.

Example 3.2.12. An **idempotent** is an endomorphism $e \colon A \to A$ of some object so that $e^2 = e$. The limit of an idempotent in Set is the set of cones with summit 1, i.e., is the set of $a \in A$ so that $ea = a$. Alternatively, applying Theorem 3.2.9, the limit L is constructed as the equalizer

$$L \overset{s}{\rightarrowtail} A \overset{1}{\underset{e}{\rightrightarrows}} A$$

The universal property of the equalizer (L, s) implies that e factors through s along a unique map r.

$$\begin{array}{ccccc}
A & & & & \\
{\scriptstyle r}\, \downarrow & \searrow^{e} & & & \\
L & \overset{s}{\rightarrowtail} & A & \overset{1}{\underset{e}{\rightrightarrows}} & A
\end{array}$$

The factorization $e = sr$ is said to **split** the idempotent. Now $srs = es = s$ implies that rs and 1_L both define factorization of the diagram

$$\begin{array}{ccccc}
L & & & & \\
{\scriptstyle rs}\, \downarrow\downarrow\, {\scriptstyle 1_L} & \searrow^{s} & & & \\
L & \overset{s}{\rightarrowtail} & A & \overset{1}{\underset{e}{\rightrightarrows}} & A
\end{array}$$

Uniqueness implies $rs = 1_L$ so L is a **retract** of A.

Limits of general shapes can also be constructed out of terminal objects and pullbacks (subject to size restrictions). For example, writing $*$ for a terminal object, the pullback of the diagram

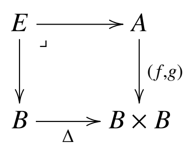
$$\begin{array}{ccc}
A \times B & \xrightarrow{\pi_B} & B \\
{\scriptstyle \pi_A}\downarrow & \lrcorner & \downarrow{\scriptstyle !} \\
A & \xrightarrow[!]{} & *
\end{array}$$

is easily seen to define the product $A \times B$. Iterating this construction, arbitrary finite products can be built out of pullbacks of diagrams constructed from smaller products and terminal objects (the empty product). Hence any category with pullbacks and a terminal object also has binary (and by iteration finite) products.

Lemma 3.2.13. *Equalizers may be defined as a pullback of the diagonal map from an object to its binary product.*

A **diagonal map** is one from an object to an iterated product of that object, all of whose components are identities.

Proof. Elements in the pullback of



are pairs $(a \in A, b \in B)$ so that $f(a) = b = g(a)$. Thus, E is isomorphic to the subset of A consisting of elements a so that $f(a) = g(a)$. This is precisely the equalizer of f and g. □

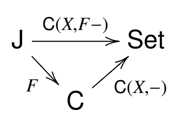
A diagram is said to be **finite** if its indexing category contains only finitely many morphisms.

Corollary 3.2.14. *Any finite limit in* Set *may be constructed from pullbacks and terminal objects.*

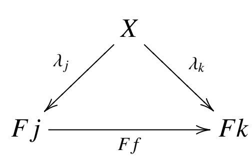
3.3. The representable nature of limits and colimits

Of course, we are not only interested in constructing limits in the category of sets; we care about limits and colimits in all categories. But the considerations just discussed will allow us to do precisely that because all limits and colimits are defined representably in terms of limits in the category of sets.

To see why, fix a diagram $F\colon \mathsf{J} \to \mathsf{C}$ in a locally small category C and an object $X \in \mathsf{C}$, and consider the functor



Theorem 3.2.2 tells us that the limit of $\mathsf{C}(X, F-)$ exists and Theorem 3.2.9 tells us how it may be constructed. An element in the set $\lim_{\mathsf{J}} \mathsf{C}(X, F-)$ is an element of the product $\prod_{j\in\mathsf{J}} \mathsf{C}(X, Fj)$, i.e., is a tuple of morphisms $(\lambda_j\colon X \to Fj)_{j\in\mathsf{J}}$, subject to some compatibility conditions. There is one condition imposed by each non-identity morphism $f\colon j \to k$ in J, namely that the diagram



commutes. In this way, we see that an element of $\lim_{\mathsf{J}} \mathsf{C}(X, F-)$ is a cone over F with summit X; hence

$$\lim_{\mathsf{J}} \mathsf{C}(X, F-) \cong \mathrm{Cone}(X, F).$$

This isomorphism is natural in X. Since the limit of F was defined to be an object that represents the functor $\mathrm{Cone}(-, F)$, we conclude that:

THEOREM 3.3.1. *For any diagram* $F\colon \mathsf{J} \to \mathsf{C}$ *in a locally small category whose limit exists there is a natural isomorphism*

$$\mathsf{C}(X, \lim_{\mathsf{J}} F) \cong \lim_{\mathsf{J}} \mathsf{C}(X, F-). \tag{3.3.2}$$

There are a number of interpretations of this result, a thorough discussion of which demands that we introduce some new terminology.

DEFINITION 3.3.3. A functor $F\colon \mathsf{C} \to \mathsf{D}$

- **preserves limits of shape** J if for any diagram $K\colon \mathsf{J} \to \mathsf{C}$ and limit cone over K, the image of this cone is a limit cone for the composite diagram $FK\colon \mathsf{J} \to \mathsf{D}$.
- **reflects limits of shape** J if any cone over a diagram $K\colon \mathsf{J} \to \mathsf{C}$, whose image under F is a limit cone for the diagram $FK\colon \mathsf{J} \to \mathsf{D}$, is a limit cone for K.
- **creates limits of shape** J if whenever $K\colon \mathsf{J} \to \mathsf{C}$ is a diagram so that $FK\colon \mathsf{J} \to \mathsf{D}$ has a limit, then K has a limit, and moreover F both preserves and reflects J-shaped limits.

The first interpretation of Theorem 3.3.1 is that covariant representable functors $\mathsf{C}(X, -)$ preserve any limits that exist in C, sending them to limits in Set: the image under $\mathsf{C}(X, -)$ of a limit of a diagram F of shape J in C is a limit in Set of the composite diagram

$$\mathsf{J} \xrightarrow{F} \mathsf{C} \xrightarrow{\mathsf{C}(X,-)} \mathsf{Set}\,.$$

Moreover, the functor $\mathsf{C}(X, -)$ preserves the legs of the limit cone because the natural isomorphism of Theorem 3.3.1, by construction, commutes with the natural maps to the product:

$$\begin{array}{ccccc} \lim_{\mathsf{J}} \mathsf{C}(X, F-) & \cong & \mathrm{Cone}(X, F) & \cong & \mathsf{C}(X, \lim_{\mathsf{J}} F) \\ & \searrow & \downarrow & \swarrow & \\ & & \prod_{j \in \mathsf{J}} \mathsf{C}(X, Fj) & & \end{array}$$

The components of the left-hand map are the legs of the limit cone for the diagram $\mathsf{C}(X, F-)$. The components of the right-hand diagram are the images of the legs of the limit cone for F under the functor $\mathsf{C}(X, -)$.

A second way to interpret Theorem 3.3.1 is that the contravariant functor $\mathsf{C}(-, \lim F)$ represented by the limit of $F\colon \mathsf{J} \to \mathsf{C}$ is the limit of the composite diagram $\mathsf{J} \xrightarrow{F} \mathsf{C} \xrightarrow{y} \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ whose objects are the representable functors $\mathsf{C}(-, Fj)$.[2] Put more concisely: the Yoneda embedding $y\colon \mathsf{C} \hookrightarrow \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ preserves all limits that exist in C.

A final observation is that, by the Yoneda lemma, the natural isomorphism (3.3.2) describes precisely the defining universal property of the limit of F. If L is an object of C with a natural isomorphism

$$\mathsf{C}(-, L) \cong \lim_{\mathsf{J}} \mathsf{C}(-, F-) \cong \mathrm{Cone}(-, F),$$

then the object L is a limit for F and the universal element $\lambda\colon L \Rightarrow F$ in $\lim_{\mathsf{J}} \mathsf{C}(L, F-) \cong \mathrm{Cone}(L, F)$ is a limit cone. This says exactly that the Yoneda embedding $y\colon \mathsf{C} \hookrightarrow \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ reflects limits: if $y\lambda\colon yL \Rightarrow yF$ is a limit cone in $\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$, then $\lambda\colon L \Rightarrow F$ is a limit cone in C.

In summary:

THEOREM 3.3.4. *Let* C *be any locally small category.*

(i) Covariant representable functors $\mathsf{C}(X, -)$ *preserve all limits that exist in* C.

[2] This makes use of Exercise 3.5.1.

(ii) The covariant Yoneda embedding $y\colon \mathsf{C} \hookrightarrow \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ *both preserves and reflects limits.*

Dualizing the preceding discussion, we obtain a representable characterization of colimits. For fixed $F\colon \mathsf{J} \to \mathsf{C}$ and $X \in \mathsf{C}$, consider the functor

$$\begin{array}{ccc} \mathsf{J}^{\mathrm{op}} & \xrightarrow{\mathsf{C}(F-,X)} & \mathsf{Set} \\ & {\scriptstyle F}\searrow \quad \nearrow{\scriptstyle \mathsf{C}(-,X)} & \\ & \mathsf{C}^{\mathrm{op}} & \end{array}$$

Again Theorem 3.2.2 tells us that a limit exists and Theorem 3.2.9 tells us how the limit may be constructed. An element in the set $\lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X)$ is an element of the product $\prod_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(Fj, X)$, i.e., is a tuple of morphisms $(\lambda_j\colon Fj \to X)_{j\in\mathsf{J}}$, subject to some conditions. There is one condition imposed by each non-identity morphism $f\colon j \to k$ in J, namely that the diagram

$$\begin{array}{ccc} Fj & \xrightarrow{Ff} & Fk \\ & {\scriptstyle \lambda_j}\searrow \quad \swarrow{\scriptstyle \lambda_k} & \\ & X & \end{array}$$

commutes. In this way, we see that an element of $\lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X)$ is a cone under F with nadir X. The isomorphism $\lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X) \cong \mathrm{Cone}(F, X)$ is again natural in X proving

Theorem 3.3.5. *For any diagram* $F\colon \mathsf{J} \to \mathsf{C}$ *in a locally small category whose colimit exists there is a natural isomorphism*

$$\mathsf{C}(\mathrm{colim}_{\mathsf{J}}\, F, X) \cong \lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X).$$

The dual of Theorem 3.3.4 is:

Theorem 3.3.6. *Let* C *be any locally small category.*

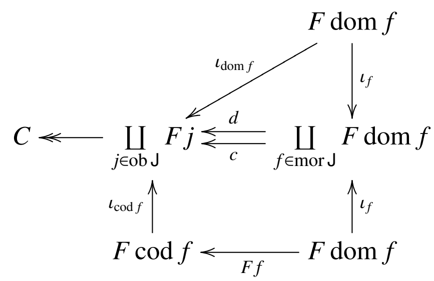
(i) Contravariant representable functors $\mathsf{C}(-, X)$ *carry colimits in* C *to limits in* Set.
(ii) The contravariant Yoneda embedding $y\colon \mathsf{C}^{\mathrm{op}} \hookrightarrow \mathsf{Set}^{\mathsf{C}}$ *both preserves and reflects limits in* C^{op}*: i.e., a cone under a diagram in* C *is a colimit cone if and only if its image defines a limit cone.*

Theorems 3.3.4.(i) and 3.3.6.(i) tell us that limits of diagrams of hom bifunctors can be "moved inside." If the action of the diagram is in the domain variable, these limits become colimits, indexed by the opposite categories. If the action of the diagram is in the codomain variable, the limits remain limits. We will prove the following theorem by applying this principle.

Theorem 3.3.7. *A locally small category* C *with coproducts and coequalizers has colimits of any shape. Dually a category with products and equalizers has all small limits.*

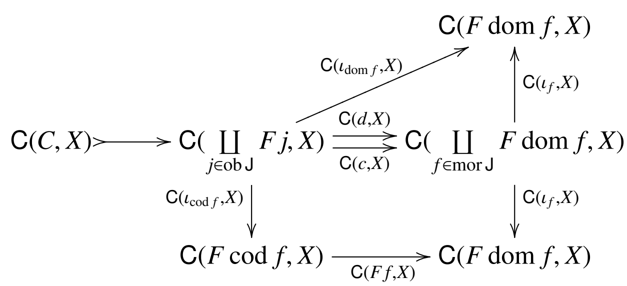
Proof. Consider a diagram $F\colon \mathsf{J} \to \mathsf{C}$. Dualizing the construction of Remark 3.2.11, define maps d, c

(3.3.8)



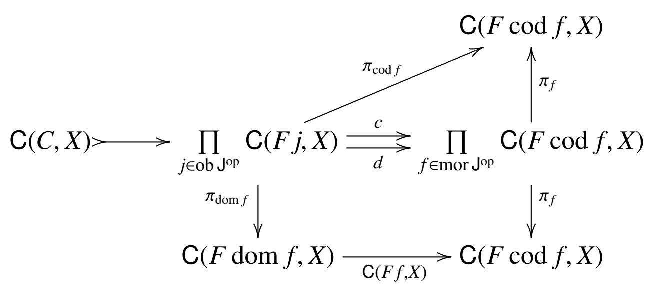
between the coproducts indexed by the morphisms and objects of J respectively. By hypothesis, their coequalizer C exists in C. Our aim is to show that C defines a colimit of F.

By Theorem 3.3.6.(ii), the Yoneda embedding $y\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}^{\mathsf{C}}$ carries the colimit diagram (3.3.8) to a limit diagram



Applying Theorem 3.3.6.(i) to each vertex in this equalizer diagram we obtain an isomorphic diagram

(3.3.9)



Note that $\mathsf{C}(-, X)$ carries the coproduct inclusion maps to product projection maps; when f is regarded as a morphism in J^{op} its domain and codomain are exchanged, so we now write "c" for the map $\mathsf{C}(d, X)$.

When X is fixed, Exercise 3.5.1 tells us that (3.3.9) defines an equalizer diagram in Set. Applying Theorem 3.2.9 and 3.2.11 to the functor $\mathsf{C}(F-, X)\colon \mathsf{J}^{\mathrm{op}} \to \mathsf{Set}$, we recognize (3.3.9) as the diagram whose equalizer defines the limit $\lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X)$. Thus we conclude that

$$\lim_{\mathsf{J}^{\mathrm{op}}} \mathsf{C}(F-, X) \cong \mathsf{C}(C, X),$$

which tells us that the coequalizer C is the colimit of $F\colon \mathsf{J} \to \mathsf{C}$. □

A direct proof of Theorem 3.3.7 simply constructs the diagram (3.3.8) in C and checks that the coequalizer of c and d has the universal property that defines $\operatorname{colim}_{\mathsf{J}} F$. Either of these strategies can also be used to prove the following:

COROLLARY 3.3.10. *Any category with pullbacks and a terminal object has all finite limits. Any category with pushouts and an initial object has all finite colimits.*

Exercises.

EXERCISE 3.3.1. Let C be a category admitting coproducts and coequalizers and consider a functor $F\colon \mathsf{J} \to \mathsf{C}$. Define a parallel pair between two coproducts whose coequalizer will equal the colimit of F.

EXERCISE 3.3.2. Explain in your own words why the Yoneda embedding $\mathsf{C} \hookrightarrow \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ preserves and reflects but does not create limits.

EXERCISE 3.3.3. Show that any equivalence of categories $F\colon \mathsf{C} \to \mathsf{D}$ creates all limits and colimits that D admits.

EXERCISE 3.3.4. For any diagram $D\colon \mathsf{J} \to \mathsf{C}$ and any functor $F\colon \mathsf{C} \to \mathsf{D}$ define a canonical map $\operatorname{colim} FD \to F \operatorname{colim} D$, assuming both colimits exist. The functor F is said to preserve the colimit of D just when this map is an isomorphism.

3.4. Examples

As we have seen, limits and colimits can be defined in any category, although they need not always exist. In this section, we consider a few examples.

EXAMPLE 3.4.1. In any poset (or preorder), a limit of a diagram is an infimum of its objects, while a colimit of a diagram is a supremum of its objects. Whether or not there are any morphisms in the diagram makes no difference because all diagrams in a preorder commute.

For instance, in the poset of natural numbers with the order relation $k \leq n$ if and only if k divides n, the limit of a set of objects is their greatest common divisor, while the colimit is their least common multiple.

Next, we consider limits of diagrams in Cat. In Examples 2.1.2.(viii) and (ix), we saw that the functors $\mathrm{ob}, \mathrm{mor}\colon \mathsf{Cat} \to \mathsf{Set}$ are both representable. By Theorem 3.3.4.(i), both functors preserve limits, which means that if the limit of a diagram of categories exists, then its set of objects must be the limit of the underlying diagrams of object-sets and its set of morphisms must be the limit of the underlying diagrams of morphism-sets. Furthermore, domains, codomains, and identities are also preserved, as each of these maps is expressible, by Exercise 2.1.1, as a natural transformation between representable functors. One might reasonably guess that composition is preserved as well. This defines candidates for the product of any categories and for the equalizer of any parallel pair of functors, allowing us to prove that:

PROPOSITION 3.4.2. *The categories* Cat *and* CAT *are complete.*

PROOF. The product, introduced in Definition 1.3.10, and higher arity versions defines products of categories. The equalizer E of a pair of parallel functors

$$\mathsf{E} \rightarrowtail \mathsf{C} \underset{G}{\overset{F}{\rightrightarrows}} \mathsf{D}$$

is the subcategory of C consisting of those objects $c \in \mathsf{C}$ so that $Fc = Gc$ and those morphisms f so that $Ff = Gf$. Now apply Theorem 3.3.7. □

EXAMPLE 3.4.3. For any functor $F\colon \mathsf{C} \to \mathsf{Set}$, the pullback of

$$\begin{array}{ccc} ? & \longrightarrow & \mathsf{Set}_* \\ \downarrow & \lrcorner & \downarrow{\scriptstyle U} \\ \mathsf{C} & \xrightarrow[F]{} & \mathsf{Set} \end{array}$$

is a category whose objects are pairs $c \in \mathsf{C}$ and $(X, x) \in \mathsf{Set}_*$ so that $Fc = X$. Morphisms are pairs $f\colon c \to c'$ and $g\colon (X, x) \to (X', x')$ so that $Ff = Ug$. Expressing this data more efficiently, we see that the pullback is the category $\int F$ of elements of F.[3]

Categories also have all colimits, which can be proven by showing they have coproducts (disjoint unions) and coequalizers. But the latter can be rather more complicated than was the case for equalizers, for reasons that we will explain in Chapter 4.

EXAMPLE 3.4.4. We can calculate the pushout

$$\begin{array}{ccc} \mathbb{1} \coprod \mathbb{1} & \rightarrowtail & \mathbb{2} \\ \downarrow & & \downarrow \\ \mathbb{1} & \longrightarrow & \mathsf{P} \end{array}$$

by describing its universal property. A functor $\mathsf{P} \to \mathsf{C}$ corresponds to a cone under the diagram $\mathbb{1} \leftarrow \mathbb{1} \coprod \mathbb{1} \to \mathbb{2}$ with nadir C. This data defines an endomorphism f in C, a morphism $\mathbb{2} \to \mathsf{C}$ so that its domain and codomain are the same object $\mathbb{1} \to \mathsf{C}$. In defining P we must take care not to impose any relations on the composite endomorphisms $f \circ f \cdots \circ f$. Thus, P must be the free category with one object and one non-identity morphism. Recalling that one-object categories can be identified with monoids, we find that P is quite familiar: it is the monoid $\mathbb{N}$ of natural numbers, with addition.

Similarly, the group $\mathbb{Z}$ of integers with addition is the pushout

$$\begin{array}{ccc} \mathbb{1} \coprod \mathbb{1} & \rightarrowtail & \mathbb{I} \\ \downarrow & & \downarrow \\ \mathbb{1} & \longrightarrow & \mathbb{Z} \end{array}$$

in which $\mathbb{I}$ is the **walking isomorphism**, the category with two objects and two non-identity morphisms pointing in opposite directions, which are inverse isomorphisms.

EXAMPLE 3.4.5. In any category with finite limits, we can define the **kernel pair** of a morphism $f\colon X \to Y$, which is the pullback of f along itself

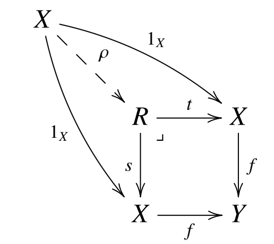
$$\begin{array}{ccc} R & \xrightarrow{t} & X \\ {\scriptstyle s}\downarrow & \lrcorner & \downarrow{\scriptstyle f} \\ X & \xrightarrow[f]{} & Y \end{array}$$

These maps define a monomorphism $(s, t)\colon R \rightarrowtail X \times X$, so the object R is always a subobject of the product $X \times X$.[4] In Set, a subset $R \subset X \times X$ defines a **relation** on X. Indeed, objects R defined in this manner are always **equivalence relations**, in the following categorical sense:

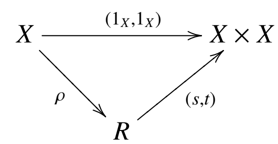
[3] This pullback construction works for contravariant functors too, in which case we define $\int F$ to be the opposite of the pullback of $U\colon \mathsf{Set}_* \to \mathsf{Set}$ along F.

[4] A subobject of an object c is a monomorphism with codomain c.

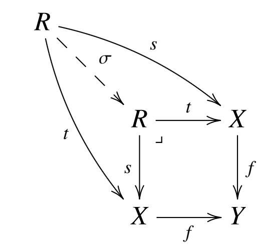
- There is a reflexivity map



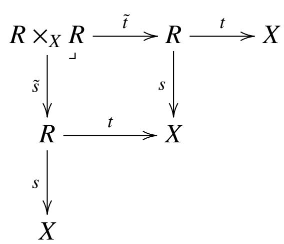
that is a section of both s and t, i.e., that defines a factorization of the diagonal



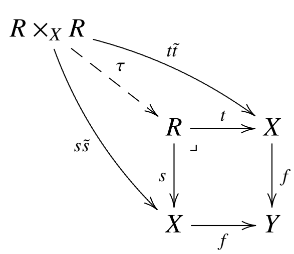
- There is a symmetry map



so that $t\sigma = s$ and $s\sigma = t$.
- There is a transitivity map whose domain is the pullback of t along s

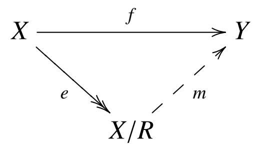


The diagram defines a cone over the pullback defining R and thus induces a map



so that $s\tau = s\tilde{s}$ and $t\tau = t\tilde{t}$.

A **equivalence relation** in a category C with finite limits is a subobject $(s,t)\colon R \rightarrowtail X \times X$ equipped with maps ρ, σ, and τ satisfying the relations with s and t described above. When it exists, the coequalizer of the maps $s,t\colon R \rightrightarrows X$ of an equivalence relation defines a quotient object $e\colon X \twoheadrightarrow X/R$. In Set, X/R is the set of R-equivalence classes of elements of X. For equivalence relations arising as kernel pairs, there is a unique factorization



In good situations, such as when C is a **regular category** or a **Grothendieck topos**, the map m is a monomorphism and this factorization is called the **image factorization** of the map f.

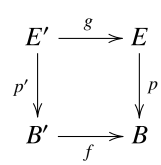
Exercises.

EXERCISE 3.4.1. Let G be a group regarded as 1-object category. Describe the colimit and the limit of a diagram $G \to \mathsf{Set}$ in group theoretic terms.

EXERCISE 3.4.2. Show that if $K\colon \mathsf{C} \to \mathsf{D}$ is an equivalence then for any D-indexed functor F, if the limit of FK exists then it also defines a limit of and limit cone for F.

EXERCISE 3.4.3. Following [**Gro58**], define a **fiber space** $p\colon E \to B$ to be a map in the category of spaces subject to no further restrictions. A map of fiber spaces is a commutative square. Thus the category of fiber spaces is the diagram category Top^2. We are also interested in the non-full subcategory $\mathsf{Top}/B \subset \mathsf{Top}^2$ of fiber spaces over B and maps whose codomain component is the identity. Prove the following:

(i) A map



of fiber spaces induces a canonical map between fibers over a point $b \in B'$ and its image $f(b) \in B$.

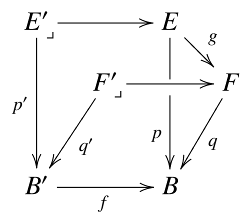
(ii) The fiber of a product of fiber spaces is the product of the fibers.

(iii) A projection $E = B \times F \to B$ defines a **trivial fiber space** over B. This definition makes sense for any space F. Note then that each fiber is canonically isomorphic to F. Characterize the isomorphisms in Top/B between two trivial fiber spaces (with a priori distinct fibers) over B.

(iv) Prove that the assignment of the set of continuous sections of a fiber space over B defines a functor Sect: $\mathsf{Top}/B \to \mathsf{Set}$.

(v) Consider the non-full subcategory $\mathsf{Top}^2_{\text{pb}}$ of fiber spaces and maps which are pullback squares. Prove that the assignment of the set of continuous sections to a fiber space defines a functor Sect: $(\mathsf{Top}^2_{\text{pb}})^{\text{op}} \to \mathsf{Set}$.

(vi) Describe the compatibility between the actions of the "sections" functors just introduced with respect to the map g of fiber spaces p and q over B and their restrictions along $f\colon B' \to B$.



3.5. Limits and colimits and diagram categories

By the universal property of products, a diagram of shape J in a product category $\prod_\alpha \mathsf{C}_\alpha$ is a family of diagrams $\mathsf{J} \to \mathsf{C}_\alpha$. It is easy to see if each component diagram has a limit or a colimit, then the same is true of the diagram $\mathsf{J} \to \prod_\alpha \mathsf{C}_\alpha$.

Proposition 3.5.1. *If* A *is small then the forgetful functor* $\mathsf{C}^{\mathsf{A}} \to \mathsf{C}^{\mathrm{ob}\,\mathsf{A}}$ *creates all limits and colimits that exist in* C*. These limits are defined pointwise, meaning that for each* $a \in \mathsf{A}$*, the evaluation functor* $\mathrm{ev}_a\colon \mathsf{C}^{\mathsf{A}} \to \mathsf{C}$ *preserves all limits and colimits existing in* C*.*

Proof. The category $\mathsf{C}^{\mathrm{ob}\,\mathsf{A}}$ of functors is isomorphic to the $\mathrm{ob}\,\mathsf{A}$-indexed product of the category C with itself. By the argument just given, $\mathsf{C}^{\mathrm{ob}\,\mathsf{A}}$ has all limits or colimits that C does, and these are preserved by the evaluation functors $\mathrm{ev}_a\colon \mathsf{C}^{\mathrm{ob}\,\mathsf{A}} \to \mathsf{C}$. The remaining details are left as Exercise 3.5.1. □

The functor category $\mathsf{C}^{\mathrm{ob}\,\mathsf{A}}$, the $\mathrm{ob}\,\mathsf{A}$-indexed product of the category C with itself, is an example of a power in the category CAT. Iterated products of an object $A \in \mathsf{C}$ are called **powers** or **cotensors**. For a set I, the I-indexed power of A is denoted $\prod_I A$ or A^I. The representable universal property is

$$\mathsf{C}(X, A^I) \cong \mathsf{C}(X, A)^I,$$

i.e., a map $h\colon X \to A^I$ determines an I-indexed family of maps $h_i\colon X \to A$ defined by composing with each product projection $\epsilon_i\colon A^I \to A$. In the case of $\mathsf{C} = \mathsf{Set}$, this notation is consistent with the exponential notation introduced previously: the power A^I is isomorphic to the set of functions from I to A, and the map $\epsilon_i\colon A^I \to A$ evaluates each function at $i \in I$.

When C has all limits of a certain shape they assemble into a functor.

Proposition 3.5.2. *If* C *has all* J*-shaped limits, then a choice of limits for each diagram defines the action on objects of a functor* $\lim\colon \mathsf{C}^{\mathsf{J}} \to \mathsf{C}$*.*

As a warning, this functor is not canonically defined but rather requires an arbitrary choice of a limit for each diagram.

Proof. Choose a limit and a limit cone for each diagram $F \in \mathsf{C}^{\mathsf{J}}$. It remains to define the action of $\lim\colon \mathsf{C}^{\mathsf{J}} \to \mathsf{C}$ on morphisms. The vertical composite of a natural transformation $\alpha\colon F \Rightarrow G$ and the limit cone $\lambda\colon \lim F \Rightarrow F$ for F defines a cone $\alpha \cdot \lambda\colon \lim F \Rightarrow F \Rightarrow G$ over G. By the universal property of $\lim G$, this cone factors uniquely through a map $\lim F \to \lim G$ in C, which we define to be $\lim \alpha$. Uniqueness of the universal properties implies that this construction is functorial. □

More generally, the construction of Proposition 3.5.2 implies that a natural transformation between diagrams gives rise to a morphism between their limits or between their colimits, whenever these exist. (Whether the codomain category has all limits or colimits of that shape is irrelevant.) Moreover, whenever the natural transformation is a natural isomorphism, the induced map between the limits or colimits is an isomorphism.

PROPOSITION 3.5.3. *A natural isomorphism between diagrams induces a naturally-defined isomorphism between their limits or colimits, whenever these exist.*

The main computational motivation for Eilenberg and Mac Lane's formulation of the concept of naturality was that it is a necessary ingredient in this result. While the general theory of limits and colimits had not yet been developed, [**EM45**] proves Proposition 3.5.3 in the special case of directed diagrams.

PROOF. The proof of Proposition 3.5.2 defines a limit functor from the full subcategory of C^{J} spanned by those functors admitting limits to C. This functor, like all functors, preserves isomorphisms (Lemma 1.3.7). □

Indeed, naturality is a necessary condition.

EXAMPLE 3.5.4. Consider the group $\mathbb{Z}/2$ and the $\mathbb{Z}/2$-sets 2, the two-element set with a trivial $\mathbb{Z}/2$-action, and $2'$, the two-element set where the generator of $\mathbb{Z}/2$ acts by exchanging the two elements. We may regard 2 and 2' as functors $\mathbb{Z}/2 \to \mathsf{Set}$ from which perspective they are pointwise isomorphic (because $|2| = |2'|$) but not naturally isomorphic (because there is no isomorphism that commutes with the $\mathbb{Z}/2$-actions).

For any G-set $X: G \to \mathsf{Set}$, the limit is the set of G-fixed points (see Example 3.2.3) and the colimit is the set of G-orbits. In the case of our pair of $\mathbb{Z}/2$-sets, $\lim 2$ has two elements while $\lim 2'$ is empty; similarly, $\operatorname{colim} 2$ has two elements while $\operatorname{colim} 2'$ is a singleton.

EXAMPLE 3.5.5. To a finite set X, we can associate the set $\mathrm{Sym}(X)$ of its permutations or the set $\mathrm{Ord}(X)$ of its total orderings. Note that these sets are isomorphic. The assignments are functorial, not with respect to all maps of finite sets, but with respect to the bijections: the functor $\mathrm{Sym}: \mathsf{Fin}_{\mathrm{iso}} \to \mathsf{Fin}$ acts by conjugation and the functor $\mathrm{Ord}: \mathsf{Fin}_{\mathrm{iso}} \to \mathsf{Fin}$ acts by translation. The pointwise isomorphism $\mathrm{Sym}(X) \cong \mathrm{Ord}(X)$ is not natural, and indeed limits or colimits of restrictions of these diagrams need not be isomorphic.

A combinatorialist would call a functor $F: \mathsf{Fin}_{\mathrm{iso}} \to \mathsf{Fin}$ a **species**.[5] The image $F(\underline{n})$ of the n-element set $\underline{n}$ is the set of labeled F-structures on $\underline{n}$. The set of unlabeled F-structures on $\underline{n}$ is defined by restricting $\mathsf{Fin}_{\mathrm{iso}}$ to the full subcategory spanned by the n-element set, i.e., to the group Σ_n, and forming the colimit of the diagram $\Sigma_n \hookrightarrow \mathsf{Fin}_{\mathrm{iso}} \xrightarrow{F} \mathsf{Fin}$. Because Ord and Sym are pointwise isomorphic, their sets of labelled F-structures are isomorphic. However, the set of unlabeled Sym-structures on $\underline{n}$ is the set of conjugacy classes of permutations of n-elements, while the set of unlabeled Ord-structures on $\underline{n}$ is trivial: all linear orders on $\underline{n}$ are isomorphic. See [**Joy81**] for more.

Exercises.

EXERCISE 3.5.1. Prove that for any small category A, the functor category C^{A} again has any limits or colimits that C does, constructed pointwise. That is, given a diagram $F: \mathsf{J} \to \mathsf{C}^{\mathsf{A}}$, with J small, show that whenever the limits of the diagrams

$$\mathsf{J} \xrightarrow{F} \mathsf{C}^{\mathsf{A}} \xrightarrow{\mathrm{ev}_a} \mathsf{C}$$

exist in C for all $a \in \mathsf{A}$, then these values define the action on objects of $\lim F \in \mathsf{C}^{\mathsf{A}}$, a limit of the diagram F.

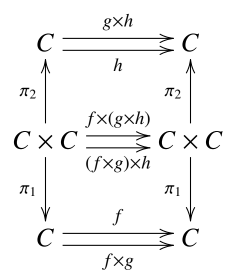
[5] In this context, "species" is singular.

3.6. Warnings

Limits, when they exist, are unique up to a unique isomorphism commuting with the maps in the limit cone. But this is not the same thing as saying that limits are unique on the nose. With the exception of skeletal categories, choices of limits of diagrams of fixed shape, as required to the define the limit functor of Proposition 3.5.2, can seldom be made compatibly. And even in skeletal categories, in which isomorphic objects are indeed equal, various natural isomorphisms involving limit constructions might not be identities.

Lemma 3.6.1. *For any triple of objects X, Y, Z in a category with binary products, there is a unique natural isomorphism $X \times (Y \times Z) \cong (X \times Y) \times Z$ commuting with the projections to X, Y, and Z.*

Lemma 3.6.1 asserts that the product is naturally associative. It follows that any iteration of binary products can be used to define n-ary products. However, even in a skeletal category, in which the objects $X \times (Y \times Z)$ and $(X \times Y) \times Z$ are necessarily equal, the natural isomorphism might not be an identity. The following example, from [**ML98a**, p. 164] is due to John Isbell. Consider sk(Set), a skeletal category of sets. Since sk(Set) is equivalent to a complete category it has all limits and in particular has products. Let C denote the countably infinite set. Its product $C \times C = C$ and the product projections $\pi_1, \pi_2 : C \to C$ are both epimorphisms. Suppose the component of the natural isomorphism $C \times (C \times C) \cong (C \times C) \times C$ were the identity. Naturality would then apply that for any triple of maps $f, g, h : C \to C$ that $f \times (g \times h) = (f \times g) \times h$. Maps



between products are equal if and only if their projections onto components are equal. Because π_1 and π_2 are epimorphisms, this implies that $f = f \times g$ and $g \times h = h$. So we conclude that any pair of maps $f, g : C \rightrightarrows C$ must be equal, which is absurd. It follows that the component of the natural associativity isomorphism $C \times (C \times C) \cong (C \times C) \times C$ is not an identity.

3.7. Size matters

In developing the theory of limits we have taken care to consider only those diagrams whose domains are small categories. In this section we explain the reason for our caution. The following definition gives a precise meaning to the cardinality of a small category.

Definition 3.7.1. The **cardinality** of a small category is the cardinality of the set of its morphisms. A category whose cardinality is less than κ is called κ-**small**.

A κ-small diagram is one whose indexing category is κ-small.

Proposition 3.7.2 (Freyd). *Any κ-small category that admits all κ-small limits is a preorder.*

In fact, this result is true under less restrictive completeness hypotheses, which the reader will have no trouble formulating.

Proof. Let λ be the cardinality of the set of morphisms in a κ-small category C, and suppose there exists a parallel pair $f, g\colon B \to A$ of morphisms with $f \neq g$. By the universal property of the power, there are 2^λ distinct morphisms $B \to A^\lambda$, whose components are either f or g. We have $|\mathsf{C}(B, A^\lambda)| \geq 2^\lambda > \lambda = |\operatorname{mor}\mathsf{C}|$, contradicting the fact that $\mathsf{C}(B, A^\lambda)$ is a subset of $\operatorname{mor}\mathsf{C}$. □

Note that it is possible for a preorder to be complete or cocomplete. A poset P is complete and cocomplete as a category if and only if it is a **complete lattice**, that is, if and only if every subset $A \subset P$ has both an infimum (greatest lower bound) and a supremum (least upper bound).

There is one notable special case of large limits and colimits that frequently exist.

Example 3.7.3. A limit of the identity functor $1_\mathsf{C}\colon \mathsf{C} \to \mathsf{C}$ is an initial object for C; recall that initial objects are small *colimits*. Dually, a colimit of the identity functor defines a terminal object, which is a small limit. The details are left as Exercise 3.7.1.

Exercises.

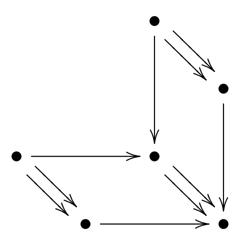
Exercise 3.7.1. Prove that a colimit of the identity functor $1_\mathsf{C}\colon \mathsf{C} \to \mathsf{C}$ defines a terminal object for C.

3.8. Interactions between limits and colimits

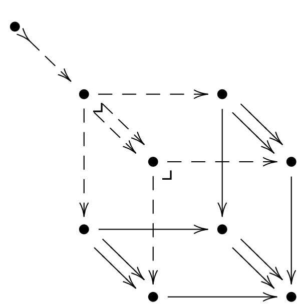
Consider a bifunctor $F\colon \mathsf{I} \times \mathsf{J} \to \mathsf{C}$. By Exercise 1.6.9, F may also be regarded as a functor $F\colon \mathsf{I} \to \mathsf{C}^\mathsf{J}$ or $F\colon \mathsf{J} \to \mathsf{C}^\mathsf{I}$.

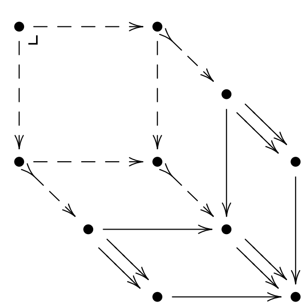
Theorem 3.8.1. *If the limits* $\lim_{i \in \mathsf{I}} \lim_{j \in \mathsf{J}} F(i, j)$ *and* $\lim_{j \in \mathsf{J}} \lim_{i \in \mathsf{I}} F(i, j)$ *associated to a diagram* $F\colon \mathsf{I} \times \mathsf{J} \to \mathsf{C}$ *exist in* C*, they are isomorphic and define the limit* $\lim_{\mathsf{I} \times \mathsf{J}} F$.

If $\mathsf{I} = \bullet \rightrightarrows \bullet$ is the category indexing an equalizer diagram and $\mathsf{J} = \bullet \to \bullet \leftarrow \bullet$ is the category indexing a pullback, then Theorem 3.8.1 says that the limit of a diagram indexed by the category $\mathsf{I} \times \mathsf{J}$



may be formed either by first forming the two pullbacks and then taking the equalizer of the induced map between them, as displayed below-left





or by taking the equalizers of the three parallel pairs, and then forming the pullback of the induced maps between them, as displayed above-right.

PROOF. By the Yoneda lemma, it suffices to prove that $\mathsf{C}(X, \lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} F(i,j)) \cong \mathsf{C}(X, \lim_{(i,j)\in\mathsf{I}\times\mathsf{J}} F(i,j)) \cong \mathsf{C}(X, \lim_{j\in\mathsf{J}} \lim_{i\in\mathsf{I}} F(i,j))$. By Theorem 3.3.4.(i), we know that

$$\mathsf{C}(X, \lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} F(i,j)) \cong \lim_{i\in\mathsf{I}} \mathsf{C}(X, \lim_{j\in\mathsf{J}} F(i,j)) \cong \lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} \mathsf{C}(X, F(i,j)).$$

Thus, we have reduced to the case of proving for any Set-valued functor $H\colon \mathsf{I}\times\mathsf{J} \to \mathsf{Set}$ that

$$\lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} H(i,j) \cong \lim_{(i,j)\in\mathsf{I}\times\mathsf{J}} H(i,j) \cong \lim_{j\in\mathsf{J}} \lim_{i\in\mathsf{I}} H(i,j).$$

On account of the isomorphism of categories $\mathsf{I}\times\mathsf{J} \cong \mathsf{J}\times\mathsf{I}$, it suffices to prove one of these isomorphisms.

The set $\lim_{(i,j)\in\mathsf{I}\times\mathsf{J}} H(i,j)$ is the set of cones with summit 1 over the $\mathsf{I}\times\mathsf{J}$-indexed diagram H. The set $\lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} H(i,j)$ is the set of cones with summit 1 over the I-indexed diagram $\lim_{j\in\mathsf{J}} H(-,j)$. Such a cone is comprised of legs $\lambda_i\colon 1 \to \lim_{j\in\mathsf{J}} H(i,j)$ that commute with the maps of limits $\lim_{j\in\mathsf{J}} H(i,j) \Rightarrow \lim_{j\in\mathsf{J}} H(i',j)$ determined by each morphism $i \to i' \in \mathsf{I}$. The map of limits is defined, as in Proposition 3.5.2, by a map of J-indexed diagrams with components $H(i,j) \to H(i',j)$ for each $j \in \mathsf{J}$.

By the universal property of the J-indexed limits, each cone leg $\lambda_i\colon 1 \to \lim_{j\in\mathsf{J}} H(i,j)$ is itself determined by legs $\lambda_{i,j}\colon 1 \to H(i,j)$, which must commute with the maps $H(i,j) \to H(i,j')$ induced by each $j \to j'$. The totality of this data is precisely a cone $(\lambda_{i,j}\colon 1 \to H(i,j))_{(i,j)\in\mathsf{I}\times\mathsf{J}}$ over the $\mathsf{I}\times\mathsf{J}$-indexed diagram H. Thus, as their elements coincide, we see that $\lim_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} H(i,j) \cong \lim_{(i,j)\in\mathsf{I}\times\mathsf{J}} H(i,j)$. □

By Theorem 3.8.1 and its dual, limits commute with limits and colimits commute with colimits. By contrast, limits do not necessarily commute with colimits.

LEMMA 3.8.2. *For any bifunctor $F\colon \mathsf{I}\times\mathsf{J} \to \mathsf{C}$, there is a canonical map*

$$\kappa\colon \operatorname*{colim}_{i\in\mathsf{I}} \lim_{j\in\mathsf{J}} F(i,j) \to \lim_{j\in\mathsf{J}} \operatorname*{colim}_{i\in\mathsf{I}} F(i,j),$$

which, however, is not necessarily an isomorphism.

PROOF. By the universal property of the colimit the map κ may be defined by specifying components $\kappa_i\colon \lim_{j\in\mathsf{J}} F(i,j) \to \lim_{j\in\mathsf{J}} \operatorname{colim}_{i'\in\mathsf{I}} F(i',j)$ that define a cone under the I-indexed diagram $\lim_{j\in J} F(-,j)$. By the universal property of the limit, each component κ_i in turn may be defined by specifying components $\kappa_{i,j}\colon \lim_{j'\in\mathsf{J}} F(i,j') \to \operatorname{colim}_{i'\in\mathsf{I}} F(i',j)$ that define a cone over the J-indexed diagram $\operatorname{colim}_{i\in\mathsf{I}} F(i,-)$. Define $\kappa_{i,j}$ to be the composite

$$\kappa_{i,j}\colon \lim_{j'\in\mathsf{J}} F(i,j') \xrightarrow{\pi_{i,j}} F(i,j) \xrightarrow{\iota_{i,j}} \operatorname*{colim}_{i'\in\mathsf{I}} F(i',j)$$

of the leg $\pi_{i,j}$ of the limit cone for $F(i,-)$ and the leg $\iota_{i,j}$ of the colimit cone for $F(-,j)$. Because the maps in the diagrams $\lim_{j\in J} F(-,j)$ and $\operatorname{colim}_{i\in\mathsf{I}} F(i,-)$ are induced by the maps $F(i,j) \to F(i',j')$ obtained by applying F to morphisms $i \to i' \in \mathsf{I}$ and $j \to j' \in \mathsf{J}$, the $\kappa_{i,j}$ assemble into a cone, defining κ_i, and the κ_i assemble into a cone defining κ. □

Specializing to the case where C is the poset category $(\mathbb{R}, \leq)$, we have the following immediate corollary:

COROLLARY 3.8.3. *For any pair of sets X and Y and any function $f\colon X\times Y \to \mathbb{R}$*

$$\sup_{x\in X} \inf_{y\in Y} f(x,y) \leq \inf_{y\in Y} \sup_{x\in X} f(x,y).$$

EXAMPLE 3.8.4. Consider the extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ with the obvious extended ordering. As a category, $\bar{\mathbb{R}}$ is complete and cocomplete: the limit of a function $f\colon X \to \bar{\mathbb{R}}$ is its inf and the colimit is its sup. These are ordinary real numbers in the case where the range of f is bounded.

Taking the domain $X = \mathbb{N}$, a function $x\colon \mathbb{N} \to \mathbb{R}$ is a sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers. We have

$$\liminf_{n\to\infty} x_n = \sup_{n\geq 0} \inf_{m\geq n} x_m = \sup_{n\geq 0} \inf_{m\geq 0} x_{n+m} = \lim_n \operatorname{colim}_m x_{n+m}$$

where we regard the sequence as a bifunctor $\mathbb{N} \times \mathbb{N} \xrightarrow{+} \mathbb{N} \xrightarrow{x} \mathbb{R}$ indexed by the discrete category $\mathbb{N} \times \mathbb{N}$. Dually,

$$\limsup_{n\to\infty} x_n = \inf_{n\geq 0} \sup_{m\geq n} x_m = \inf_{n\geq 0} \sup_{m\geq 0} x_{n+m} = \operatorname{colim}_n \lim_m x_{n+m}$$

By Corollary 3.8.3, $\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$. The limit of this sequence exists if and only if this inequality is an equality.

CHAPTER 4

Adjunctions

> It appears ... that there exists a kind of duality between the tensor product and the ... functor Hom ...
>
> Daniel Kan, "Adjoint functors"

From a set S one can build a vector space defined over any field $\Bbbk$. The most natural way to do this is to let the elements of S serve as a basis for the vector space: vectors are then finite formal sums $k_1 s_1 + \cdots + k_n s_n$ with $k_i \in \Bbbk$ and $s_i \in S$. The dimension of this vector space, often denoted by $\Bbbk[S]$, is equal to the cardinality of S. This construction is functorial — a function $f\colon S \to T$ induces a linear map $\Bbbk[f]\colon \Bbbk[S] \to \Bbbk[T]$ defined on the basis elements in the evident way — defining a functor $\Bbbk[-]\colon \mathsf{Set} \to \mathsf{Vect}_{\Bbbk}$. The vector space $\Bbbk[S]$ is the **free vector space on the set** S.

There are many instances of free constructions in mathematics. One can define the free (abelian) group on a set, the free ring on a set,[1] the free module over some ring R on an abelian group, and so on. Sometimes there are competing notions of free construction: does the free graph on a set have a single edge between every pair of vertices or none? All of these constructions, and many others besides, are clarified by the concept of an **adjunction**, introduced by Daniel Kan [**Kan58**].

The universal property of the free vector space functor $\Bbbk[-]\colon \mathsf{Set} \to \mathsf{Vect}_{\Bbbk}$ is expressed by saying that it is left adjoint to the forgetful functor $U\colon \mathsf{Vect}_{\Bbbk} \to \mathsf{Set}$ that carries a vector space to its underlying set: linear maps $\Bbbk[S] \to V$ correspond to functions $S \to U(V)$, which specify the image of the basis vectors $S \subset \Bbbk[S]$. This correspondence is natural in the set S and in the vector space V, and so, by the Yoneda lemma, this universal property can be used to define the free vector space functor $\Bbbk[-]$. The forgetful functor $U\colon \mathsf{Vect}_{\Bbbk} \to \mathsf{Set}$ has no other adjoints; unlike the situation for graphs, there are no competing notions of free construction.

In this chapter we explore the general theory of adjoint functors, both in the abstract and aided by a plethora of examples.

4.1. Adjoint functors

In this chapter we consider certain opposing pairs of functors $F\colon \mathsf{C} \rightleftarrows \mathsf{D}\colon G$.

Definition 4.1.1. An **adjunction** consists of a pair of functors $F\colon \mathsf{C} \to \mathsf{D}$ and $G\colon \mathsf{D} \to \mathsf{C}$ together with a natural isomorphism

$$\mathsf{D}(Fc, d) \cong \mathsf{C}(c, Gd),$$

[1] But not the free field!

in which case one says that F is **left adjoint** to G, or, equivalently, that G is **right adjoint** to F. The morphisms

$$Fc \xrightarrow{f} d \qquad \leftrightsquigarrow \qquad c \xrightarrow{f^\sharp} Gd$$

corresponding under the bijection are **adjunct**, or **transposes** of each other.[2]

Naturality asserts that the isomorphisms defining an adjunction assemble into a natural isomorphism between the functors

$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \underset{\mathsf{C}(-,G-)}{\overset{\mathsf{D}(F-,-)}{\underset{\Downarrow\cong}{\rightrightarrows}}} \mathsf{Set}.$$

More explicitly, naturality in D says that for any morphism $k\colon d \to d'$, the left-hand diagram displayed below commutes in Set:

$$\begin{array}{ccc} \mathsf{D}(Fc,d) & \xrightarrow{\cong} & \mathsf{C}(c,Gd) \\ {\scriptstyle k_*}\downarrow & & \downarrow{\scriptstyle (Gk)_*} \\ \mathsf{D}(Fc,d') & \xrightarrow[\cong]{} & \mathsf{C}(c,Gd') \end{array} \qquad \forall Fc \xrightarrow{f} d \quad \rightsquigarrow \quad \begin{array}{ccc} c & \xrightarrow{f^\sharp} & Gd \\ & {\scriptstyle (kf)^\sharp}\searrow & \downarrow{\scriptstyle Gk} \\ & & Gd' \end{array}$$

On elements this says that for any $f\colon Fc \to d$, the transpose of $kf\colon Fc \to d'$ is equal to the composite of the transpose of f with Gk.

Dually, naturality in C says that for any morphism $h\colon c \to c'$, the left-hand diagram displayed below commutes in Set.

$$\begin{array}{ccc} \mathsf{D}(Fc',d) & \xrightarrow{\cong} & \mathsf{C}(c',Gd) \\ {\scriptstyle (Fh)^*}\downarrow & & \downarrow{\scriptstyle h^*} \\ \mathsf{D}(Fc,d) & \xrightarrow[\cong]{} & \mathsf{C}(c,Gd) \end{array} \qquad \forall Fc' \xrightarrow{f} d \quad \rightsquigarrow \quad \begin{array}{ccc} c & & \\ {\scriptstyle h}\downarrow & \searrow{\scriptstyle (f\cdot Fh)^\sharp} & \\ c' & \xrightarrow[f^\sharp]{} & Gd \end{array}$$

On elements, this asserts that the transpose of $f \cdot Fh\colon Fc \to d$ is the composite of the transpose of f with h.

LEMMA 4.1.2. *Suppose F and G are adjoint functors, i.e., $\mathsf{D}(Fc,d) \cong \mathsf{C}(c,Gd)$ naturally in $c \in \mathsf{C}$ and $d \in \mathsf{D}$. Then for any morphisms with domains and codomains as displayed below*

$$\begin{array}{ccc} Fc & \xrightarrow{f} & d \\ {\scriptstyle Fh}\downarrow & & \downarrow{\scriptstyle k} \\ Fc' & \xrightarrow[g]{} & d' \end{array} \qquad \leftrightsquigarrow \qquad \begin{array}{ccc} c & \xrightarrow{f^\sharp} & Gd \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle Gk} \\ c' & \xrightarrow[g^\sharp]{} & Gd' \end{array} \tag{4.1.3}$$

the left-hand square commutes in D if and only if the right-hand transposed square commutes in C.

REMARK 4.1.4. Conversely, if $F\colon \mathsf{C} \rightleftarrows \mathsf{D}\colon G$ are opposing functors equipped with isomorphisms $\mathsf{D}(Fc,d) \cong \mathsf{C}(c,Gd)$ for all $c \in \mathsf{C}$ and $d \in \mathsf{D}$, then naturality of these chosen isomorphisms is equivalent to the assertion of lemma 4.1.2, that a square as displayed on the left of (4.1.3) commutes if and only if the square on the right commutes. The proof is left as Exercise 4.1.1.

---

[2] The notation "$(-)^\sharp$" is meant to signal any adjunct, with no preference as to which of the adjunct pair is denoted in this way.

Notation 4.1.5. A turnstyle "⊣" is used to designate that an opposing pair of functors are adjoints either in text or in displayed equations. The expressions $F \dashv G$ and $G \vdash F$ and the diagrams

$$\mathsf{C} \mathrel{\substack{\xrightarrow{F} \\ \bot \\ \xleftarrow[G]{}}} \mathsf{D} \qquad \mathsf{C} \mathrel{\substack{\xleftarrow{G} \\ \top \\ \xrightarrow[F]{}}} \mathsf{D} \qquad \mathsf{D} \mathrel{\substack{\xleftarrow{F} \\ \bot \\ \xrightarrow[G]{}}} \mathsf{C} \qquad \mathsf{D} \mathrel{\substack{\xrightarrow{G} \\ \top \\ \xleftarrow[F]{}}} \mathsf{C}$$

all assert that $F \colon \mathsf{C} \to \mathsf{D}$ is left adjoint to $G \colon \mathsf{D} \to \mathsf{C}$.

Example 4.1.6. The forgetful functor $U \colon \mathsf{Top} \to \mathsf{Set}$ admits both left and right adjoints. To define the left adjoint, one needs to construct a topological space from a set S so that continuous maps from this space to another space T correspond naturally and bijectively to functions $S \to UT$. The discrete topology on S has this universal property: writing $D(S)$ for this space, any function $S \to U(T)$ defines a continuous map $D(S) \to T$. Similarly, to define the right adjoint, one needs to construct a topological space from a set S so that continuous maps from T to this space correspond naturally and bijectively to functions $UT \to S$. The indiscrete topology on S has this universal property: write $I(S)$ for this space, any function $U(T) \to S$ defines a continuous function $T \to I(S)$. So we have natural isomorphisms

$$\mathsf{Top}(D(S), T) \cong \mathsf{Set}(S, U(T)) \qquad \mathsf{Set}(U(T), S) \cong \mathsf{Top}(T, I(S)),$$

and thus the discrete, forgetful, and indiscrete functors define a pair of adjunctions

$$\mathsf{Top} \mathrel{\substack{\xleftarrow{D} \\ \bot \\ \xrightarrow{U} \\ \bot \\ \xleftarrow[I]{}}} \mathsf{Set}$$

Similarly, the forgetful functors ob: $\mathsf{Cat} \to \mathsf{Set}$ and Vert: $\mathsf{Graph} \to \mathsf{Set}$ (for any variety of graphs) admit both left and right adjoints. We leave it to the reader to work out what these are.

An adjunction between poset categories is called a (monotone) **Galois connection**. If A and B are posets, functors $F \colon \mathsf{A} \to \mathsf{B}$ and $G \colon \mathsf{B} \to \mathsf{A}$ are simply order-preserving functions. We have $F \dashv G$ if and only if

$$Fa \leq b \quad \text{if and only if} \quad a \leq Gb$$

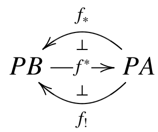
for all $a \in \mathsf{A}$ and $b \in \mathsf{B}$. In this context, F is often called the **lower adjoint** and G is called the **upper adjoint**.

Example 4.1.7. Both adjoints to inclusion of posets $\mathbb{Z} \hookrightarrow \mathbb{R}$ (with the usual $\leq$ ordering) exist defining the ceiling and floor functions. For any integer n and real number r, $n \leq r$ if and only if $n \leq \lfloor r \rfloor$, where $\lfloor r \rfloor$ denotes the greatest integer less than or equal to r. This function is order-preserving, defining a functor $\lfloor - \rfloor \colon \mathbb{R} \to \mathbb{Z}$ that is right adjoint to the inclusion $\mathbb{Z} \hookrightarrow \mathbb{R}$. Dually, $r \leq n$ if and only if $\lceil r \rceil \leq n$, and this order-preserving function defines a functor $\lceil - \rceil \colon \mathbb{R} \to \mathbb{Z}$ that is left adjoint to the inclusion.

$$\mathbb{Z} \mathrel{\substack{\xleftarrow{\lceil - \rceil} \\ \bot \\ \hookrightarrow \\ \bot \\ \xleftarrow[\lfloor - \rfloor]{}}} \mathbb{R}$$

Example 4.1.8. Consider a function $f \colon A \to B$ between sets. The subsets of A and subsets of B form posets, PA and PB, ordered by inclusion. The map f induces direct image and inverse image functors $f_*: PA \to PB$ and $f^*: PB \to PA$; as functions both maps are order-preserving and thus define functors between the poset categories. The direct image is left adjoint to the inverse image: for $A' \subset A$ and $B' \subset B$, $f(A') \subset B'$ if and only if $A' \subset f^{-1}(B')$.

Moreover, the inverse image functor has a further right adjoint $f_!$ that carries a subset $A' \subset A$ to the set of elements of B whose fibers lie entirely in A'. With this definition, $B' \subset f_!(A')$ if and only if $f^{-1}(B') \subset A'$. So we again have a pair of adjunctions



EXAMPLE 4.1.9. A propositional function is a function $P: X \to \Omega = \{\perp, \top\}$, which decides, for each $x \in X$, whether $P(x)$ is true or false. The logical operations of universal and existential quantification define functions $\forall_{x \in X}, \exists_{x\in X}: \Omega^X \to \Omega$ in the expected way: $\forall_{x\in X} P(x) = \top$ if and only if $P(x) = \top$ for all $x \in X$. These functions become functors between poset categories when Ω is given the partial order $\perp \leq \top$ and Ω^X inherits a pointwise defined order: $P \leq Q$ if and only if $P(x) \leq Q(x)$ for all $x \in X$. There is also a constant "dummy variable" functor $\Delta_{x\in X}: \Omega \to \Omega^X$. One can verify that $\exists_{x\in X} \dashv \Delta_{x\in X} \dashv \forall_{x\in X}$. See [**Awo96**] for more.

EXAMPLES 4.1.10. There is a large and very important family of "free-forgetful" adjunctions, with the forgetful functor defining the right adjoint and the free functor defining the left adjoint. In general, if one can construct the "free" object of type X on an object of type Y, this construction most likely defines a left adjoint to the forgetful functor from the category of Xs to the category of Ys. The following forgetful functors admit left adjoints, defining "free" constructions.

(i) U: Group $\to$ Set
(ii) U: Ab $\to$ Set
(iii) U: Ring $\to$ Set
(iv) U: $\mathsf{Mod}_R \to$ Set, generalizing the case U: $\mathsf{Vect}_{\Bbbk} \to$ Set considered in the introduction to this Chapter.
(v) U: $\mathsf{Mod}_R \to$ Ab, forgetting the scalar multiplication
(vi) U: Ring $\to$ Ab, forgetting the multiplicative structure[3]
(vii) $(-)^\times$: Ring $\to$ Group, carrying a ring to its group of units.
(viii) ϕ^*: $\mathsf{Mod}_S \to \mathsf{Mod}_R$ induced from a ring homomorphism $\phi: R \to S$. The forgetful functor is called **restriction of scalars**.
(ix) ϕ^*: $\mathsf{Set}^G \to \mathsf{Set}^H$ induced from a group homomorphism $\phi: H \to G$, another restriction-of-scalars functor
(x) U: $\mathsf{Set}_* \to$ Set

Each of these free-forgetful adjunctions is an instance of a **monadic adjunction**, which will be introduced in Chapter 5.

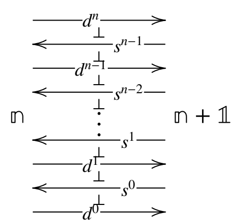
EXAMPLE 4.1.11. None of the functors U: Field $\to$ Ring, U: Field $\to$ Ab, $(-)^\times$: Field $\to$ Ab, U: Field $\to$ Set that forget algebraic structure on fields admit left adjoints.

EXAMPLE 4.1.12. The forgetful functor U: Cat $\to$ DirGraph also admits a left adjoint F, defining the **free category** on a directed graph. A directed graph G consists of a set V of

[3] While a unital ring has an underlying (multiplicative) monoid, containing every element except the zero element, this construction does not define a functor Ring $\to$ Mon because a ring homomorphism might have a non-trivial kernel. Field homomorphisms differ in this regard; see Example 4.1.11.

vertices, a set E of edges, and two functions $s, t\colon E \to V$ defining the source and target of each directed edge. The free category on G has V as its set of objects. The set of morphisms consists of identities for each vertex together with finite paths of edges; here a "path" must respect direction. Composition is defined by concatenation of paths. The adjunction supplies a natural bijection between functors $F(G) \to \mathsf{C}$ and morphisms $G \to U(\mathsf{C})$ of directed graphs. The latter presentation defines a diagram in C with no commutativity requirements, since directed graphs do not encode composites. For instance, diagrams indexed by the directed graph ω, first introduced in Example 1.1.3.(iv), have this form. The adjunction tells us that diagrams without commutativity requirements are precisely diagrams indexed by free categories.

Example 4.1.13. Let $\mathbb{n} + \mathbb{1}$ denote the ordinal category, freely generated by the graph $0 \to 1 \to 2 \to \cdots \to n$. For each $0 \leq i \leq n$, there is an injective functor $d^i\colon \mathbb{n} \to \mathbb{n} + \mathbb{1}$ where $i \in \mathbb{n}$ is the unique object missing from the image. For each $0 \leq i < n$, there is also a surjective functor $s^i\colon \mathbb{n} + \mathbb{1} \to \mathbb{n}$ for which $i \in \mathbb{n}$ is the unique object with two preimages. These functors define a sequence of $2n + 1$ adjoints



Definition 4.1.1 can be dualized in three ways, by replacing C, D, or both C and D by their opposite categories. The latter dualization recovers the original notion of adjunction, with $G^{\mathrm{op}}\colon \mathsf{D}^{\mathrm{op}} \to \mathsf{C}^{\mathrm{op}}$ left adjoint to $F^{\mathrm{op}}\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}^{\mathrm{op}}$ if and only if $F \dashv G$. In particular, any theorem about left adjoints has a dual theorem about right adjoints, which is a very useful duality principle for adjunctions. The other two dualizations lead to new (but still dual) types of adjoint functors.

Definition 4.1.14. A pair of contravariant functors $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{D}$ and $G\colon \mathsf{D}^{\mathrm{op}} \to \mathsf{C}$ are **mutually left adjoint** if there exists a natural isomorphism

$$\mathsf{D}(Fc, d) \cong \mathsf{C}(Gd, c),$$

or **mutually right adjoint** if there exists a natural isomorphism

$$\mathsf{D}(d, Fc) \cong \mathsf{C}(c, Gd).$$

Example 4.1.15. Let $\mathsf{Axiom}_{\mathcal{L}}$ be a set of **axioms**, i.e., sentences in a fixed first-order language $\mathcal{L}$. Let $\mathsf{Model}_{\mathcal{L}}$ be a set of **models** for that language, i.e., sets with interpretations of the constants, relations, and functions in the language. For instance, the language of the natural numbers has a constant symbol "0", a binary function symbol "+", and a binary relation symbol "≤", so a model would be any set with a specified constant, binary function, and binary relation. Given a set of models M and a set of axioms A we write $M \vDash A$ if each of the axioms in A is **satisfied by**, that is, true in, each of the models in M. For instance, an axiom might assert the transitivity of the relation symbol "≤", in which case a model satisfies this axiom if and only if its interpretation of this relation is transitive.

Form the poset categories $P(\mathsf{Axiom}_{\mathcal{L}})$ and $P(\mathsf{Model}_{\mathcal{L}})$ ordered by inclusion. There are contravariant functors

$$\text{True in}\colon P(\mathsf{Axiom}_{\mathcal{L}})^{\mathrm{op}} \to P(\mathsf{Model}_{\mathcal{L}}) \qquad \text{Satisfying}\colon P(\mathsf{Model}_{\mathcal{L}})^{\mathrm{op}} \to P(\mathsf{Axiom}_{\mathcal{L}})$$

which send a set of axioms A to the set of models that satisfy those axioms and send a set of models to the set of axioms that they satisfy. These are mutual right adjoints, forming what is called the **Galois connection between syntax and semantics** [**Smi**].

Exercises.

EXERCISE 4.1.1. Prove Lemma 4.1.2 and remark 4.1.4.

EXERCISE 4.1.2. Define left and right adjoints to $\mathrm{ob}\colon \mathsf{Cat} \to \mathsf{Set}$ and to $\mathrm{Vert}\colon \mathsf{Graph} \to \mathsf{Set}$, the functor that takes a graph to its vertex set, for either directed or undirected graphs, as you prefer.

EXERCISE 4.1.3. Suppose that $F\colon \mathsf{A} \times \mathsf{B} \to \mathsf{C}$ is a bifunctor so that for each object $a \in \mathsf{A}$, the induced functor $F(a,-)\colon \mathsf{B} \to \mathsf{C}$ admits a right adjoint $G_a\colon \mathsf{C} \to \mathsf{B}$.

(i) Show that these right adjoints assemble into a unique bifunctor $G\colon \mathsf{A}^{\mathrm{op}} \times \mathsf{C} \to \mathsf{B}$, defined so that $G(a,c) = G_a(c)$ so that the isomorphisms

$$\mathsf{C}(F(a,b),c) \cong \mathsf{B}(b,G(a,c))$$

are natural in all three variables.

(ii) Suppose further that for each $b \in \mathsf{B}$, the induced functor $F(-,b)\colon \mathsf{A} \to \mathsf{C}$ admits a right adjoint $H_b\colon \mathsf{C} \to \mathsf{A}$. Conclude that there is a unique bifunctor $H\colon \mathsf{B}^{\mathrm{op}} \times \mathsf{C} \to \mathsf{A}$ so that $H(b,c) = H_b(c)$ and the isomorphism

$$\mathsf{A}(a,H(b,c)) \cong \mathsf{C}(F(a,b),c) \cong \mathsf{B}(b,G(a,c)) \tag{4.1.16}$$

is natural in all three variables.

(iii) Conclude that for each $c \in \mathsf{C}$, the functors $G(-,c)\colon \mathsf{A}^{\mathrm{op}} \to \mathsf{B}$ and $H(-,c)\colon \mathsf{B}^{\mathrm{op}} \to \mathsf{A}$ are mutual right adjoints.

A triple of bifunctors F, G, and H equipped with a natural isomorphism (4.1.16) is called a **two-variable adjunction**.

EXERCISE 4.1.4. What are some examples of two-variable adjunctions?

4.2. The unit and counit as universal arrows

Consider an adjunction

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\leftarrow}{\overset{\rightarrow}{\bot}}}} \mathsf{D} \qquad\qquad \mathsf{D}(Fc,d) \cong \mathsf{C}(c,Gd).$$

Fixing $c \in \mathsf{C}$, the defining natural isomorphism says that the object $Fc \in \mathsf{D}$ represents the functor $\mathsf{C}(c,G-)\colon \mathsf{D} \to \mathsf{Set}$. By the Yoneda lemma, the natural isomorphism $\mathsf{D}(Fc,-) \cong \mathsf{C}(c,G-)$ is determined by an element of $\mathsf{C}(c,GFc)$, the transpose of 1_{Fc}, that we denote by η_c. By Lemma 4.1.2, the maps η_c assemble into the components of a natural transformation $\eta\colon 1_{\mathsf{C}} \Rightarrow GF$.

LEMMA 4.2.1. *Given an adjunction $F \dashv G$, there is a natural transformation $\eta\colon 1_{\mathsf{C}} \Rightarrow GF$, called the **unit** of the adjunction, whose component $\eta_c\colon c \to GFc$ at c is defined to be the transpose of the identity morphism 1_{Fc}.*

Proof. To prove that η is natural, we must show that the left-hand square commutes for every $f : c \to c'$ in C.

$$\begin{array}{ccc} c & \xrightarrow{\eta_c} & GFc \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle GFf} \\ c' & \xrightarrow[\eta_{c'}]{} & GFc' \end{array} \quad \leftrightsquigarrow \quad \begin{array}{ccc} Fc & \xrightarrow{1_{Fc}} & Fc \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Ff} \\ Fc' & \xrightarrow[1_{Fc'}]{} & Fc' \end{array}$$

This follows from Lemma 4.1.2 and the obvious commutativity of the right-hand transposed square. □

Dually, fixing $d \in \mathsf{D}$ the defining natural isomorphism of an adjunction $F \dashv G$ says that the object $Gd \in \mathsf{C}$ represents the functor $\mathsf{D}(F-, d) : \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$. By the Yoneda lemma, the natural isomorphism $\mathsf{C}(-, Gd) \cong \mathsf{D}(F-, d)$ is determined by an element of $\mathsf{D}(FGd, d)$, the transpose of 1_{Gd}, that we denote by ϵ_d. By Lemma 4.1.2, the maps ϵ_d assemble into the components of a natural transformation $\epsilon : FG \Rightarrow 1_{\mathsf{D}}$. We have the following dual of Lemma 4.2.1

Lemma 4.2.2. *Given an adjunction $F \dashv G$, there is a natural transformation $\epsilon : FG \Rightarrow 1_{\mathsf{D}}$, called the* ***counit*** *of the adjunction, whose component $\epsilon_d : FGd \to d$ at d is defined to be the transpose of the identity morphism 1_{Gd}.*

Lemmas 4.2.1 and 4.2.2 show that any adjunction has a unit and a counit. Conversely, if $F : \mathsf{C} \rightleftarrows \mathsf{D} : G$ are opposing functors equipped with natural transformations $\eta : 1_{\mathsf{C}} \Rightarrow GF$ and $\epsilon : FG \Rightarrow 1_{\mathsf{D}}$ satisfying a dual pair of conditions, then this data provides an alternate encoding of the notion of an adjunction.

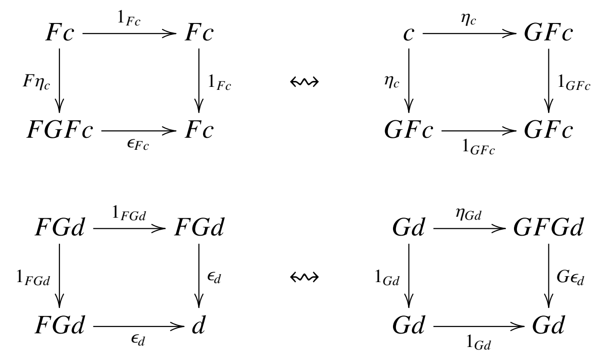
Definition 4.2.3. An **adjunction** consists of an opposing pair of functors $F : \mathsf{C} \rightleftarrows \mathsf{D} : G$, together with natural transformations $\eta : 1_{\mathsf{C}} \Rightarrow GF$ and $\epsilon : FG \Rightarrow 1_{\mathsf{D}}$ that satisfy the **triangle identities**:

$$\begin{array}{ccc} F & \xRightarrow{F\eta} & FGF \\ & {\scriptstyle 1_F}\searrow & \Downarrow{\scriptstyle \epsilon F} \\ & & F \end{array} \qquad\qquad \begin{array}{ccc} G & \xRightarrow{\eta G} & GFG \\ & {\scriptstyle 1_G}\searrow & \Downarrow{\scriptstyle G\epsilon} \\ & & G \end{array}$$

The left-hand triangle asserts that a certain diagram commutes in D^{C}, while the right-hand triangle asserts that the dual diagram commutes in C^{D}. The natural transformations $F\eta$, ϵF, ηG, and $G\epsilon$ are defined by whiskering; recall Remark 1.6.15. Together these equations assert that "the counit is a left inverse of the unit modulo translation." They can't literally be inverses because the components of the unit $\eta_c : c \to GFc$ lie in C while the components of the counit $\epsilon_d : FGd \to d$ lie in D; these morphisms aren't composable. But if we apply F to the unit, we obtain a morphism $F\eta_c : Fc \to FGFc$ in D whose left inverse is ϵ_{Fc}. And if we apply G to the counit, we obtain a morphism $G\epsilon_d : GFGd \to Gd$ whose right inverse is η_{Gd}.

Proposition 4.2.4. *A pair of functors $F : \mathsf{C} \rightleftarrows \mathsf{D} : G$ is adjoint in the sense of Definition 4.1.1 if and only if they are adjoint in the sense of Definition 4.2.3.*

Proof. Lemmas 4.2.1 and 4.2.2 demonstrate that a natural isomorphism $\mathsf{D}(Fc, d) \cong \mathsf{C}(c, Gd)$ can be used to define natural transformations $\eta : 1_{\mathsf{C}} \Rightarrow GF$ and $\epsilon : FG \Rightarrow 1_{\mathsf{D}}$, whose components are transposes of identity morphisms. It remains to demonstrate the triangle identities, for which it suffices by Lemma 1.6.12 to consider the components of each natural transformation. By Lemma 4.1.2, the upper-left-hand and lower-right-hand squares

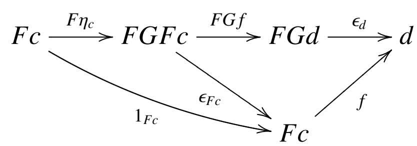


commute because the transposed squares manifestly do. This is what we wanted to show.

Conversely, the unit and counit can be used to define a natural bijection $\mathsf{D}(Fc, d) \cong \mathsf{C}(c, Gd)$. Given $f\colon Fc \to d$ and $g\colon c \to Gd$ their adjuncts are defined to be the composites:

$$f^\flat := \ c \xrightarrow{\eta_c} GFc \xrightarrow{Gf} Gd \qquad\qquad g^\sharp := \ Fc \xrightarrow{Fg} FGd \xrightarrow{\epsilon_d} d$$

The triangle identities are used to show that these operations are inverses. By definition $(f^\flat)^\sharp$ is equal to the top composite



By naturality of ϵ and one triangle identity, this equals f. The dual diagram chase, which demonstrates that $(g^\sharp)^\flat = g$, follows or is left as an exercise for the reader. □

Corollary 4.2.5. *If* A *and* B *are posets and* $F\colon \mathsf{A} \to \mathsf{B}$ *and* $G\colon \mathsf{B} \to \mathsf{A}$ *form a Galois connection, with* $F \dashv G$*, then* F *and* G *satisfy the following fixed point formulas*

$$FGF = F \qquad GFG = G.$$

Proof. By the triangle identities $F(a) \leq FGF(a) \leq F(a)$ for all $a \in \mathsf{A}$, whence $F = FGF$. The other formula is dual. □

Example 4.2.6. The left adjoint F to the forgetful functor $U\colon \mathsf{Group} \to \mathsf{Set}$ defines the **free group** on a set S. Elements of the group $F(S)$ are finite words written in the letters $s \in S$ and formal "inverses" s^{-1}, modulo some evident relations. The empty word serves as the identity element and multiplication is by concatenation. The component of the unit of the adjunction $F \dashv U$ at a set S is the function $S \to UF(S)$ that sends an element to the corresponding singleton word. The component of the counit of the adjunction at a group G is the group homomorphism $FU(G) \to G$ that sends a word, whose letters are elements of the group G and formal inverses, to the product of those symbols, interpreted using the multiplication, inverses, and identity present in the group G. Note that this function sends a singleton word to that group element; this proves one of the triangle identities. We leave the interpretation of the other triangle identity to the reader.

Exercises.

Exercise 4.2.1. Explain each step needed to convert the statement of Lemma 4.2.1 into the statement of Lemma 4.2.2.

Exercise 4.2.2. Pick your favorite forgetful functor from Example 4.1.10 and prove that it is a right adjoint by defining its left adjoint, the unit, and the counit, and demonstrating that the triangle identities hold.

Exercise 4.2.3. Each component of the counit of an adjunction is a terminal object in some category. What category?

Exercise 4.2.4. A **morphism of adjunctions**, from $F \dashv G$ to $F' \dashv G'$ is comprised of a pair of functors

$$\begin{array}{ccc} \mathsf{C} & \xrightarrow{H} & \mathsf{C}' \\ F \downarrow \dashv \uparrow G & & F' \downarrow \dashv \uparrow G' \\ \mathsf{D} & \xrightarrow[K]{} & \mathsf{D}' \end{array}$$

so that the square with the left adjoints and the square with the right adjoints both commute (i.e., $KF = F'H$ and $HG = G'K$) and satisfying one additional condition, which takes a number of equivalent forms. Prove that the following are equivalent:

(i) $H\eta = \eta' H$, where η and η' denote the respective units of the adjunctions.
(ii) $K\epsilon = \epsilon' K$, where ϵ and ϵ' denote the respective counits of the adjunctions.
(iii) Transposition across the adjunctions commutes with application of the functors H and K, i.e., for every $c \in \mathsf{C}$ and $d \in \mathsf{D}$, the diagram

$$\begin{array}{ccc} \mathsf{D}(Fc, d) & \xrightarrow{\cong} & \mathsf{C}(c, Gd) \\ {\scriptstyle K}\downarrow & & \downarrow{\scriptstyle H} \\ \mathsf{D}'(KFc, Kd) & & \mathsf{C}'(Hc, HGd) \\ \| & & \| \\ \mathsf{D}'(F'Hc, Kd) & \xrightarrow[\cong]{} & \mathsf{C}'(Hc, G'Kd) \end{array}$$

commutes.

4.3. Formal facts about adjunctions

"Formal" means 2-categorical, meaning that they can be proven syntactically using objects (categories), morphisms (functors), and morphisms of morphisms (natural transformations), which can be composed horizontally and vertically, as we saw in Section 1.6. This is the reason for giving diagrammatic proofs of results that can be proven more efficiently by appealing to the Yoneda lemma: the diagrammatic proofs apply more generally.[4]

Proposition 4.3.1. *If F and F' are left adjoint to U, then $F \cong F'$. Moreover, there is a unique such isomorphism commuting with the unit and counit of the adjunction.*

[4]Unfortunately, we won't be able to describe the other contexts in which the diagrammatic proofs apply.

Proof 1. To define a natural transformation $\theta\colon F \Rightarrow F'$, it suffices, by Lemma 4.1.2, to define a transposed natural transformation, which we take to be $\eta'\colon 1_{\mathsf{C}} \Rightarrow UF'$

$$\begin{array}{ccc} Fc & \xrightarrow{\theta_c} & F'c \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle F'f} \\ Fc' & \xrightarrow[\theta_{c'}]{} & F'c' \end{array} \qquad \leftrightsquigarrow \qquad \begin{array}{ccc} c & \xrightarrow{\eta'_c} & UF'c \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle UF'f} \\ c' & \xrightarrow[\eta'_{c'}]{} & UF'c' \end{array}$$

The proof of Proposition 4.2.4 provides a formula for θ:

$$\theta := \ F \xRightarrow{F\eta'} FUF' \xRightarrow{\epsilon F'} F'.$$

Exchanging the roles of F' and F, we also define a natural transformation $\theta'\colon F' \Rightarrow F$ to be the transpose of $\eta\colon 1_{\mathsf{C}} \to UF$, i.e., by the formula

$$\theta' := \ F' \xRightarrow{F'\eta} F'UF \xRightarrow{\epsilon' F} F.$$

The hope is that $\theta' = \theta^{-1}$, so that θ is a natural isomorphism. To prove that

$$F \xRightarrow{F\eta'} FUF' \xRightarrow{\epsilon F'} F' \xRightarrow{F'\eta} F'UF \xRightarrow{\epsilon' F} F \ = \ F \xRightarrow{1_F} F$$

it suffices to prove that the transposed natural transformations are equal, i.e., to show that $\eta\colon 1 \Rightarrow UF$ equals the composite

$$1 \xRightarrow{\eta} UF \xRightarrow{UF\eta'} UFUF' \xRightarrow{U\epsilon F'} UF' \xRightarrow{UF'\eta} UF'UF \xRightarrow{U\epsilon' F} UF\ .$$

By naturality of η, this equals

$$1 \xRightarrow{\eta'} UF' \xRightarrow{\eta UF'} UFUF' \xRightarrow{U\epsilon F'} UF' \xRightarrow{UF'\eta} UF'UF \xRightarrow{U\epsilon' F} UF\ .$$

By the triangle identity $U\epsilon F' \cdot \eta UF' = 1_{UF'}$, this simplifies to

$$1 \xRightarrow{\eta'} UF' \xRightarrow{UF'\eta} UF'UF \xRightarrow{U\epsilon' F} UF\ .$$

By naturality of η', this equals

$$1 \xRightarrow{\eta} UF \xRightarrow{\eta' UF} UF'UF \xRightarrow{U\epsilon' F} UF\ ,$$

and by the triangle identity $U\epsilon' F \cdot \eta' UF = 1_{UF}$ this simplifies to:

$$1 \xRightarrow{\eta} UF$$

as desired. The other diagram chase, proving that $\theta \cdot \theta' = 1_{F'}$, is dual.

From the formula for θ, there are easy diagram chases, left to the reader, that verify that the diagrams of natural transformations commute

$$\begin{array}{ccc} 1_{\mathsf{C}} & \xRightarrow{\eta} & UF \\ & {\scriptstyle \eta'}\searrow & \Downarrow{\scriptstyle U\theta} \\ & & UF' \end{array} \qquad\qquad \begin{array}{ccc} FU & \xRightarrow{\epsilon} & 1_{\mathsf{D}} \\ {\scriptstyle \theta U}\Downarrow & \nearrow{\scriptstyle \epsilon'} & \\ F'U & & \end{array}$$

The left-hand triangle asserts that the transpose of θ across the adjunction $F \dashv U$ is η', proving uniqueness. ∎

PROOF 2. By the Yoneda lemma, the composite natural isomorphism

$$\mathsf{D}(F'c, d) \cong \mathsf{C}(c, Ud) \cong \mathsf{D}(Fc, d),$$

defines a natural isomorphism $\theta_c \colon Fc \cong F'c$. The component θ_c is defined to be the image of $1_{F'c}$. The first isomorphism carries this to $\eta'_c \colon c \to UF'c$. The second isomorphism carries it to its transpose along $F \dashv U$. Thus

$$\theta_c := \; Fc \xrightarrow{F\eta'_c} FUF'c \xrightarrow{\epsilon F'_c} F'c$$

This is the formula we discovered in Proof 1. Lemma 4.1.2 can be used to show commutativity with the units and counits. □

PROPOSITION 4.3.2. *Given adjunctions*

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D} \underset{G'}{\overset{F'}{\underset{\perp}{\rightleftarrows}}} \mathsf{E}$$

the composite $F'F$ is left adjoint to the composite GG'.

There are two proofs: a Yoneda-style proof and a 2-categorical proof.

PROOF 1. There are natural isomorphisms

$$\mathsf{E}(F'Fc, e) \cong \mathsf{D}(Fc, G'e) \cong \mathsf{C}(c, GG'e),$$

the first defined using $F' \dashv G'$ and the second defined using $F \dashv G$. □

PROOF 2. The unit and counit of $F'F \dashv GG'$ are defined to be the composites

$$\bar{\eta} := \; 1_{\mathsf{C}} \xRightarrow{\eta} GF \xRightarrow{G\eta' F} GG'F'F \qquad\qquad \bar{\epsilon} := \; F'FGG' \xRightarrow{F'\epsilon G'} F'G' \xRightarrow{\epsilon'} 1_{\mathsf{E}}$$

The proof of the triangle identities is an entertaining diagram chase left to the reader. □

Using the formula for adjunct arrows given in the proof of Proposition 4.2.4, the first proof gives the formulas for the unit and counit provided by the second.

PROPOSITION 4.3.3. *Any equivalence*

$$\mathsf{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathsf{D} \qquad \eta \colon \mathbb{1}_{\mathsf{C}} \overset{\cong}{\Rightarrow} GF, \quad \epsilon \colon FG \overset{\cong}{\Rightarrow} 1_{\mathsf{D}}$$

can be promoted to an adjoint equivalence, in which the natural isomorphisms satisfy the triangle identities, by redefining either the unit or counit.

The proof is left as Exercise 4.3.1.

PROPOSITION 4.3.4. *Given an adjunction*

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D}$$

post-composition with F and G define a pair of adjoint functors

$$\mathsf{C}^{\mathsf{J}} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D}^{\mathsf{J}}$$

for any small category J.

This can be proven by appealing to either of the definitions of an adjunction.

Exercises.

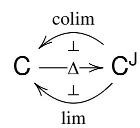
EXERCISE 4.3.1. Prove Proposition 4.3.3.

EXERCISE 4.3.2. Use the natural bijection between hom-sets to prove Proposition 4.3.4.

EXERCISE 4.3.3. Use the unit and counit associated to an adjunction to prove Proposition 4.3.4.

4.4. Adjunctions, limits, and colimits

PROPOSITION 4.4.1. *A category* C *admits all limits of diagrams indexed by a small category* J *if and only if the constant diagram functor* $\Delta\colon \mathsf{C} \to \mathsf{C}^{\mathsf{J}}$ *admits a right adjoint. Dually,* C *admits all colimits of* J*-indexed diagrams if and only if* Δ *admits a left adjoint.*



Recall from Proposition 3.5.2 that the axiom of choice is needed to define the action of the limit or colimit functor on objects.

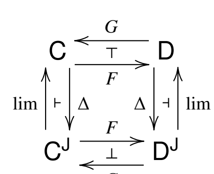
PROOF. These dual statements follow immediately from the defining universal properties of a limit and colimit. For $c \in \mathsf{C}$ and $F \in \mathsf{C}^{\mathsf{J}}$, the hom-set $\mathsf{C}^{\mathsf{J}}(\Delta c, F)$ is the set of natural transformations from the constant J-diagram at c to the diagram F. This, by Definition 3.1.2, is precisely the set of cones over F with summit c. There is an object $\lim F \in \mathsf{C}$ together with a natural isomorphism

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

if and only if this limit exists. □

THEOREM 4.4.2 (RAPL). *Right adjoints preserve limits.*

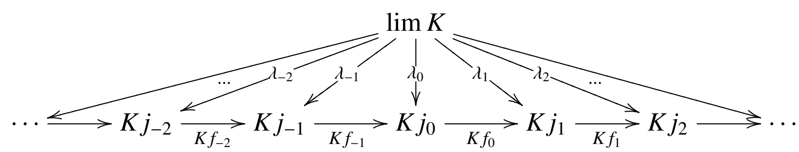
If $G\colon \mathsf{D} \to \mathsf{C}$ has a left adjoint F and if D and C admit all limits indexed by a category J, then there is a slick proof of this result. The left adjoint functors



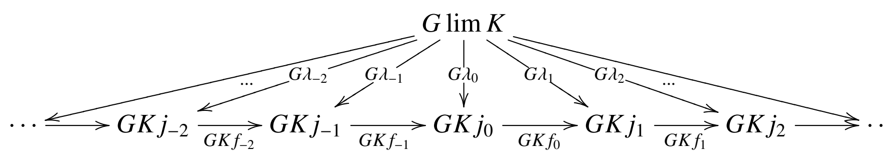
manifestly commute. By Proposition 4.3.2, the functors $G \lim$ and $\lim G$ are both right adjoints to the composite $F\Delta = \Delta F$. By Proposition 4.3.1 they must therefore be isomorphic, proving that G preserves J-indexed limits.

However, we want a more precise version of Theorem 4.4.2 that tells us that a right adjoint preserves any limit that happens to exist in its domain, regardless of whether or not all limits of that shape exist.

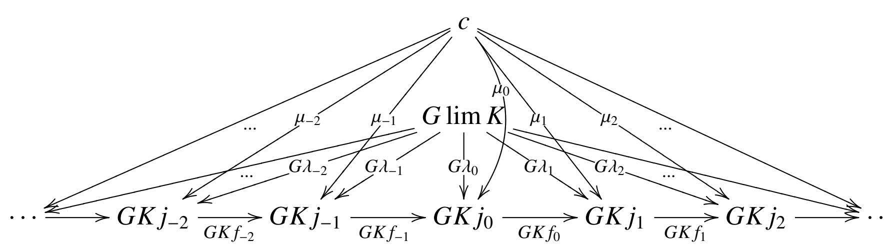
Proof. Consider a diagram $K\colon \mathsf{J} \to \mathsf{D}$ admitting a limit cone, illustrated here in the case where J is the poset of integers.



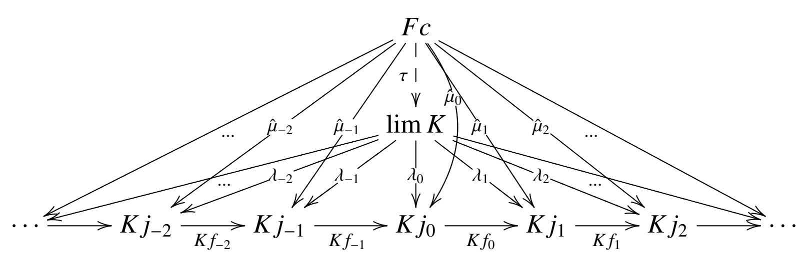
Applying a right adjoint $G\colon \mathsf{D} \to \mathsf{C}$ we obtain a cone



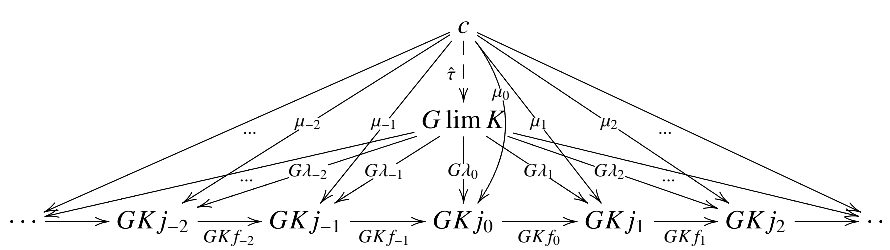
in C. We claim that this is a limit cone over the diagram $GK\colon \mathsf{J} \to \mathsf{C}$. To prove this, consider another cone:



The legs of this cone transpose to define maps in D



which define a cone over K by Lemma 4.1.2. The universal property of the limit cone induces a unique factorization τ of the cone $\hat{\mu}$ through the cone λ. The map τ transposes to define a factorization



of the cone μ through the cone $G\lambda$, again by applying Lemma 4.1.2. This factorization is clearly unique: another such map would transpose to define a factorization of $\hat{\mu}$ through λ, which would necessarily equal τ, and $\hat{\tau}$ is the unique adjunct of this map. Thus, $G\lambda\colon G\lim K \Rightarrow GK$ is a limit cone for GK. ◻

In summary, we have just argued that

$$\mathsf{C}^{\mathsf{J}}(\Delta c, GK) \cong \mathsf{D}^{\mathsf{J}}(F\Delta c, K) \cong \mathsf{D}^{\mathsf{J}}(\Delta Fc, K) \cong \mathsf{D}(Fc, \lim_{\mathsf{J}} K) \cong \mathsf{C}(c, G \lim_{\mathsf{J}} K),$$

which, by the defining universal property of the limit, says that $G \lim_{\mathsf{J}} K$ defines a limit for the diagram GK.

Dually, of course, we have:

Theorem 4.4.3 (LAPC). *Left adjoints preserve colimits.*

These results have myriad useful corollaries. For instance:

Corollary 4.4.4. *For any function $f \colon A \to B$, the inverse image $f^{-1} \colon P(B) \to P(A)$, a function between the powersets of A and B, preserves both unions and intersections, while the direct image $f_* \colon P(A) \to P(B)$ only preserves unions.*

Corollary 4.4.5. *For any vector spaces U, V, W, $U \otimes (V \oplus W) \cong (U \otimes V) \oplus (U \otimes W)$.*

Similarly, Theorems 4.4.2 and 4.4.3 supply proofs of many of the basic operations in arithmetic on account of the adjunctions $A \times - \dashv (-)^A$ between the product and exponential on Set or its subcategory Fin.

Corollary 4.4.6. *For finite sets A, B, C we have natural isomorphisms*

$$A \times (B + C) \cong (A \times B) + (A \times C) \qquad (B \times C)^A \cong B^A \times C^A \qquad A^{B+C} \cong A^B \times A^C$$

Proof. The left adjoint $A \times -$ preserves coproducts, the right adjoint $(-)^A$ preserves products, and the functor $A^- \colon \mathsf{Fin}^{\mathrm{op}} \to \mathsf{Fin}$, which is mutually right adjoint to itself, carries coproducts in Fin to products in Fin. □

The forgetful functors of Example 4.1.10 carry any limits that exist in the categories of groups, rings, modules, and so forth to corresponding limits of their underlying sets. Indeed, we will show in Chapter 5 that these forgetful functors create, and not merely preserve, all limits. Conversely, the fact that the left adjoints preserve all colimits in Set tells us that the construction of colimits in groups, rings, modules, and so forth is necessarily more complicated. For instance:

Corollary 4.4.7. *The free group on the set $X \coprod Y$ is the free product of the free groups on the sets X and Y.*

We now study a certain important class of adjoint functors.

Definition 4.4.8. A **reflective subcategory** of a category C is a full subcategory D so that the inclusion $\mathsf{D} \hookrightarrow \mathsf{C}$ admits a left adjoint $R \colon \mathsf{C} \to \mathsf{D}$, called the *reflector*.

In studying this notion, we will make use of the following straightforward lemma.

Lemma 4.4.9. *Consider an adjunction*

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D}$$

with counit $\epsilon \colon FG \Rightarrow 1_{\mathsf{D}}$. Then:

(i) G is faithful if and only if each component of ϵ is an epimorphism.
(ii) G is full if and only if each component of ϵ is a split monomorphism.
(iii) G is full and faithful if and only if ϵ is an isomorphism.

A dual result characterizes full or faithful left adjoints.

In particular, any object in a reflective subcategory $\mathsf{D} \hookrightarrow \mathsf{C}$ is isomorphic, via the counit $Rd \xrightarrow{\cong} d$ of the adjunction, to its reflection back into that subcategory. The components of the unit have the form $c \to Rc$; the mnemonic is that "an object looks at its reflection."

EXAMPLES 4.4.10. The following define reflective subcategories:

- The inclusion $\mathsf{kHaus} \hookrightarrow \mathsf{Top}$ of compact Hausdorff spaces into the category of all topological spaces. The reflector is the functor $\beta\colon \mathsf{Top} \to \mathsf{kHaus}$ sending a space to its **Stone-Čech compactification**.
- $U\colon \mathsf{Ab} \to \mathsf{Group}$. The reflector carries a group G to its **abelianization**, the quotient $G/[G,G]$ by the **commutator subgroup**, the normal subgroup generated by elements $ghg^{-1}h^{-1}$. The quotient maps $G \to G/[G,G]$ define the components of the unit. For any abelian group A, we have an isomorphism $A \cong A/[A,A]$; the commutator subgroup is trivial if and only if the group is abelian. The adjunction asserts that a homomorphism from a group G to an abelian group A necessarily factors through the abelianization of G.
- The inclusion $\mathsf{tfAb} \to \mathsf{Ab}$ of torsion-free abelian groups is reflective. The reflector sends an abelian group A to the quotient A/TA by its torsion subgroup. The quotient maps $A \to A/TA$ define the components of the unit, and any map from A to a torsion-free group necessarily factors through this quotient homomorphism.
- As described above, for any ring homomorphism $\phi\colon R \to T$, there exist adjoint functors

$$\mathsf{Mod}_T \underset{\phi^*}{\overset{T\otimes_R -}{\underset{\perp}{\leftrightarrows}}} \mathsf{Mod}_R,$$

  the right adjoint being restriction of scalars and the left adjoint being extension of scalars. The restriction of scalars functor is always faithful — as both categories of modules admit faithful forgetful functors to Ab — and is full if and only if $\phi\colon R \to T$ is an epimorphism. Note that epimorphisms in Ring include all surjections but are not only the surjections. An important class of epimorphisms include the **localizations**. Let $S \subset R$ be a monoid under multiplication; that is, S is a multiplicatively closed subset of the ring R. The localization $R \to S^{-1}R$ is an initial object in the category whose objects are ring homomorphisms $R \to T$ that carry all of the elements of S to units in T. For integral domains, the ring $S^{-1}R$ can be constructed as a field of fractions. A similar construction exists for general commutative rings. For non-commutative rings, the localization may or may not exist.

PROPOSITION 4.4.11. *If* $\mathsf{D} \hookrightarrow \mathsf{C}$ *is a reflective subcategory then*

(i) D *has all limits that* C *admits, formed as in* C.
(ii) D *has all colimits that* C *does, formed by applying the reflector to the colimit in* C.

By Theorems 4.4.2 and 4.4.3, if D has limits or colimits, they must be constructed in the way described in Proposition 4.4.11. The real content of this result is that these (co)limits necessarily exist in D. We defer the proof of (i) to Exercise 5.5.1, where it follows as a special case of a more general result to be proven there.

PROOF OF (II). For clarity, we introduce notation $i\colon \mathsf{D} \hookrightarrow \mathsf{C}$ for the inclusion, the right adjoint to the reflector R. Consider a diagram $F\colon \mathsf{J} \to \mathsf{D}$ and let $\lambda\colon iF \Rightarrow c$ be a colimit cone for the diagram iF in C. The left adjoint R sends this to a colimit cone $R\lambda\colon RiF \Rightarrow Rc$

in D. Now the counit supplies a natural isomorphism $Ri \cong 1_{\mathsf{D}}$. Transporting along these isomorphisms, we obtain a colimit cone $F \cong RiF \Rightarrow Rc$ for the original diagram in D. □

Exercises.

EXERCISE 4.4.1. When does the functor $!\colon \mathsf{C} \to \mathbb{1}$ have a left adjoint? When does it have a right adjoint?

EXERCISE 4.4.2. Suppose the diagonal functor $\Delta\colon \mathsf{C} \to \mathsf{C}^{\mathsf{J}}$ admits both left and right adjoints. Describe the units and counits of these adjunctions.

EXERCISE 4.4.3. Prove Lemma 4.4.9.

EXERCISE 4.4.4. Consider a reflective subcategory inclusion $\mathsf{D} \hookrightarrow \mathsf{C}$ with reflector $R\colon \mathsf{C} \to \mathsf{D}$.

(i) Show that $\eta R = R\eta$, and that these natural transformations are isomorphisms.
(ii) Show that an object $c \in \mathsf{C}$ is in the **essential image** of the inclusion $\mathsf{D} \hookrightarrow \mathsf{C}$, i.e., is isomorphic to an object in the subcategory D, if and only if η_c is an isomorphism.
(iii) Show that the essential image of D consists of those objects c that are **local** for the class of morphisms that is inverted by R. That is, c is in the essential image if and only if the precomposition functions

$$\mathsf{C}(b, c) \xrightarrow{f^*} \mathsf{C}(a, c)$$

are isomorphisms for all $f\colon a \to b$ in C for which Rf is an isomorphism in D.

4.5. Existence of adjoint functors

Does the inclusion $\mathsf{Ring} \hookrightarrow \mathsf{Rng}$ of unital rings into the category of possibly non-unital rings have any adjoints? A first strategy to probe a question of this form might be called the "initial and terminal objects test": by Theorems 4.4.2 and 4.4.3 a functor admitting a left adjoint must necessarily preserve all limits while a functor admitting a right adjoint must necessarily preserve all colimits. The ring $\mathbb{Z}$ is initial in Ring but not in Rng, so there can be no right adjoint. The zero ring is terminal in both categories, so a left adjoint to the inclusion might be possible. Indeed, Ring has and the inclusion preserves all limits. A limit-preserving functor is called **continuous**; a colimit-preserving functor is called **cocontinuous**.

The search for a left adjoint to $\mathsf{Ring} \hookrightarrow \mathsf{Rng}$ might be thought of as some sort of formulaic (that is to say functorial) optimization problem, whose aim is to adjoin a multiplicative unit to a possibly non-unital ring R in the "most efficient way possible." The Yoneda lemma can be used to make this intuition precise. We seek a unital ring R^* together with a natural isomorphism

$$\mathsf{Ring}(R^*, S) \cong \mathsf{Rng}(R, S)$$

for all unital rings S. That is, we seek a representation for the functor $\mathsf{Rng}(R, -)\colon \mathsf{Ring} \to \mathsf{Set}$. By Proposition 2.4.4, we seek an initial object in the category of elements $\int \mathsf{Rng}(R, -)$, whose objects are homomorphisms $R \to S$ whose codomain is a unital ring and whose morphisms are commutative triangles whose leg opposite R is a unital ring homomorphism. This category of elements is isomorphic to the comma category $R \downarrow \mathsf{Ring}$ of non-unital homomorphisms from R to a unital ring. The optimization problem is solved if we can find a unital ring R^* and ring homomorphism $R \to R^*$ that is initial in this category.[5]

The same line of reasoning proves the following general result.

Lemma 4.5.1. *A functor* $U\colon \mathsf{A} \to \mathsf{S}$ *admits a left adjoint if and only if for each* $s \in \mathsf{S}$ *the comma category* $s \downarrow U$ *has an initial object.*

Proof. The comma category $s \downarrow U$ is isomorphic to the category of elements for the functor $\mathsf{S}(s, U-)\colon \mathsf{A} \to \mathsf{Set}$. If a left adjoint F exists, then the component at s of its unit defines an initial object $\eta_s\colon s \to UFs$ in this category; see Exercise 3.7.1. Conversely, suppose each $s \downarrow U$ admits an initial object, which we suggestively denote by $\eta_s\colon s \to UFs$. This defines the value of a function $F\colon \mathrm{ob}\,\mathsf{S} \to \mathrm{ob}\,\mathsf{A}$ which we now extend to a functor. For each morphism $f\colon s \to s'$ in S, define $Ff\colon Fs \to Fs'$ to be the unique morphism in A making the square

$$\begin{array}{ccc} s & \xrightarrow{\eta_s} & UFs \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle UFf} \\ s' & \xrightarrow[\eta_{s'}]{} & UFs' \end{array}$$

commute; the existence and uniqueness of such a map makes use of the fact that η_s is initial in $s \downarrow U$. It is easy to verify that these unique choices are functorial, and so we have defined a functor $F\colon \mathsf{S} \to \mathsf{A}$ together with a natural transformation $\eta\colon 1_{\mathsf{S}} \Rightarrow UF$.

This data allows us to define a natural transformation

$$\phi_{s,a}\colon \mathsf{A}(Fs, a) \to \mathsf{S}(s, Ua),$$

as in the proof of Proposition 4.2.4. Given $g\colon Fs \to a$ in A, define $\phi_{s,a}(g)$ to be $s \xrightarrow{\eta_s} UFs \xrightarrow{Ug} Ua$. Injectivity and surjectivity of $\phi_{s,a}$ follows immediately from the uniqueness and existence of morphisms from η_s to any particular $s \to Ua$ in $s \downarrow U$. This natural isomorphism proves that $F \dashv U$ with unit η. □

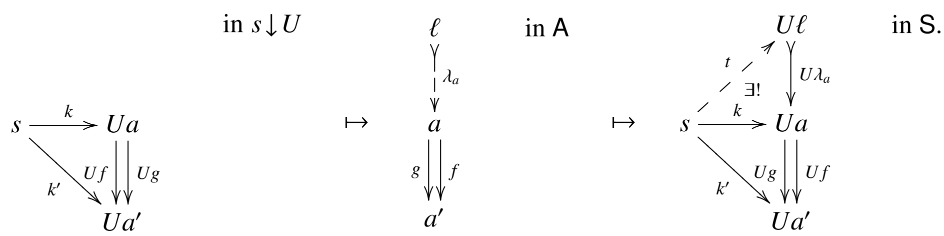
Lemma 4.5.1 reduces the problem of finding a left adjoint to a continuous functor $U\colon \mathsf{A} \to \mathsf{S}$ to the problem of finding an initial object in the comma category $s \downarrow U$ defined for each $s \in \mathsf{S}$. This comma category, as the category of elements for $\mathsf{S}(s, U-)\colon \mathsf{A} \to \mathsf{Set}$, comes with a canonical forgetful functor $\Pi\colon s \downarrow U \to \mathsf{A}$ that carries an object $s \to Ua$ to the object a.

Lemma 4.5.2. *For any functor* $U\colon \mathsf{A} \to \mathsf{S}$ *and object* $s \in \mathsf{S}$*, the associated forgetful functor* $\Pi\colon s \downarrow U \to \mathsf{A}$ *creates the limit of any diagram whose limit exists in* A *and is preserved by* U*. In particular, if* A *is complete and* U *is continuous, then* $s \downarrow U$ *is complete.*

Proof. Consider a diagram $\mathsf{J} \xrightarrow{K} s \downarrow U$ so that $\mathsf{J} \xrightarrow{K} s \downarrow U \xrightarrow{\Pi} \mathsf{A}$ admits a limit cone $\lambda\colon \ell \Rightarrow \Pi K$. By hypothesis, the functor U carries this limit cone to a limit cone $U\lambda\colon U\ell \Rightarrow U\Pi K$ for the diagram $\mathsf{J} \xrightarrow{K} s \downarrow U \xrightarrow{\Pi} \mathsf{A} \xrightarrow{U} \mathsf{S}$. Now the original diagram $\mathsf{J} \xrightarrow{K} s \downarrow U$ can be thought of as a cone over this diagram in S with summit s, which then

[5]The optimization problem intuition for the construction of adjoint functors is explained very well on Wikipedia's adjoint functors entry (retrieved on April 14, 2015). However, their suggestion that "picking the right category [to express the universal property of the adjoint construction] is something of a knack" is incorrect. A left adjoint to a functor $U\colon \mathsf{A} \to \mathsf{S}$ at an object $s \in \mathsf{S}$ defines an initial object in the category of elements of the functor $\mathsf{S}(s, U-)\colon \mathsf{A} \to \mathsf{Set}$; dually, a right adjoint defines a terminal object in the category of elements of $\mathsf{S}(U-, s)\colon \mathsf{A}^{\mathrm{op}} \to \mathsf{Set}$.

factors uniquely along a map $t\colon s \to U\ell$ to the limit. In the case where J is the parallel pair category, this process is illustrated below:



The verification that $t\colon s \to U\ell$ is a limit for the original diagram K in $s \downarrow U$ is straightforward. Note also, that $\Pi\colon s \downarrow U \to \mathsf{A}$ preserves the limit, by construction. □

Lemma 4.5.2 can be used to help find an initial object in $s \downarrow U$. An initial object is equally the *limit* of the identity functor; see Example 3.7.3. Applying Lemma 4.5.2, a limit of the identity functor on $s \downarrow U$ exists if and only if the limit of the forgetful functor $\Pi\colon s \downarrow U \to \mathsf{A}$ exists in A. Following this line of reasoning, it seems that all continuous functors whose domains are complete should admit left adjoints.

The problem is that $s \downarrow U$ is not in general a small category, so even if A admits all small limits it may not admit a limit of the large diagram $s \downarrow U \to \mathsf{A}$. The adjoint functor theorems supply conditions under which this large limit can be reduced to a small one that A possesses. We discuss two of the common variants of this theorem.

Theorem 4.5.3 (General Adjoint Functor Theorem). *Let $U\colon \mathsf{A} \to \mathsf{S}$ be a continuous functor whose domain is locally small and complete. Suppose U satisfies the following* ***solution set condition****:*

- *For every $s \in \mathsf{S}$ there exists a set of morphisms $f_i\colon s \to Ua_i$ so that any $f\colon s \to Ua$ factors through some f_i along a morphism $a_i \to a$ in A.*

Then U admits a left adjoint.

The solution set condition says exactly that the set $f_i\colon s \to Ua_i$ is **jointly weakly initial** in the category $s \downarrow U$, in the sense that any other object in this category admits a map from one of these objects. Because A has all limits and U preserves them, Lemma 4.5.2 implies that the category $s \downarrow U$ is also complete, and a particular limit construction will be used to turn this jointly weakly initial set into an initial object.

Lemma 4.5.4. *If C is complete, locally small, and has a jointly weakly initial set of objects Φ, so that for any $C \in \mathsf{C}$ there is some $K \in \Phi$ and a morphism $K \to C$, then C has an initial object.*

Proof. Let P be the product of the objects in Φ. Then P is **weakly initial**: for each $C \in \mathsf{C}$, there is some morphism $P \to C$ but likely more than one. Let $\ell\colon L \to P$ be the limit cone for the diagram consisting of P and all of its endomorphisms.

Because P is weakly initial, L is too. Suppose there is some $C \in \mathsf{C}$ with two morphisms $f, g\colon L \rightrightarrows C$ and let $e\colon E \to L$ be their equalizer. Because P is weakly initial there is some morphism $p\colon P \to E$. The composite $\ell e p$ is an endomorphism of P so $\ell e p \ell = \ell$. The universal property of the limit L implies that ℓ is monic, and so $ep\ell = 1_L$, and in particular e is split epi. But e is also mono; as a split epi and mono is necessarily an isomorphism, we conclude that $f = g$. □

The proof of the general adjoint functor theorem proceeds directly by assembling the lemmas in this section.

Proof of Theorem 4.5.3. By Lemma 4.5.1, U admits a left adjoint if and only if for each $s \in \mathsf{S}$ the comma category $s \downarrow U$ has an initial object; these initial objects define the value of the left adjoint on objects and the components of the unit of the adjunction. The solution set condition says that $s \downarrow U$ has a jointly weakly initial set of objects. Because A is locally small, $s \downarrow U$ is as well. Because A is complete and U is continuous, Lemma 4.5.2 tells us that $s \downarrow U$ is also complete. So we may apply Lemma 4.5.4 to construct an initial object in $s \downarrow U$. □

Example 4.5.5. For example, consider the functor $U \colon \mathsf{Group} \to \mathsf{Set}$ and a set S. The following construction of the free group on S can be found in [**Lan02**, §I.12]. Let Φ' be the set of representatives for isomorphism classes of groups which can be generated by $|S|$ or fewer elements. Define Φ to be the set of functions $S \to UG$, with $G \in \Phi'$, whose image generates the group G. Form the product, indexed by the set Φ, of the groups appearing as codomains and consider the induced map

$$\eta \colon S \to U(\prod_{S \to UG \in \Phi} G).$$

Let Γ be the subgroup of the product $\prod_{S \to UG \in \Phi} G$ generated by the image of η. We show that Γ is the free group on S by proving that the restriction $\eta \colon S \to U\Gamma$ is initial in $S \downarrow U$.

Given a group H and a function $\psi \colon S \to U(H)$ let $G \subset H$ be the subgroup generated by the image of S. By construction ψ factors through $\phi \colon S \to U(G)$ and $\phi \in \Phi$. Here ϕ is the composite of $\eta \colon S \to U(\Gamma)$ with the projection to the component indexed by ϕ, so we have found a way to factor ψ through η via a homomorphism $\Gamma \to H$. The uniqueness of this factorization is clear because Γ is generated by the image of S. As $\psi \colon S \to U(H)$ specifies the images of these generators no alternate choices are possible.

Our next adjoint functor theorem requires a few preliminary definitions.

Definition 4.5.6. A **generating set** or **generator** for a category C is a set Φ of objects that can distinguish distinct parallel morphisms in the following sense: given $f, g \colon x \rightrightarrows y$, if $f \neq g$ then there exists some $h \colon c \to x$ with $c \in \Phi$ so that $fh \neq gh$. A **cogenerating set** in C is a generating set in C^{op}.

Definition 4.5.7. A **subobject** of an object $c \in \mathsf{C}$ is a monomorphism $c' \rightarrowtail c$ with codomain c. Isomorphic subobjects, that is, subobjects $c' \rightarrowtail c \leftarrowtail c''$ with a commuting isomorphism $c' \cong c''$ are typically identified. The **intersection** of a family of subobjects of c is the limit of the diagram of monomorphisms. The induced map from the limit to c is again a monomorphism. So the intersection is the minimal subobject that is contained in every member of this family.

Theorem 4.5.8 (Special Adjoint Functor Theorem). *Let $U \colon \mathsf{A} \to \mathsf{S}$ be a continuous functor whose domain is complete and whose domain and codomain are locally small. If furthermore A has a small cogenerating set and every collection of subobjects of a fixed object in A admits an intersection, then U admits a left adjoint.*

The purpose for these hypotheses is that they will allow us to construct the initial objects sought for in Lemma 4.5.1.

Lemma 4.5.9. *Suppose C is locally small, complete, has a small cogenerating set Φ, and has the property that every collection of subobjects has an intersection. Then C has an initial object.*

Proof. To say that Φ is cogenerating is to say that for every $C \in \mathsf{C}$ the canonical map

$$C \rightarrowtail \prod_{K \in \Phi} K^{\mathrm{Hom}(C,K)}$$

is a monomorphism. Form the product $P = \prod_{K \in \Phi} K$ of the objects in the cogenerating set and form the intersection $I \hookrightarrow P$ of all of the subobjects of P. We claim that I is initial.

For any $C \in \mathsf{C}$, there is a map $P \to \prod_{K \in \Phi} K^{\mathrm{Hom}(C,K)}$ induced by the maps $K \to K^{\mathrm{Hom}(C,K)}$ defined to be the identity 1_K on each component of the power $K^{\mathrm{Hom}(C,K)}$. The pullback

$$\begin{array}{ccc} P_C & \longrightarrow & C \\ \downarrow & \lrcorner & \downarrow \\ P & \longrightarrow & \prod_{K \in \Phi} K^{\mathrm{Hom}(C,K)} \end{array}$$

defines a subobject P_C of P and thus a map $I \to P_C \to C$ from the intersection to C. There cannot be more than one arrow from I to C because the equalizer of two distinct such would define a smaller subobject of P, contradicting minimality of the intersection I. □

Proof of Theorem 4.5.8. Again, for each $s \in \mathsf{S}$, the category $s \downarrow U$ is locally small and complete. Monomorphisms in $s \downarrow U$ are preserved and reflected by $\Pi\colon s \downarrow U \to \mathsf{A}$, and so the comma categories has intersections of subobjects created by the forgetful functor. If Φ is a cogenerating set for A, then the set

$$\Phi' = \{s \to Ua \mid a \in \Phi\}$$

is cogenerating for $s \downarrow U$; because S is locally small, Φ' is again a set. Applying Lemma 4.5.9, we can construct an initial object in each comma category $s \downarrow U$, which provides a left adjoint to U by Lemma 4.5.1. □

Example 4.5.10. The special adjoint functor theorem is an abstraction of the construction of the Stone-Čech compactification $\beta\colon \mathsf{Top} \to \mathsf{kHaus}$, which defines a left adjoint to the inclusion $\mathsf{kHaus} \hookrightarrow \mathsf{Top}$. The unit interval $I = [0,1]$ is a cogenerating object in kHaus: if $f \neq g\colon X \rightrightarrows Y$, there must be some $x \in X$ with $f(x) \neq g(x)$. Then Urysohn's lemma can be used to define a continuous function $h\colon Y \to I$ with $hf(x) = 0$ and $hg(x) = 1$; hence, $hf \neq hg$. Given a topological space X, Theorem 4.5.8 constructs an initial object in $X \downarrow \mathsf{kHaus}$ in the following manner. A cogenerating family in $X \downarrow \mathsf{kHaus}$ is given by the set of all maps $X \to I$. The product of these maps is a map

$$\hat{\eta}\colon X \to \prod_{\mathrm{Hom}(X,I)} I.$$

A subobject of this object is a compact Hausdorff subspace $K \subset \prod_{\mathrm{Hom}(X,I)} I$ containing the image of $\hat{\eta}$. Because $\prod_{\mathrm{Hom}(X,I)} I$ is compact Hausdorff, a subspace is compact Hausdorff if and only if it is closed. Thus, the intersection of all subobjects of $\hat{\eta}\colon X \to \prod_{\mathrm{Hom}(X,I)} I$ is simply the codomain restriction $\eta\colon X \to \beta(X)$, where $\beta(X)$ is the closure of the image of $\hat{\eta}$. This is the Stone-Čech compactification.

The adjoint functor theorems have a number of useful corollaries.

Corollary 4.5.11. *Suppose* C *is locally small, complete, has a small cogenerating set, and has the property that every collection of subobjects of a fixed object has an intersection. Then* C *is also cocomplete.*

PROOF. For any small category J, as C is locally small, C^{J} is again locally small. The constant diagram functor $\Delta\colon \mathsf{C} \to \mathsf{C}^{\mathsf{J}}$ preserves limits by Exercise 3.5.1. Applying Theorem 4.5.8, Δ has a left adjoint, which by Proposition 4.4.1 demonstrates that C has all J-shaped colimits. □

The adjoint functor theorems also specialize to give conditions under which a set-valued limit-preserving functor is representable.

COROLLARY 4.5.12. *Suppose* C *is locally small, complete, has a small cogenerating set, and has the property that every collection of subobjects of a fixed object has an intersection. Then any continuous functor* $F\colon \mathsf{C} \to \mathsf{Set}$ *is representable.*

PROOF. By Theorem 4.5.8, F has a left adjoint $L\colon \mathsf{Set} \to \mathsf{C}$. In particular, we have a natural isomorphism

$$\mathsf{C}(L(*), c) \cong \mathsf{Set}(*, Fc) \cong Fc$$

where $* \in \mathsf{Set}$ denotes the singleton set. Thus $L(*)$ represents F. □

COROLLARY 4.5.13 (Freyd's representability theorem). *Let* $F\colon \mathsf{C} \to \mathsf{Set}$ *be a continuous functor and suppose that* C *is complete and locally small. If* F *satisfies the solution set condition:*

- *There exists a set* S *of objects of* C *so that for any* $c \in \mathsf{C}$ *and any element* $x \in Fc$ *there exists an* $s \in S$*, an element* $y \in Fs$*, and a morphism* $f\colon s \to c$ *so that* $Ff(y) = x$*.*

then F *is representable.*

PROOF. The solution set defines a jointly weakly initial set of objects in the comma category $* \downarrow F \cong \int F$, where $* \in \mathsf{Set}$ is a singleton set. By Lemma 4.5.2, this category is complete, and so by Lemma 4.5.4 it has an initial object. By Proposition 2.4.4, this defines a representation for F. □

Exercises.

EXERCISE 4.5.1. Suppose C is a locally small category with coproducts. Show that a functor $F\colon \mathsf{C} \to \mathsf{Set}$ is representable if and only if it admits a left adjoint.

CHAPTER 5

Monads and their Algebras

> A monad is just a monoid in the category of endofunctors — what's the problem?
>
> Barbie

Consider an adjunction

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\longleftarrow}{\overset{\longrightarrow}{\perp}}}} \mathsf{D} \qquad\qquad \eta\colon 1_{\mathsf{C}} \Rightarrow GF, \quad \epsilon\colon FG \Rightarrow 1_{\mathsf{D}}.$$

If we restrict this data to the category C, forgetting about D entirely, we obtain a **monad** on C. Dually, if we restrict to D and forget entirely about the category C, we obtain a **comonad** on D. Perhaps unexpectedly from this perspective, monads are interesting in their own right, as syntactic representations of algebraic structure. If a particular variety of algebra is **monadic**, i.e., can be represented as the **algebras for a monad**, there are a number of pleasing consequences. Monoids, groups, rings, modules over a fixed ring, sets with an action of a fixed group, pointed sets, compact Hausdorff spaces, lattices, and so on all arise as algebras for a monad on the category of sets. Fields do not, which explains why the category of fields shares few of the properties of the categories just described.

5.1. Monads from adjunctions

Definition 5.1.1. A **monad** on a category C consists of

- an endofunctor $T\colon \mathsf{C} \to \mathsf{C}$,
- a **unit** natural transformation $\eta\colon 1_{\mathsf{C}} \Rightarrow T$, and
- a **multiplication** natural transformation $\mu\colon T^2 \Rightarrow T$,

so that the following diagrams commute in C^{C}:

$$\begin{array}{ccc} T^3 & \overset{T\mu}{\Longrightarrow} & T^2 \\ {\scriptstyle \mu T}\Downarrow & & \Downarrow{\scriptstyle \mu} \\ T^2 & \underset{\mu}{\Longrightarrow} & T \end{array} \qquad\qquad \begin{array}{ccccc} T & \overset{\eta T}{\Longrightarrow} & T^2 & \overset{T\eta}{\Longleftarrow} & T \\ & {\scriptstyle 1_T}\searrow & \Downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_T} & \\ & & T & & \end{array}$$

Remark 5.1.2. The diagrams in Definition 5.1.1 are reminiscent of the diagrams in Definition 1.6.3. This is no coincidence. Monads, like (topological) monoids, unital rings, and $\Bbbk$-algebras are all instances of **monoids in a monoidal category**. A monoidal category V is a category equipped with a binary functor $\mathsf{V} \times \mathsf{V} \to \mathsf{V}$ and a unit object, together with some additional coherence natural isomorphisms satisfying conditions that we decline to describe. A monad on C is a monoid in the monoidal category C^{C} of endofunctors on C. The binary functor $\mathsf{C}^{\mathsf{C}} \times \mathsf{C}^{\mathsf{C}} \to \mathsf{C}^{\mathsf{C}}$ is composition and the unit object is the identity endofunctor $1_{\mathsf{C}} \in \mathsf{C}^{\mathsf{C}}$. This is one explanation for the name "monad."

Lemma 5.1.3. *Any adjunction*

$$\mathsf{C} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D} \qquad\qquad \eta\colon 1_{\mathsf{C}} \Rightarrow GF, \quad \epsilon\colon FG \Rightarrow 1_{\mathsf{D}}$$

gives rise to a monad on the category C *serving as the domain of the left adjoint, with*

- *the endofunctor* T *defined to be* GF,
- *the unit* $\eta\colon 1_{\mathsf{C}} \Rightarrow GF$ *serving as the unit* $\eta\colon 1_{\mathsf{C}} \Rightarrow T$ *for the monad, and*
- $G\epsilon F\colon GFGF \Rightarrow GF$ *serving as the multiplication* $\mu\colon T^2 \Rightarrow T$.

Proof. The triangles

$$\begin{array}{ccc} GFGFGF & \xRightarrow{GFG\epsilon F} & GFGF \\ {\scriptstyle G\epsilon FGF}\Downarrow & & \Downarrow{\scriptstyle G\epsilon F} \\ GFGF & \xRightarrow[G\epsilon F]{} & GF \end{array} \qquad\qquad \begin{array}{ccccc} GF & \xRightarrow{\eta GF} & GFGF & \xLeftarrow{GF\eta} & GF \\ & {\scriptstyle 1_{GF}}\searrow & \Downarrow{\scriptstyle G\epsilon F} & \swarrow{\scriptstyle 1_{GF}} & \\ & & GF & & \end{array}$$

commute by the triangle identities for the adjunction. The square commutes by naturality of the vertical natural transformation $G\epsilon\colon GFG \Rightarrow G$. □

Examples 5.1.4. For instance:

(i) The free forgetful adjunction between pointed sets and ordinary sets induces a monad on Set whose endofunctor $(-)_+\colon \mathsf{Set} \to \mathsf{Set}$ adds a new point. The components of the unit are given by the obvious inclusions $\eta_A\colon A \to A_+$. The components of the multiplication $\mu_A\colon (A_+)_+ \to A_+$ are defined to be the identity on the subset A and send the two new points in $(A_+)_+$ to the new point in A_+. By Lemma 5.1.3, or by a direct verification, the diagrams

$$\begin{array}{ccc} ((A_+)_+)_+ & \xrightarrow{(\mu_A)_+} & (A_+)_+ \\ {\scriptstyle \mu_{A_+}}\downarrow & & \downarrow{\scriptstyle \mu_A} \\ (A_+)_+ & \xrightarrow[\mu_A]{} & A_+ \end{array} \qquad\qquad \begin{array}{ccccc} A_+ & \xrightarrow{\eta_{A_+}} & (A_+)_+ & \xleftarrow{(\eta_A)_+} & A_+ \\ & {\scriptstyle 1_{A_+}}\searrow & \downarrow{\scriptstyle \mu_A} & \swarrow{\scriptstyle 1_{A_+}} & \\ & & A_+ & & \end{array}$$

commute. Particularly in computer science, this monad is called the **maybe monad**, for reasons that we will explore later. There is a similar monad on Top, or any category with coproducts, which acts by adjoining a copy of a fixed object (in this case a point).

(ii) The **free monoid monad** is induced by the free-forgetful adjunction between monoids and sets. The endofunctor $T\colon \mathsf{Set} \to \mathsf{Set}$ is defined by

$$TA := \coprod_{n\geq 0} A^n,$$

that is, TA is the set of finite lists of elements in A; in computer science contexts, this monad is often called the **list monad**. The components of the unit $\eta_A\colon A \to TA$ are defined by the evident coproduct inclusions. The components of the multiplication $\mu_A\colon T^2A \to TA$ are the concatenation functions, sending a list of lists to the composite list. See Exercise 5.1.1 for a more categorical description that demonstrates that the free monoid monad T can also be defined in any category with coproducts and a well-behaved binary functor such as $\times$.

(iii) The free-forgetful adjunction between sets and the category of R-modules induces the **free R-module monad** $R[-]\colon \mathsf{Set} \to \mathsf{Set}$. Define $R[A]$ to be the set of finite formal R-linear combinations of elements of A. Formally, a finite R-linear combination is a function $\chi\colon A \to R$, for which only finitely many of the values are non-zero; it might be written as $\coprod_{a \in A} \chi(a) \cdot a$. The components $\eta_A\colon A \to R[A]$ of the unit send an element $a \in A$ to the singleton formal R-linear combination corresponding to the function $\chi_a\colon A \to R$ that sends a to $1 \in R$ and every other element to zero. The components $\mu_A\colon R[R[A]] \to R[A]$ are defined by distributing the coefficients in a formal sum of formal sums. Special cases of interest include the **free abelian group monad** and the **free vector space monad**.

(iv) The free-forgetful adjunction between sets and groups induces the **free group monad** $F\colon \mathsf{Set} \to \mathsf{Set}$ that sends a set A to the set $F(A)$ of finite words in the letters $a \in A$ together with formal inverses a^{-1}.

(v) The composite adjunction

$$\mathsf{kHaus} \underset{\longrightarrow}{\overset{\beta}{\underset{\perp}{\longleftarrow}}} \mathsf{Top} \underset{\underset{U}{\longrightarrow}}{\overset{D}{\underset{\perp}{\longleftarrow}}} \mathsf{Set}$$

induces a monad $\beta\colon \mathsf{Set} \to \mathsf{Set}$ that sends a set to the underlying set of the Stone-Čech compactification of the discrete space on that set. There is a simpler description: $\beta(A)$ is the set of **ultrafilters** on A; an ultrafilter is a set of subsets of A that is upward closed, closed under finite intersections, and for any subset of A contains exactly one of that subset or its complement. See Exercise 5.1.2.

(vi) The contravariant powerset functor is its own mutual right adjoint:

$$\mathsf{Set} \underset{\underset{P}{\longleftarrow}}{\overset{P}{\underset{\perp}{\longrightarrow}}} \mathsf{Set}^{\mathrm{op}} \qquad\qquad \mathsf{Set}(A, P(B)) \cong \mathsf{Set}(B, P(A))$$

A function from A to the power-set of B, or equally a function from B to the powerset of A, can be encoded as a function $A \times B \to \Omega$, i.e., as a relation on $A \times B$. This monad takes a set A to $P(P(A))$. The components of the unit are the functions $\eta_A\colon A \to P(P(A))$ that send an element a to the set of subsets of A that contain a. The components of the multiplication take a set of sets of sets of subsets to the set of subsets of A with the property that one of the sets of sets of subsets is the set of all sets of subsets of A that include that particular subset as an element.[1] There is a similar **double dual monad** on $\mathsf{Vect}_{\Bbbk}$.

EXAMPLES 5.1.5. Monads also arise in nature:

(i) The covariant power-set functor $P\colon \mathsf{Set} \to \mathsf{Set}$ is also a monad. The unit $A \to P(A)$ sends an element to the singleton subset. The multiplication $P(P(A)) \to P(A)$ takes the union of a set of subsets. Naturality of the unit, with respect to a function $f\colon A \to B$, makes use of the fact that $f_*\colon P(A) \to P(B)$ is the direct image (rather than inverse image) function. Naturality of the multiplication maps makes use of Corollary 4.4.4; the direct image function, as a left adjoint,

[1]This is one of those instances where it is easier to speak mathematics than to speak English; the multiplication is the inverse image function associated to $\eta_{PA}\colon PA \to P(P(P(A)))$.

preserves unions.

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & PA \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f_*} \\ B & \xrightarrow[\eta_B]{} & PB \end{array} \qquad\qquad \begin{array}{ccc} P^2(A) & \xrightarrow{\mu_A} & PA \\ {\scriptstyle (f_*)_*}\downarrow & & \downarrow{\scriptstyle f_*} \\ P^2(B) & \xrightarrow[\mu_B]{} & P(B) \end{array}$$

(ii) A modification of the free monoid monad yields the **free commutative monoid monad** defined by

$$TA := \coprod_{n \geq 0} (A^{\times n})_{/\Sigma_n}.$$

Elements are finite unordered lists of elements, or equally functions $A \to \mathbb{N}$ with finite support (only finitely many non-zero values).

(iii) There is a monad $- \times \mathbb{N} \colon \mathsf{Set} \to \mathsf{Set}$. We might think of the second component of an element $(a, n) \in A \times \mathbb{N}$ as a discrete time variable. The unit $A \to A \times \mathbb{N}$ is defined by $a \mapsto (a, 0)$ and the multiplication $A \times \mathbb{N} \times \mathbb{N} \to A \times \mathbb{N}$ is defined by $(a, m, n) \mapsto (a, m + n)$. A similar monad exists for any monoid M in place of $\mathbb{N}$, and this example can be generalized further from Set to any monoidal category.

A monad on C^{op} defines a **comonad** on C, consisting of an endofunctor $K \colon \mathsf{C} \to \mathsf{C}$ together with natural transformations $\epsilon \colon K \Rightarrow 1_{\mathsf{C}}$ and $\delta \colon K \Rightarrow K^2$ so that the diagrams dual to Definition 5.1.1 commute in C^{C}. A comonad is a comonoid in the category of endofunctors of C. By the dual to Lemma 5.1.3, any adjunction induces a comonad on the domain of its right adjoint.

EXAMPLE 5.1.6. A monad on a preorder P is given by an order-preserving function $T \colon \mathsf{P} \to \mathsf{P}$ that is so that $p \leq Tp$ and $T^2 p \leq Tp$. If P is a poset, so that isomorphic objects are equal, these two conditions imply that $T^2 p = Tp$. An order-preserving function T so that $p \leq Tp$ and $T^2 p = Tp$ is called a **closure operator**. Dually, a comonad on a poset category P defines a **kernel operator**: an order-preserving function K so that $Kp \leq p$ and $Kp = K^2 p$.

For example, the poset PX of subsets of a topological space X admits a closure operator $TA = \bar{A}$, where $\bar{A}$ is the closure of $A \subset X$, and a kernel operator $KA = A^o$, where A^o is the interior of $A \subset X$.[2]

Exercises.

EXERCISE 5.1.1. Suppose V is a category with a bifunctor $\otimes \colon \mathsf{V} \times \mathsf{V} \to \mathsf{V}$ that is associative up to coherent natural isomorphism.[3] Suppose also that V has countable coproducts and that the bifunctor $\otimes$ preserves them in each variable.[4] Show that $T(X) = \coprod_{n \geq 0} X^{\otimes n}$ defines a monad on V by defining natural transformations $\eta \colon 1_{\mathsf{V}} \Rightarrow T$ and $\mu \colon T^2 \Rightarrow T$ that satisfy the required conditions.

EXERCISE 5.1.2. Show that the functor $\beta \colon \mathsf{Set} \to \mathsf{Set}$ that carries a set to the set of ultrafilters on that set is a monad by defining unit and multiplication natural transformations that satisfy the unit and associativity laws.

[2] The closure of A is the smallest closed set containing A, equally the intersection of all closed sets containing A; the interior is defined dually.

[3] Rather than worry about what this means, feel free to assume that there is a well-defined n-ary functor $\mathsf{V}^{\times n} \xrightarrow{\otimes^n} \mathsf{V}$ built from the bifunctor $\otimes$. Such a functor exists in any monoidal category.

[4] In particular $(v \coprod v') \otimes (w \coprod w') \cong v \otimes w \coprod v' \otimes w \coprod v \otimes w' \coprod v' \otimes w'$.

Exercise 5.1.3. The adjunction associated to a reflective subcategory of C induces an **idempotent monad** on C. Prove that the following three characterizations of an idempotent monad (T, η, μ) are equivalent:

(i) The multiplication $\mu \colon T^2 \Rightarrow T$ is a natural isomorphism.
(ii) The natural transformations $\eta T, T\eta \colon T \Rightarrow T^2$ are equal.
(iii) Each component of $\mu \colon T^2 \Rightarrow T$ is a monomorphism.

5.2. Adjunctions from monads

We have seen, in Lemma 5.1.3, that any adjunction gives rise to a monad on the category serving as the domain of its left adjoint. It's natural to ask whether all monads arise this way. For instance, is there any adjunction that gives rise to the power-set monad on Set? Perhaps surprisingly, the answer is yes: any monad T in C arises as the shadow cast by two (usually distinct) adjunctions, that are moreover universal with this property. The initial such adjunction, in a category to be introduced shortly, is between C and the **Kleisli category** of T. The terminal such adjunction is between C and the **Eilenberg-Moore category** of T, also called the category of ***T*-algebras**. The definitions of the Kleisli and Eilenberg-Moore categories are elementary[5] and could be given immediately, but we prefer to start by giving an example so that when they appear, these definitions won't look so strange.

Recall that an affine space, over a fixed field $\Bbbk$, is like a "vector space that has forgotten its origin." It's possible to translate a point in n-dimensional affine space A by a n-dimensional vector, in which case one obtains a unique point in the affine space. Conversely, any two points in affine space can be subtracted to obtain a unique vector. Choosing an origin $\mathbf{o} \in A$ defines a bijection between points $\mathbf{a}$ in the affine space and vectors $\vec{\mathbf{v}} = \mathbf{a} - \mathbf{o}$. In summary:

Definition 5.2.1. Given a vector space V over a field $\Bbbk$, an **affine space** is a non-empty set A together with a "translation" function $V \times A \xrightarrow{+} A$ so that

- $\vec{0} + \mathbf{a} = \mathbf{a}$ for all $\mathbf{a} \in A$.
- $(\vec{\mathbf{v}} + \vec{\mathbf{w}}) + \mathbf{a} = \vec{\mathbf{v}} + (\vec{\mathbf{w}} + \mathbf{a})$ for all $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in V$ and $\mathbf{a} \in A$.
- for any $\mathbf{a} \in A$, the function $- + \mathbf{a} \colon V \to A$ is a bijection.

There is another simpler definition that allows us to define an affine space without making use of the auxiliary vector space V. If we temporarily fix an origin $\mathbf{o} \in A$, then for any scalar $\lambda \in \Bbbk$ and two elements $\mathbf{a}, \mathbf{b} \in A$ of the affine space, we can exploit the bijection $- + \mathbf{o} \colon V \to A$ to see that there is a unique $\mathbf{c} \in A$ so that

$$\mathbf{c} - \mathbf{o} = \lambda(\mathbf{a} - \mathbf{o}) + (1 - \lambda)(\mathbf{b} - \mathbf{o}).$$

This element $\mathbf{c}$ is sensibly denoted by $\lambda\mathbf{a} + (1 - \lambda)\mathbf{b}$. This element $\lambda\mathbf{a} + (1 - \lambda)\mathbf{b} \in A$ is independent of the choice of origin. More generally, for any n-tuple $\mathbf{a}_1, \ldots, \mathbf{a}_n \in A$ and scalars $\lambda_1, \ldots, \lambda_n \in \Bbbk$ with $\lambda_1 + \cdots + \lambda_n = 1$, there is a unique element $\lambda_1\mathbf{a}_1 + \cdots + \lambda_n\mathbf{a}_n \in A$ defined analogously as an **affine linear combination** of $\mathbf{a}_i$. This leads to a second equivalent definition of affine space.

Definition 5.2.2. An **affine space** is a non-empty set A in which affine linear combinations can be evaluated.

[5] This is not to say that the constructions demonstrating the existence of adjunctions giving rise to a generic monad aren't rather clever.

More precisely, define $\mathrm{Aff}_{\Bbbk}(A)$ to be the set of finite formal sums[6] $\sum_{i=1}^{n} \lambda_i \mathbf{a}_i$ so that $\sum_{i=1}^{n} \lambda_i = 1$; a sum is **affine** precisely when the coefficients sum to $1 \in \Bbbk$. To say "these can be evaluated" means there is a function $\mathrm{ev}_A \colon \mathrm{Aff}_{\Bbbk}(A) \to A$. For this evaluation to define a reasonable affine linear combination function, we need a few axioms, namely:

- If $\eta_A \colon A \to \mathrm{Aff}_{\Bbbk}(A)$ is the "singleton" function and $\mu_A \colon \mathrm{Aff}_{\Bbbk}(\mathrm{Aff}_{\Bbbk}(A)) \to \mathrm{Aff}_{\Bbbk}(A)$ is the "distributivity" function, then the following diagrams

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & \mathrm{Aff}_{\Bbbk}(A) \\ & \searrow^{1_A} & \downarrow \mathrm{ev}_A \\ & & A \end{array} \qquad \begin{array}{ccc} \mathrm{Aff}_{\Bbbk}(\mathrm{Aff}_{\Bbbk}(A)) & \xrightarrow{\mu_A} & \mathrm{Aff}_{\Bbbk}(A) \\ {\scriptstyle \mathrm{Aff}_{\Bbbk}(\mathrm{ev}_A)} \downarrow & & \downarrow \mathrm{ev}_A \\ \mathrm{Aff}_{\Bbbk}(A) & \xrightarrow[\mathrm{ev}_A]{} & A \end{array}$$

  commute in Set.

The first condition says that the value of a singleton sum $1 \cdot \mathbf{a}$ is the element $\mathbf{a}$. The second condition says that an affine linear combination of affine linear combinations

$$\lambda_1 \cdot (\mu_{11}\mathbf{a}_{11} + \cdots + \mu_{1n_1}\mathbf{a}_{1n_1}) + \cdots + \lambda_k \cdot (\mu_{k1}\mathbf{a}_{k1} + \cdots + \mu_{kn_k}\mathbf{a}_{kn_k})$$

can be evaluated by first distributing and then evaluation — note that $\sum_{ij} \lambda_i \mu_{ij} = 1$ — or by first evaluating inside each of the k sets of parentheses and then evaluating the resulting affine linear combination. In summary, an affine space is an **algebra** for the affine linear combination monad $\mathrm{Aff}_{\Bbbk}(-) \colon \mathsf{Set} \to \mathsf{Set}$.

Let us now introduce the general definition.

Definition 5.2.3. Let C be a category with a monad (T, η, μ). The **Eilenberg-Moore category** for T or the **category of T-algebras** is given by:

- objects are pairs $(A \in \mathsf{C}, a \colon TA \to A)$, so that the diagrams

(5.2.4)
$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & TA \\ & \searrow^{1_A} & \downarrow a \\ & & A \end{array} \qquad \begin{array}{ccc} T^2A & \xrightarrow{\mu_A} & TA \\ {\scriptstyle Ta} \downarrow & & \downarrow a \\ TA & \xrightarrow[a]{} & A \end{array}$$

  commute in C and
- morphisms $f \colon (A, a) \to (B, b)$ are given by $f \colon A \to B$ in C so that the square

$$\begin{array}{ccc} TA & \xrightarrow{Tf} & TB \\ {\scriptstyle a} \downarrow & & \downarrow b \\ A & \xrightarrow[f]{} & B \end{array}$$

  commutes.

Various notations are common for the Eilenberg-Moore category, among them C^T, $\mathsf{C}[T]$, or Alg_T (with varying typography). Here we'll write C^T for conciseness.

Lemma 5.2.5. *For any category C with a monad (T, η, μ), there is an adjunction*

$$\mathsf{C} \underset{U^T}{\overset{F^T}{\rightleftarrows}} \mathsf{C}^T \quad (F^T \dashv U^T)$$

[6] Two finite sums are considered to be identical when they differ only up to reordering of terms, up to consolitating repeated instances of the same term by adding their coefficients, or up to inclusion or deletion of terms whose coefficients are zero.

so that the induced monad is (T, η, μ)*.*

Proof. The functor $F^T : \mathsf{C} \to \mathsf{C}^T$ carries an object $A \in \mathsf{C}$ to the **free T-algebra** $(TA, \mu_A : T^2A \to TA)$ and carries a morphism $f : A \to B$ to the free T-algebra morphism $Tf : (TA, \mu_A) \to (TB, \mu_B)$. The functor $U^T : \mathsf{C}^T \to \mathsf{C}$ is the evident forgetful functor. Note that $U^T F^T = T$. The unit of the adjunction $F^T \dashv U^T$ is given by the natural transformation $\eta : 1_{\mathsf{C}} \Rightarrow T$. The components of the counit $\epsilon : F^T U^T \Rightarrow 1_{\mathsf{C}^T}$ are defined as follows:

$$\epsilon_{(A,a)} := (TA, \mu_A) \xrightarrow{a} (A, a) \qquad \begin{array}{ccc} T^2A & \xrightarrow{Ta} & TA \\ {\scriptstyle \mu_A}\downarrow & & \downarrow{\scriptstyle a} \\ TA & \xrightarrow[a]{} & A \end{array}$$

That is, the component at an algebra (A, a) is given by the algebra structure map $a : TA \to A$; the square displayed above-right demonstrates that this map defines a morphism of algebras $a : (TA, \mu_A) \to (A, a)$. Note, in particular, that $U^T \epsilon F^T_A = \mu_A$, so that the monad underlying the adjunction $F^T \dashv U^T$ is (T, η, μ).

The triangle identities can be demonstrated by straightforward verifications left to the reader. □

Note that algebras for the affine space monad $\mathrm{Aff}_{\Bbbk}$ on Set are precisely affine spaces in the sense of Definition 5.2.2. This is a representative example: the abstract definition of 5.2.3, which is a priori a bit strange, precisely captures familiar notions of “algebra,” of the variety encoded by the monad (T, η, μ).

Examples 5.2.6. For instance:

(i) Consider the free pointed set monad of Example 5.1.4.(i). An algebra is a set A with a map $a : A_+ \to A$ so that the diagrams (5.2.4) commute. The square imposes no additional conditions, but the triangle says that the map $a : A_+ \to A$ restricts to the identity on the A component of the disjoint union $A_+ = A \sqcup \{*\}$. Thus, the data of an algebra is a set with a specified basepoint $a \in A$, the image of the extra point $*$ under the map $a : A_+ \to A$. A morphism $f : (A, a) \to (B, b)$ is a map $f : A \to B$ so that

$$\begin{array}{ccc} A_+ & \xrightarrow{f_+} & B_+ \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle b} \\ A & \xrightarrow[f]{} & B \end{array}$$

commutes. The map f_+ carries the extra point in A_+ to the extra point in B_+. Thus, this condition says exactly that $f(a) = b$. In this way we see that the Eilenberg-Moore category is isomorphic to Set_*, the category of pointed sets.

(ii) An algebra for the free monoid monad is a set A with a map $\alpha : \coprod_{n \geq 0} A^n \to A$, which simultaneously specifies an n-ary operation on A for each n, satisfying certain conditions. The unit triangle says that the unary operation defined by α

is the identity. The commutative square

$$\begin{array}{ccc} \coprod_{n\geq 0}(\coprod_{m\geq 0} A^m)^n & \xrightarrow{\mu_A} & \coprod_{k\geq 0} A^k \\ {\scriptstyle \coprod_{n\geq 0}\alpha^n}\downarrow & & \downarrow{\scriptstyle \alpha} \\ \coprod_{n\geq 0} A^n & \xrightarrow[\alpha]{} & A \end{array} \tag{5.2.7}$$

imposes an associativity condition on the operations. It says that given a list

$$((a_{11}, \ldots, a_{1m_1}), \ldots, (a_{n1}, \ldots, a_{nm_n}))$$

of n lists of elements of A of varying length, then the result of applying the $(m_1 + \cdots + m_n)$-ary operation to the concatenated list is the same as applying the n-ary operation to the results of applying the m_i-ary operations to each sublist.

Recall, a **monoid** is a set A with an associative binary operation $A^2 \to A$ and a unit element $e \in A$. It's clear that a monoid defines an algebra for the free monoid monad: the nullary operation $1 \to A$ picks out e, the unary operation is necessarily the identity, the binary operation is given, and the operations of higher arity are determined by iterating the binary operation; associativity implies that this is well-defined.

Conversely, any algebra defines a monoid, whose unit and binary operations are defined to be the 0th and 2nd components of the map $\alpha\colon \coprod_{n\geq 0} A^n \to A$. The associativity condition (5.2.7) implies in particular that the ternary product of three elements $a_1, a_2, a_3 \in A$ is equal to both of the iterated binary products $(a_1 \cdot a_2) \cdot a_3$ and $a_1 \cdot (a_2 \cdot a_3)$. Similarly, it implies that the composite of the binary operation with the nullary operation in one of its inputs is equal to the identity unary operation. Thus, we conclude that the category of algebras for the free monoid monad is isomorphic to the category Mon of associative unital monoids.

(iii) An algebra for the closure closure operator on the poset of subsets of a topological space X is exactly a closed subset of X. Dually, a coalgebra[7] for the interior kernel operator is exactly an open subset.

(iv) Consider a reflective subcategory $\mathsf{D} \hookrightarrow \mathsf{C}$ with reflector R. The induced endofunctor $R\colon \mathsf{C} \to \mathsf{C}$ defines a monad on C with unit $\eta_C\colon C \to RC$ and multiplication a natural isomorphism $R^2C \cong RC$; see Exercise 4.4.4. An algebra is an object $C \in \mathsf{C}$ together with a map $c\colon RC \to C$ that is a retraction of the unit component η_C. In fact, η_C and c are inverse isomorphisms. By naturality of η, $\eta_C \cdot c = Rc \cdot \eta_{RC}$, but $\eta_{RC} = R\eta_C$, as both maps are left inverse to the isomorphism μ_C, and so $Rc \cdot \eta_{RC} = Rc \cdot R\eta_C = 1_{RC}$. An easy diagram chase shows that the multiplication condition on $c = \eta_C^{-1}$ is automatic. We conclude that the map $c\colon RC \to C$ provides no additional data; its existence is instead a condition on C. So the category of R-algebras is isomorphic to the essential image of $\mathsf{D} \hookrightarrow \mathsf{C}$. See Exercises 4.4.4 and 5.1.3.

A second solution to the problem of finding an adjunction that induces a particular monad on a category is given by the Kleisli category construction.

Definition 5.2.8. Let C be a category with a monad (T, η, μ). The **Kleisli category** C_T has the same objects as C but a morphism from A to B in C_T is a morphism $A \to TB$ in C; we

[7]Interpreting Definition 5.2.3 for a monad (T, η, μ) on C^{op} defines the category of **coalgebras** for the comonad on C.

might write $A \rightsquigarrow B$ for morphisms in the Kleisli category to distinguish them. The monad structure is used to define identities and composition, so that the Kleisli category C_T is a category. The unit $\eta_A \colon A \to TA$ defines the identity morphism at $A \in \mathsf{C}_T$. The composite of a morphism $f \colon A \to TB$ from A to B with a morphism $g \colon B \to TC$ from B to C is defined to be

$$A \xrightarrow{f} TB \xrightarrow{Tg} T^2C \xrightarrow{\mu_C} TC$$

It is straightforward to verify that these operations are associative and unital.

Lemma 5.2.9. *For any category* C *with a monad* (T, η, μ)*, there is an adjunction*

$$\mathsf{C} \underset{U_T}{\overset{F_T}{\underset{\perp}{\rightleftarrows}}} \mathsf{C}_T$$

so that the induced monad is (T, η, μ)*.*

Proof. The functor F_T is the identity on objects and carries a morphism $f \colon A \to B$ to the morphism from A to B in C_T defined by $A \xrightarrow{f} B \xrightarrow{\eta_B} TB$. The functor U_T sends an object $A \in \mathsf{C}_T$ to TA and sends a morphism $g \colon A \to TB$ from A to B in C_T to $TA \xrightarrow{Tg} T^2B \xrightarrow{\mu_B} TB$. We leave it to the reader to see that both assignments are functorial. Note in particular that $U_TF_T = T$. From the definition of the hom-sets in C_T, we have isomorphisms

$$\mathsf{C}_T(F_TA, B) \cong \mathsf{C}(A, U_TB),$$

which can be seen to be natural in both variables. This demonstrates the adjunction $F_T \dashv U_T$. The remaining details are left as an exercise. □

Examples 5.2.10. For instance:

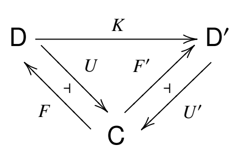
(i) Objects in the Kleisli category for the maybe monad on Set are sets. A map from A to B is a function $A \to B_+$, which may be thought of as a partially defined function from A to B: the elements of A that are sent to the free basepoint have "undefined" output. The composite of two partial functions is the maximal partially defined function. Thus, the Kleisli category is the category of sets and partially-defined functions.

(ii) Objects in the Kleisli category for the free monoid monad on Set are sets and a map from A to B is a function $A \to \coprod_{n\geq 0} B^n$. For each input element $a \in A$, the output is an element $(b_1, \ldots, b_k) \in \coprod_{n\geq 0} B^n$, i.e., a list of elements of B.

For any monad (T, η, μ) on C there is a category Adj_T whose objects are fully-specified adjunctions

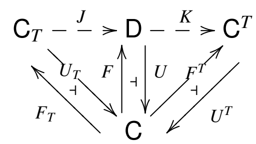
$$\mathsf{C} \underset{U}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathsf{D} \qquad \eta \colon 1_{\mathsf{C}} \Rightarrow UF, \quad \epsilon \colon FU \Rightarrow 1_{\mathsf{D}}$$

inducing the monad (T, η, μ) on C. A morphism



is a functor $K \colon \mathsf{D} \to \mathsf{D}'$ commuting with both the left and right adjoints, i.e., so that $KF = F'$ and $U'K = U$. These conditions imply further that the whiskered composites of K with the counits of $F \dashv U$ and $F' \dashv U'$ coincide, and that K carries the transpose in D of a morphism $c \to Ud = U'Kd$ to the transpose of this morphism in D'; see Exercise 4.2.4.

PROPOSITION 5.2.11. *The Kleisli category* C_T *is initial in* Adj_T *and the Eilenberg-Moore category* C^T *is terminal. That is, for any adjunction* $F \dashv U$ *inducing the monad* T *on* C*, there exist unique functors*



commuting with the left and right adjoints. Moreover, J *and* K *commute with the counits of the adjunctions.*

PROOF. On objects $c \in \mathsf{C}_T$, we are forced to define $Jc = Fc$ so that $JF_T = F$. A morphism f from c to c' in C_T is the transpose, under $F_t \dashv U_T$, of the representing morphism $f\colon c \to Tc' = U_Tc'$ in C, so the fact that J commutes with transposition forces us to define Jf to be the composite $Fc \xrightarrow{Ff} FTc' = FUFc' \xrightarrow{\epsilon F_{c'}} Fc'$, i.e., to be the transpose of f under $F \dashv U$, where ϵ is the counit. The proof that these definitions are functorial makes use of the fact that $U\epsilon F = \mu$, the multiplication for the monad.

To define the functor K on an object $d \in \mathsf{D}$, we must find a suitable T-algebra structure for the object $Ud \in \mathsf{C}$. For any algebra $(c, \gamma\colon Tc \to c) \in \mathsf{C}^T$, the algebra structure map γ can be recognized as the morphism $\gamma\colon (Tc, \mu_c) \to (c, \gamma)$ from the free T-algebra on c to the given T-algebra, that is the transpose of the identity on $c = U^T(c, \gamma)$. The fact that K preserves transposes now tells us that we should define $Kd = (Ud, U\epsilon_d)$; the proof that $U\epsilon_d\colon UFUd \to Ud$ is an algebra structure map makes use of the fact that $U\epsilon F = \mu$. On morphisms, the condition $U^TK = U$ forces us to define the image of $f\colon d \to d'$ to be $Uf\colon (Ud, U\epsilon_d) \to (Ud', U\epsilon_{d'})$, and indeed this is a morphism of algebras. Functoriality in this case is obvious. □

Proposition 5.2.11 implies in particular that there is a unique functor from the Kleisli category for any monad to the Eilenberg-Moore category that commutes with the free and forgetful functors from and to the underlying category. The following result characterizes its image.

LEMMA 5.2.12. *Let* (T, η, μ) *be a monad on* C*. The canonical functor* $K\colon \mathsf{C}_T \to \mathsf{C}^T$ *from the Kleisli category to the Eilenberg-Moore category is full and faithful and its image consists of the free* T*-algebras.*

Recall, a **free** T**-algebra** is an object $F^Tc = (Tc, \mu_c)$ that is in the image of the free functor $F^T\colon \mathsf{C} \to \mathsf{C}^T$.

PROOF. The proof of Proposition 5.2.11 supplies the definition of the functor: $Kc := (Tc, \mu_c)$, Tc being the object U_Tc and μ_c being U_T of the component of the counit of the Kleisli adjunction at the object $c \in \mathsf{C}_T$.

Applying Exercise 4.2.4, for each pair $c, c' \in \mathsf{C}_T$, the action of the functor K on the hom-set from c to c'

$$\mathsf{C}_T(c, c') \xrightarrow{K} \mathsf{C}^T(Kc, Kc') = \mathsf{C}^T((Tc, \mu_c), (Tc', \mu_{c'}))$$

commutes with the transposition natural isomorphisms from each hom-set to $\mathsf{C}(c, Tc')$. In particular, this map must also be an isomorphism, demonstrating that the functor $\mathsf{C}_T \to \mathsf{C}^T$ is full and faithful. □

The upshot of Lemma 5.2.12 is that the Kleisli category for a monad embeds as the full subcategory of free T-algebras and all maps between such. Lemma 5.2.12 also tells us precisely when the Kleisli and Eilenberg-Moore categories are equivalent: this is the case when all algebras are free.

Example 5.2.13. Example 1.5.4 demonstrates that the canonical comparison from the Kleisli to the Eilenberg-Moore categories associated to the maybe monad on Set are equivalent

Exercises.

Exercise 5.2.1. Fill in the remaining details in the proof of Lemma 5.2.5 to show that the free and forgetful functors relating a category C with a monad T to the category C^T of T-algebras are adjoints, inducing the given monad T.

Exercise 5.2.2. Fill in the remaining details in the proof of Proposition 5.2.11: verify the functoriality of J and K and show that the whiskered composites of the counits of these three adjunctions with K and J agree.

Exercise 5.2.3. Verify that the Kleisli category is a category by checking that the composition operation of Definition 5.2.8 is associative and unital.

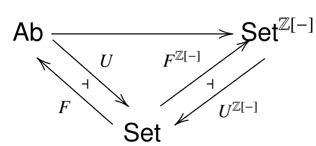
Exercise 5.2.4. Fill in the remaining details in the proof of Lemma 5.2.9 to show that the functors F_T and U_T relating a category C with a monad T to the Kleisli category C_T are adjoints, inducing the given monad T.

5.3. Free algebras and canonical presentations

The free-forgetful adjunction

$$\mathsf{Set} \underset{U}{\overset{F}{\rightleftarrows}} \mathsf{Ab} \quad (F \dashv U)$$

between sets and abelian groups induces a monad $\mathbb{Z}[-]$ on Set. Given a set S, $\mathbb{Z}[S]$ is defined to be the set of finite $\mathbb{Z}$-linear combinations of elements of S, and given a function $f\colon S \to T$, the map $\mathbb{Z}[f]\colon \mathbb{Z}[S] \to \mathbb{Z}[T]$ carries a formal sum $n_1 s_1 + \cdots + n_k s_k$, with $n_i \in \mathbb{Z}$ and $s_i \in S$, to the formal sum $n_1 f(s_1) + \cdots + n_k f(s_k)$. By Proposition 5.2.11, there is a unique functor



from the category of abelian groups to the category of algebras that commutes with the left and right adjoints. The image of an abelian group A is the algebra consisting of the underlying set of A together with the "evaluation" map $\epsilon_A\colon \mathbb{Z}[A] \to A$ that interprets a finite formal sum as a single element of A. This evaluation map is the component of the counit of the adjunction $F \dashv U$ at $A \in \mathsf{Ab}$, from which point it follows easily that the diagrams (5.2.4) commute.

In fact, this functor defines an isomorphism of categories. A $\mathbb{Z}[-]$-algebra is a set A with an "evaluation" function $\alpha\colon \mathbb{Z}[A] \to A$, so that diagrams

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & \mathbb{Z}[A] \\ & {\scriptstyle 1_A}\searrow & \downarrow{\scriptstyle \alpha} \\ & & A \end{array} \qquad\qquad \begin{array}{ccc} \mathbb{Z}[\mathbb{Z}[A]] & \xrightarrow{\mu_A} & \mathbb{Z}[A] \\ {\scriptstyle \mathbb{Z}[\alpha]}\downarrow & & \downarrow{\scriptstyle \alpha} \\ \mathbb{Z}[A] & \xrightarrow[\alpha]{} & A \end{array}$$

commute. This tells us that the evaluation function sends a singleton sum $a \in \mathbb{Z}[A]$ to the element a and that the evaluation map is associative: in particular, the values assigned to the sums of sums $(a_1 + a_2) + (a_3)$ and $(a_1) + (a_2 + a_3)$ are equal. In this way, we see that the map $\alpha \colon \mathbb{Z}[A] \to A$ gives the set A the structure of an abelian group. Moreover, an algebra homomorphism

$$\begin{array}{ccc} \mathbb{Z}[A] & \xrightarrow{\mathbb{Z}[f]} & \mathbb{Z}[B] \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\beta} \\ A & \xrightarrow[f]{} & B \end{array}$$

is a function $f \colon A \to B$ that preserves the evaluation of formal sums, i.e., a group homomorphism. Thus, we see that the canonical comparison functor defines an isomorphism from the category of abelian groups to the category of algebras for the free abelian group monad.

Our aim in the next section will be to state and prove the **monadicity theorem**, which characterizes when the comparison functor from the domain of a right adjoint to the category of algebras for the induced monad on the domain of the left adjoint is an equivalence of categories. This theorem allows us to recognize when a generic adjunction is **monadic**, i.e., equivalent to the adjunction between the base category and the category of algebras for a monad. A main ingredient in this result is also of independent interest: any algebra for a monad admits a canonical presentation as a quotient of free algebras.

Before proving this result in general, let us explore its meaning in the case of abelian groups. Any abelian group has a presentation that can be defined in terms of the free and forgetful functors

$$\mathsf{Set} \underset{U}{\overset{\mathbb{Z}[-]}{\underset{\perp}{\rightleftarrows}}} \mathsf{Ab}.$$

If G is any set of elements of an abelian group A, there is a canonical map $\mathbb{Z}[G] \to A$ that sends a $\mathbb{Z}$-linear combination of these elements to the element of A that it evaluates to. The set G is a set of **generators** for A precisely when this map is surjective. **Relations** involving these generators are elements of the group $\mathbb{Z}[G]$, so again there is a canonical "evaluation" map $\mathbb{Z}[R] \to \mathbb{Z}[G]$ from the free group on a set R of relations to the free group on the generators. The set $G \subset A$ of generators and $R \subset \mathbb{Z}[G]$ of relations defines a **presentation** of A if the quotient map $\mathbb{Z}[G] \twoheadrightarrow A$ is a coequalizer of the evaluation map and the zero group homomorphism

$$\mathbb{Z}[R] \underset{0}{\overset{\text{evaluation}}{\rightrightarrows}} \mathbb{Z}[G] \twoheadrightarrow A. \tag{5.3.1}$$

One often writes $A = \langle G \mid R \rangle$ when this is the case.

Ad hoc presentations as described by (5.3.1) can be a useful way to describe abelian groups, but they are unlikely to be functorial: a homomorphism $\varphi \colon A \to A'$ is unlikely to carry the presentation for A to any sets of chosen generators and relations for A'. There is, however, a canonical, by which we mean functorial, presentation of any abelian group. Rather than choose a proper subset of generators, we take all of the elements of A to be generators; the canonical evaluation map $\alpha \colon \mathbb{Z}[A] \twoheadrightarrow A$ is certainly surjective. Similarly, rather than choose any particular set of relations in $\mathbb{Z}[A]$, we take all of the elements of $\mathbb{Z}[A]$ to be "relations." Here we don't intend to send every formal sum of elements of A to zero; the result would be the trivial group. Instead, we generalize the meaning of presentation, making use of the fact that coequalizers are more flexible than cokernels. The diagram

$$\mathbb{Z}[\mathbb{Z}[A]] \underset{\mu_A}{\overset{\mathbb{Z}[\alpha]}{\rightrightarrows}} \mathbb{Z}[A] \overset{\alpha}{\twoheadrightarrow} A$$

is always a coequalizer. That is, any abelian group is the quotient of the free abelian group on its underlying set modulo the relation that identifies a formal sum with the element that it evaluates to.

In general:

Proposition 5.3.2. *Let (T, η, μ) be a monad on C and let $(A, TA \overset{\alpha}{\to} A)$ be a T-algebra. Then*

$$(T^2A, \mu_{TA}) \underset{\mu_A}{\overset{T\alpha}{\rightrightarrows}} (TA, \mu_A) \overset{\alpha}{\longrightarrow} (A, \alpha) \tag{5.3.3}$$

is a coequalizer diagram in C^T.

We'll prove this Proposition as a special case of a more general result, which requires some new definitions.

Definition 5.3.4. A **split coequalizer** diagram consists of maps

$$x \underset{g}{\overset{f}{\rightrightarrows}} y \overset{h}{\longrightarrow} z, \qquad t\colon y \to x, \qquad s\colon z \to y$$

so that $hf = hg$, $hs = 1_z$, $gt = 1_y$, and $ft = sh$.

The condition $hf = hg$ says that this triple of maps defines a **fork**, i.e., h is a cone under the parallel pair $f, g\colon x \rightrightarrows y$.

Lemma 5.3.5. *The underlying fork of a split coequalizer diagram is a coequalizer. Moreover, it is an **absolute colimit**: any functor preserves this coequalizer.*

Proof. Given a map $k\colon y \to w$ so that $kf = kg$, we must show that k factors through h; uniqueness of a hypothetical factorization is immediate, as h is a split epimorphism. The factorization is given by the map $ks\colon z \to w$, as demonstrated by the following easy diagram chase:

$$ksh = kft = kgt = k.$$

Now clearly split coequalizers are preserved by any functor, so the universal property of the underlying fork defining a colimit diagram is also preserved. □

Example 5.3.6. For any algebra (A, α) for a monad (T, η, μ) on C, the diagram

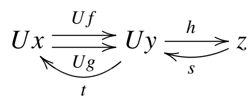
$$T^2A \underset{\mu_A}{\overset{T\alpha}{\rightrightarrows}} TA \overset{\alpha}{\longrightarrow} A, \qquad \eta_{TA}\colon TA \to T^2A, \qquad \eta_A\colon A \to TA$$

defines a split coequalizer diagram in C. Note that while the fork lifts to maps (5.3.3) of algebras, the splittings η_A and η_{TA} do not.

The situation of Example 5.3.6 is captured by the following general definition.

Definition 5.3.7. Given a functor $U\colon \mathsf{D} \to \mathsf{C}$:

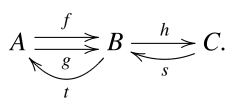
- a **U-split coequalizer** is a parallel pair $f, g\colon x \rightrightarrows y$ in D together with an extension of the pair $Uf, Ug\colon Ux \rightrightarrows Uy$ to a split coequalizer diagram



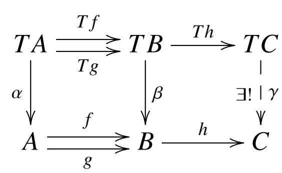
  in C.
- U **creates coequalizers of U-split pairs** if any U-split coequalizer admits a coequalizer in D whose image under U is isomorphic to the U-split coequalizer diagram in C.
- U **strictly creates coequalizers of U-split pairs** if any U-split coequalizer admits a unique lift to a coequalizer in D for the given parallel pair.

Proposition 5.3.8. *For any monad (T, η, μ) on C, the monadic forgetful functor $U^T\colon \mathsf{C}^T \to \mathsf{C}$ strictly creates coequalizers of U^T-split pairs.*

Proof. Suppose given a parallel pair $f, g\colon (A, \alpha) \rightrightarrows (B, \beta)$ in C^T that admits a U^T-splitting:



We must show that C lifts to an algebra (C, γ) and h lifts to an algebra map that is a coequalizer of f and g in C^T, and that moreover these lifts are unique with this property. To define the algebra structure map γ, note that the functor T preserves the split coequalizer diagram; in particular, by Lemma 5.3.5, Th is the coequalizer of Tf and Tg. The algebra structure maps α and β define a diagram

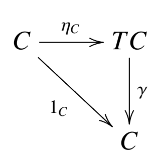


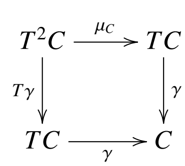
in which the square with the fs and the square with the gs both commute. Thus

$$h \cdot \beta \cdot Tf = h \cdot f \cdot \alpha = h \cdot g \cdot \alpha = h \cdot \beta \cdot Tg,$$

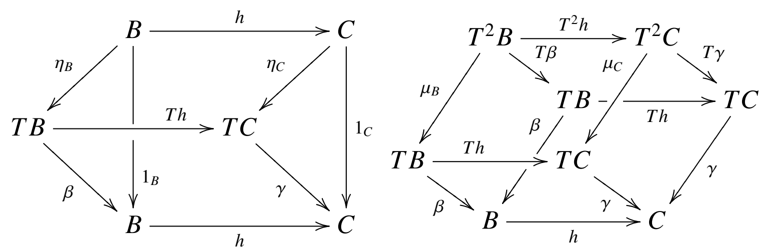
which says that $h\beta$ defines a cone under the pair $Tf, Tg\colon TA \rightrightarrows TB$. By the universal property of their coequalizer, there is a unique map $\gamma\colon TC \to C$ so that the right-hand square above commutes. Once we show that the pair (C, γ) is a T-algebra, this commutative square will demonstrate that $h\colon (B, \beta) \to (C, \gamma)$ is a T-algebra morphism.

We need to check that the diagrams



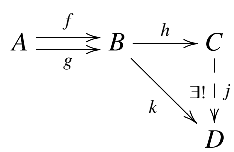


commute. This will follow from the corresponding conditions for the T-algebra (B,β) and the fact that the coequalizer maps are epimorphisms. Specifically, we have diagrams



in which all but the right-most face of each prism is known to commute. It follows that $\gamma \cdot \eta_C \cdot h = 1_C \cdot h$ and $\gamma \cdot \mu_C \cdot T^2h = \gamma \cdot T\gamma \cdot T^2h$, and we conclude that the top faces commute by canceling the epimorphisms h and T^2h. Thus (C,γ) is a T-algebra.

Finally, we show that $h\colon (B,\beta) \to (C,\gamma)$ is a coequalizer in C^T. Given a cone $k\colon (B,\beta) \to (D,\delta)$ so that $kf = kg$, there is a unique factorization



using the universal property of the coequalizer in C. We need only check that j lifts to a map of T-algebras, and again it suffices to verify that $j \cdot \gamma = \delta \cdot Tj$ after precomposing with the epimorphism Th. The result follows from an easy diagram chase, using the fact that h and k are algebra maps:

$$j \cdot \gamma \cdot Th = j \cdot h \cdot \beta = k \cdot \beta = \delta \cdot Tk = \delta \cdot Tj \cdot Th. \qquad \square$$

Proof of Proposition 5.3.2. Example 5.3.6 shows that the fork (5.3.3) is part of a U^T-split coequalizer. In particular, $\alpha\colon TA \to A$ is an absolute coequalizer of the pair $T\alpha, \mu_A\colon T^2A \rightrightarrows TA$ in C. The proof of Proposition 5.3.8 demonstrates that this coequalizer lifts to a coequalizer in C^T. □

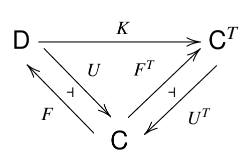
Exercises.

Exercise 5.3.1. The coequalizer of a parallel pair of morphisms f and g in the category Ab is equally the cokernel of the map $f - g$. Explain how the canonical presentation of an abelian group described in Proposition 5.3.2 defines a presentation of that group, in the usual sense.

5.4. Recognizing categories of algebras

There are many versions of the monadicity theorem. For space reasons, we present just one, due to Jon Beck.

Theorem 5.4.1 (monadicity theorem). *Given an adjunction $F \dashv U$ inducing a monad T on C, consider the canonical comparison functor:*



The following are equivalent:

(i) K is an isomorphism of categories.
(ii) U strictly creates coequalizers of U-split pairs.

Proof. Proposition 5.3.8 proves the implication (i)⇒(ii), so we assume (ii) and use it to construct an inverse isomorphism $L: \mathsf{C}^T \to \mathsf{D}$. We have $U^T K = U$ and $KF = F^T$, so if L is to be inverse to K, we must have $U^T = UL$ and $F = LF^T$. Guided by these conditions, we see how to define L on free algebras, namely

$$L(TA, \mu_A) := FA.$$

Similarly, L must carry a free map $Tf: (TA, \mu_A) \to (TB, \mu_B)$ between free algebras to the map $Ff: FA \to FB$. By Exercise 4.2.4, the equations $U^T = UL$ and $F = LF^T$ imply that L will also commute with the counits of the adjunctions $F \dashv U$ and $F^T \dashv U^T$.

An isomorphism (or equivalence) of categories preserves all limits and colimits, by Proposition 4.3.3 for example. In particular, L must carry the canonical coequalizer diagram (5.3.3) to a coequalizer

$$FUFA \underset{\epsilon_{FA}}{\overset{F\alpha}{\rightrightarrows}} FA \dashrightarrow L(A, \alpha), \tag{5.4.2}$$

which exists by hypothesis (ii). We take this to be the definition of $L(A, \alpha)$. The action of L on a morphism $f: (A, \alpha) \to (B, \beta)$ is defined to be the unique map between coequalizers induced by a commutative diagram of parallel pairs:

$$\begin{array}{ccccc} FUFA & \underset{\epsilon_{FA}}{\overset{F\alpha}{\rightrightarrows}} & FA & \longrightarrow & L(A,\alpha) \\ {\scriptstyle FUFf}\downarrow & & \downarrow{\scriptstyle Ff} & & \downarrow{\scriptstyle Lf} \\ FUFB & \underset{\epsilon_{FB}}{\overset{F\beta}{\rightrightarrows}} & FB & \longrightarrow & L(B,\beta) \end{array}$$

Uniqueness of factorizations through colimit cones implies that this definition is functorial.

By construction $F = LF^T$. The pair (5.4.2) is U-split, by Example 5.3.6, so the hypothesis (ii) that U strictly creates coequalizers of U-split pairs tells us that $UL(A, \alpha) = A$; thus $UL = U^T$. Now L is a morphism in the category Adj_T and so we can use the fact that the Eilenberg-Moore adjunction is terminal to see that $KL = 1_{\mathsf{C}^T}$.

It remains only to prove that $LK = 1_{\mathsf{D}}$. Given $D \in \mathsf{D}$, the object LKD is defined to be the coequalizer of

$$FUFUD \underset{\epsilon_{FUD}}{\overset{FU\epsilon_D}{\rightrightarrows}} FUD$$

The map $\epsilon: FUD \to D$ defines a cone under this parallel pair so that when we apply U, we have a split coequalizer diagram.

$$UFUFUD \underset{U\epsilon_{FUD}}{\overset{UFU\epsilon_D}{\rightrightarrows}} UFUD \underset{\eta_{UD}}{\overset{U\epsilon_D}{\rightleftarrows}} UD \qquad \eta_{UFUD}: UFUD \to UFUFUD$$

This implies that $\epsilon: FUD \to D$ is the coequalizer of the pair, and so $D = LKD$ as claimed. ∎

Exercises.

EXERCISE 5.4.1. Recall a monoid is a set M together with maps $\eta\colon 1 \to M$ and $\mu\colon M \times M \to M$ so that certain diagrams commute in Set; see Definition 1.6.2. A monoid homomorphism is a function $f\colon M \to M'$ so that the diagrams

$$\begin{array}{ccc} 1 & & \\ {\scriptstyle \eta}\downarrow & \searrow^{\eta'} & \\ M & \xrightarrow[f]{} & M' \end{array} \qquad \begin{array}{ccc} M \times M & \xrightarrow{f \times f} & M' \times M' \\ {\scriptstyle \mu}\downarrow & & \downarrow{\scriptstyle \mu'} \\ M & \xrightarrow[f]{} & M' \end{array}$$

commute in Set. Prove that the functor $U\colon \mathsf{Monoid} \to \mathsf{Set}$ is monadic by appealing to the monadicity theorem.

EXERCISE 5.4.2. For any group G, the forgetful functor $\mathsf{Set}^G \to \mathsf{Set}$ admits a left adjoint that sends a set X to the G-set $G \times X$, with G acting on the left. Prove that this adjunction is monadic by appealing to the monadicity theorem.

EXERCISE 5.4.3. Generalizing Exercise 5.4.2, for any small category J and any cocomplete category C the forgetful functor $\mathsf{C}^{\mathsf{J}} \to \mathsf{C}^{\mathrm{ob}\,\mathsf{J}}$ admits a left adjoint $\mathrm{Lan}\colon \mathsf{C}^{\mathrm{ob}\,\mathsf{J}} \to \mathsf{C}^{\mathsf{J}}$ that sends a functor $F \in \mathsf{C}^{\mathrm{ob}\,\mathsf{J}}$ to the functor $\mathrm{Lan}F \in \mathsf{C}^{\mathsf{J}}$ defined by

$$\mathrm{Lan}F(j) = \coprod_{x \in \mathsf{J}} \coprod_{\mathsf{J}(x,j)} Fx.$$

(i) Define $\mathrm{Lan}F$ on morphisms in J.
(ii) Define Lan on morphisms in $\mathsf{C}^{\mathrm{ob}\,\mathsf{J}}$.
(iii) Use the Yoneda lemma to show that Lan is left adjoint to the forgetful (restriction) functor $\mathsf{C}^{\mathsf{J}} \to \mathsf{C}^{\mathrm{ob}\,\mathsf{J}}$.
(iv) Prove that this adjunction is monadic by appealing the monadicity theorem.

EXERCISE 5.4.4. Describe a more general class of functors $K\colon \mathsf{I} \to \mathsf{J}$ between small categories so that for any cocomplete C the restriction functor $\mathrm{res}_K\colon \mathsf{C}^{\mathsf{J}} \to \mathsf{C}^{\mathsf{I}}$ strictly creates colimits of res_K-split parallel pairs. All such functors admit left adjoints and are thus monadic. Challenge: describe the left adjoint.

5.5. Limits and colimits in categories of algebras

A category A is **monadic over** C if there is an adjunction

$$\mathsf{C} \underset{U}{\overset{F}{\underset{\bot}{\rightleftarrows}}} \mathsf{A}$$

so that the right adjoint U creates coequalizers of U-split pairs or equivalently, by Theorem 5.4.1 if the canonical comparison functor from A to the category of algebras for the monad UF on C is an equivalence. A functor $U\colon \mathsf{A} \to \mathsf{C}$ is **monadic** if it admits a right adjoint and if it creates coequalizers of U-split pairs.

EXAMPLES 5.5.1. The following categories are monadic over Set via the free-forgetful adjunctions described in Example 4.1.10.

(i) Monoid, Group, Ab, Ring, and other variants, such as commutative rings or monoids, or non-unital versions of the preceding.
(ii) Mod_R, $\mathsf{Vect}_{\Bbbk}$, $\mathsf{Aff}_{\Bbbk}$, Set^G.
(iii) $\mathsf{Lattice}$ or the categories of sup or meet semi-lattices.
(iv) Set_*.

(v) kHaus, the category of compact Hausdorff spaces.

The categories listed in Examples 5.5.1.(i)-(iv) are categories of **models of an algebraic theory**. More generally, any category of models for an algebraic theory is monadic over Set.

EXAMPLES 5.5.2. Other examples of monadic adjunctions include:

(i) The forgetful functor $U\colon \mathsf{Ring} \to \mathsf{Ab}$ is monadic. The induced monad on Ab is the free monoid monad $TA := \oplus_{n\geq 0} A^{\otimes n}$.
(ii) The forgetful functor $U\colon \mathsf{Mod}_R \to \mathsf{Ab}$ is monadic. The induced monad on Ab is given by $R \otimes_{\mathbb{Z}} -\colon \mathsf{Ab} \to \mathsf{Ab}$.
(iii) If C has coproducts and J is small then Exercise 5.4.3 demonstrates that the restriction functor $\mathsf{C}^{\mathsf{J}} \to \mathsf{C}^{\mathrm{ob}\,\mathsf{J}}$, which carries a J-indexed diagram in C to the ob J-indexed family of objects in its image, is monadic. The underlying set functor $\mathsf{Set}^G \to \mathsf{Set}$ of Examples 5.5.1.(ii) is a special case of this.
(iv) $U\colon \mathsf{Cat} \to \mathsf{DirGraph}$ is also monadic. The monad forms (the underlying graph of) the free category on a directed graph, where identities and composites of all finite directed paths are formally added. This functor factors through a forgetful functor $U\colon \mathsf{Cat} \to \mathsf{rDirGraph}$ whose objects are **reflexive directed graphs**, with specified "identity" endoarrows of each vertex, that is also monadic.

Our aim in this section is to present the common properties shared by categories A that arise in this way.

LEMMA 5.5.3. *If $U\colon \mathsf{A} \to \mathsf{C}$ is monadic, then U is conservative: that is for any morphism $f\colon a \to a'$ in A, if Uf is an isomorphism in C, then f is an isomorphism in A.*

PROOF. A monadic functor $U\colon \mathsf{A} \to \mathsf{C}$ is equivalent to the forgetful functor $U^T\colon \mathsf{C}^T \to \mathsf{C}$ from the category of algebras for the induced monad on C, and so it suffices to demonstrate that U^T has this property. Recall that a morphism $f\colon (A,\alpha) \to (A',\alpha')$ in C^T is a map $f\colon A \to A'$ in C so that the left-hand diagram

$$\begin{array}{ccc} TA & \xrightarrow{Tf} & TA' \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha'} \\ A & \xrightarrow[f]{} & A' \end{array} \qquad\qquad \begin{array}{ccc} TA' & \xrightarrow{Tf^{-1}} & TA \\ {\scriptstyle\alpha'}\downarrow & & \downarrow{\scriptstyle\alpha} \\ A' & \xrightarrow[f^{-1}]{} & A \end{array}$$

commutes in C, whence the right-hand diagram also commutes whenever the inverse exists in C. □

Lemma 5.5.3 answers the question motivated by Examples 1.1.6.

COROLLARY 5.5.4. *Any bijective continuous function between compact Hausdorff spaces is a homeomorphism.*[8]

By contrast, the categories Top or Poset are not monadic over sets: there exists non-invertible maps that act as the identity on underlying sets.

COROLLARY 5.5.5. *Any bijective homomorphism between models of an algebraic theory is an isomorphism.*

[8]More generally any continuous bijection from a compact space to a Hausdorff one is a homeomorphism; but compactness of the domain and surjectivity of the map imply that the codomain is also compact, while Hausdorffness of the codomain and injectivity of the map imply that the domain is also Hausdorff.

For instance, the inverse of a bijective homomorphism between groups is also a homomorphism. This of course, is not difficult to prove. Corollary 5.5.5 eliminates the redundancy of proving the same result over and over again in many similar contexts.

We now turn our focus to limits and colimits in a category of algebras.

Proposition 5.5.6. *A monadic functor $U\colon \mathsf{A} \to \mathsf{C}$ creates*

(i) any limits that C *has and*
(ii) any colimits that C *has and the monad and its square preserve.*

Proof. We start with (i). Again it suffices to prove this result for the forgetful functor $U^T\colon \mathsf{C}^T \to \mathsf{C}$ for a monad (see Exercise 3.3.3). Consider a diagram $D\colon \mathsf{J} \to \mathsf{C}^T$, spanning objects $(Dj, \gamma_j) \in \mathsf{C}^T$, so that the underlying diagram $U^T D\colon \mathsf{J} \to \mathsf{C}$ admits a limit cone $\mu\colon L \Rightarrow Dj$ in C. We wish to lift the summit L and the legs of the limit cone to a limit cone for the diagram D in C^T. The fact that D is a diagram of algebras implies that the algebra structure maps assemble into the components of a natural transformation $\gamma\colon TD \Rightarrow D\colon \mathsf{J} \to \mathsf{C}$ to the underlying C-valued diagram D from the composite C-valued diagram TD. Composing,

$$TL \overset{T\mu}{\Rightarrow} TD \overset{\gamma}{\Rightarrow} D$$

defines a cone with summit TL over $D\colon \mathsf{J} \to \mathsf{C}$, which factors through the limit cone to define a unique map $\lambda\colon TL \to L$ in C so that the diagrams

$$\begin{array}{ccc} TL & \xrightarrow{T\mu_j} & TDj \\ \downarrow{\scriptstyle\lambda} & & \downarrow{\scriptstyle\gamma_j} \\ L & \xrightarrow[\mu_j]{} & Dj \end{array}$$

commute for each $j \in \mathsf{J}$. Appealing again to the universal property of the limit L in C, it's easy to verify that (L, λ) is a T-algebra; the point is that the required diagrams abut to L, permitting the application of its universal property. Thus, μ lifts to a cone $\mu\colon (L, \lambda) \Rightarrow D$ over the diagram $D\colon \mathsf{J} \to \mathsf{C}^T$. The verification that this is a limit cone proceeds similarly and is left as an exercise.

A similar argument works for (ii) with the additional hypothesis that T and T^2 preserve the colimit cone in C under consideration. If $L \in \mathsf{C}$ is the nadir of a colimit cone under a diagram $U^T D\colon \mathsf{J} \to \mathsf{C}$ then its algebra structure map $\lambda\colon TL \to L$ will be induced by the universal property of the colimit cone under $TU^T D$ with nadir TL. □

Corollary 5.5.7. *Any category, such as* Ab, Group, Ring, Mod_R, *that is monadic over* Set *is complete. Moreover, limits in these categories are created from their underlying sets.*

Proof. Theorem 3.2.2 proves that Set is complete. Proposition 5.5.6 demonstrates that is monadic over Set is also complete. □

Example 5.5.8. For instance, the p-adic integers are defined to be the limit of an ω^{op}-indexed diagram of rings

$$\mathbb{Z}_p := \lim_n \mathbb{Z}/p^n \longrightarrow \cdots \longrightarrow \mathbb{Z}/p^3 \longrightarrow \mathbb{Z}/p^2 \longrightarrow \mathbb{Z}/p$$

We can use the fact that $U\colon \mathsf{Ring} \to \mathsf{Set}$ preserves limits to describe the p-adic integers, using our familiarity with the construction of limits in the category of sets: as a set we have

$$\mathbb{Z}_p = \{(a_1 \in \mathbb{Z}/p, a_2 \in \mathbb{Z}/p^2, a_3 \in \mathbb{Z}/p^3, \ldots) \mid a_n \equiv a_m \bmod p^{\min(n,m)}\}$$

That is, a p-adic integer is a sequence of elements $a_n \in \mathbb{Z}/p^n$ which are compatible modulo congruence.

The underlying set functor also creates the limit cone, which tells us that the ring structure on this set of elements must be defined in such a way so that the projection functions $\mathbb{Z}_p \to \mathbb{Z}/p^n$ are ring homomorphisms. This tells us that addition and multiplication of elements is "componentwise."

Proposition 5.5.6 shows that Field is not monadic over Set. The category of fields does not admit products of any fields of different characteristic.

Corollary 5.5.9. Set *is cocomplete.*

Proof. The contravariant power-set functor $P\colon \mathsf{Set}^{\mathrm{op}} \to \mathsf{Set}$ is monadic, so the colimit of a diagram is created from the limit of the corresponding diagram on powersets. □

Corollary 5.5.10. $\mathsf{Mod}_R \to \mathsf{Ab}$ *creates all colimits that* Ab *admits.*

Proof. The monad $R \otimes_{\mathbb{Z}} -\colon \mathsf{Ab} \to \mathsf{Ab}$ has a right adjoint $\mathrm{Hom}_{\mathbb{Z}}(R, -)$, which carries an abelian group A to the group of homomorphisms $R \to A$. The natural isomorphism

$$\mathrm{Hom}(R \otimes_{\mathbb{Z}} A, A') \cong \mathrm{Hom}(A, \mathrm{Hom}_{\mathbb{Z}}(R, A'))$$

can be deduced from the defining universal property of the tensor product; both the left-hand and right-hand sides correspond to bilinear maps $R \times A \to A'$. □

In fact, these categories are all cocomplete. However, unless the induced monad preserves colimits of a particular shape, they will not be created by the monadic forgetful functor U. To explore this type of setting, consider the monadic forgetful functor $U\colon \mathsf{Group} \to \mathsf{Set}$. We've seen that both Set and Group admit coproducts; in Set these are simply disjoint unions, while in Group they are given by the free product. The **free product** of groups G and H is the group $G * H$ of words in G and H modulo relations defined using the group operations in each group. Note in particular that $U\colon \mathsf{Group} \to \mathsf{Set}$ does not preserve (and so in particular does not create) coproducts.

Let us now define the free product $G * H$ more precisely. Our description will be given entirely in terms of the free-forgetful adjunction

$$\mathsf{Group} \underset{U}{\overset{F}{\underset{\longrightarrow}{\overset{\longleftarrow}{\bot}}}} \mathsf{Set}$$

In particular, it will be generalizable to any monadic adjunction.

A first approximation to the free product $G * H$ is given by $F(UG \coprod UH)$, the free group on the disjoint union of the underlying sets of G and H. Elements of this group are words in the letters G and H but we have not yet imposed the relations generated by words in G and words in H. Now words in G and words in H are elements of the set $UFUG \coprod UFUH$ defined to be the disjoint union of the underlying sets of the free groups on the underlying sets of G and H. The free group on this set, $F(UFUG \coprod UFUH)$ is the group of words of words in the letters G and H so that each subword is exclusively comprised of letters drawn from a single one of these groups. The desired relations, that define $G * H$ as a quotient of $F(UG \coprod UH)$, identify a word of words of this type with the word obtained by evaluating each subword using the group structure of G or of H.

To encode the desired relations, we define a natural pair of group homomorphisms.

$$\begin{array}{ccc} F(UFUG \coprod UFUH) & \xrightarrow{F(U\epsilon_G \coprod U\epsilon_H)} & F(UG \coprod UH) \\ & {\scriptstyle F\kappa}\searrow \quad \nearrow{\scriptstyle \epsilon_{F(UG \coprod UH)}} & \\ & FUF(UG \coprod UH) & \end{array}$$

The fact that G and H are groups is encoded by a pair of homomorphisms $\epsilon_G \colon FUG \to G$ and $\epsilon_H \colon FUH \to H$ that evaluate a word in the letters of G and formal inverses to the group element that it represents. The top map sends an element in the group $F(UFUG \coprod UFUH)$, a word of words in which these subwords are exclusively in the group G or in the group H, to a word in $UG \coprod UH$ by evaluating each subword to the corresponding group element.

The bottom composite map first makes use of the natural map κ of Exercise 3.3.4, that compares the coproduct of the image of two objects under a functor, in this case UF, with the image of the coproduct under that functor. Here this has the effect of regarding a word of words, with subwords exclusively in either G or H, as a word of words whose letters might belong to either G or H, but happen in this case to belong to only one or the other. The map $\epsilon_{F(UG \coprod UH)}$ then concatenates to produce a single word in letters of G and H. From this point, we see that the coequalizer of these two group homomorphisms imposes exactly the relations desired to define the free product:

$$\begin{array}{cccc} F(UFUG \coprod UFUH) & \xrightarrow{F(U\epsilon_G \coprod U\epsilon_H)} & F(UG \coprod UH) & \twoheadrightarrow G * H \\ & {\scriptstyle F\kappa}\searrow \quad \nearrow{\scriptstyle \epsilon_{F(UG \coprod UH)}} & & \\ & FUF(UG \coprod UH) & & \end{array}$$

Theorem 5.5.11. *Suppose* C *is cocomplete and* $U \colon \mathsf{A} \to \mathsf{C}$ *is monadic. Then the following are equivalent:*

(i) A *is cocomplete*
(ii) A *has coequalizers*

Proof. The implication $(i) \Rightarrow (ii)$ is trivial. For $(ii) \Rightarrow (i)$ it suffices by Theorem 3.3.7 to prove that a category of algebras that has coequalizers also has coproducts. The coproduct (A, α) of a family $(A_i, \alpha_i) \in \mathsf{C}^T$ is defined to be the following coequalizer

$$\begin{array}{cccc} T(\coprod_i T(A_i), \mu_{\coprod_i T(A_i)}) & \xrightarrow{T(\coprod_i \alpha_i)} & (T(\coprod_i A_i), \mu_{\coprod_i A_i}) & \xrightarrow{q} (A, \alpha) \\ & {\scriptstyle T\kappa}\searrow \quad \nearrow{\scriptstyle \mu_{\coprod_i A_i}} & & \\ & (T^2(\coprod_i A_i), \mu_{T(\coprod_i A_i)}) & & \end{array}$$

Note that the objects in the coequalizer diagram are all free algebras. Using the adjunction $F^T \dashv U^T$, it is straightforward to show that (A, α) has the desired universal property of the coproduct of the objects (A_i, α_i). □

Under certain conditions on C it is possible to prove that the category of algebras for any monad on C has coequalizers and is therefore cocomplete.

Definition 5.5.12. A category C is **regular** when:

- every arrow has a kernel pair (see Example 3.4.5)
- every kernel pair has a coequalizer

- the pullback of a regular epimorphism along any morphism exists and is again a regular epimorphism

A **regular epimorphism** is a morphism that appears as the coequalizer of some parallel pair of maps into its domain. Exercise 3.1.3 demonstrates that any such map is an epimorphism.

Examples of regular categories include Set, $\mathsf{Set}^{\mathsf{J}}$, and many other categories, as we shall soon discover.

THEOREM 5.5.13. *Suppose* C *is a locally small complete and cocomplete regular category in which every regular epimorphism has a section.*[9] *Then for every monad on* C *its category of algebras is complete, cocomplete, and regular.*

The axioms of a regular category imply that any morphism may be factored (uniquely up to isomorphism) as a regular epimorphism followed by a monomorphism. The intermediate object in this factorization is called the **image** of the map, and the factorization is called the **image factorization**. Rather than prove this, the reader may wish to interpret the following proof in the familiar case of $\mathsf{C} = \mathsf{Set}$.

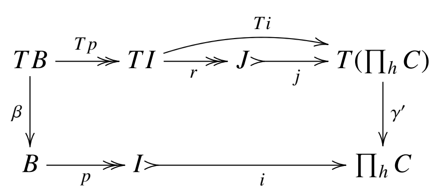
PROOF. By Theorem 5.5.13, to show that C^T is cocomplete, it suffices to construct the coequalizer of a pair of maps $f, g\colon (A, \alpha) \rightrightarrows (B,\beta)$. Consider all morphisms $h\colon (B,\beta) \to (C, \gamma)$ in C^T so that $hf = hg$ and $h\colon B \to C$ is a regular epimorphism in C. This can be thought of as the set of "quotients" of (B,β) on which the maps f and g agree. By hypothesis, any regular epimorphism h in C has a section s. It's easy to see that any epimorphism h with a section s is then the coequalizer of the idempotent sh and the identity. This allows us to conclude that the set of "quotients" of (B,β) is really only a set: the object B has at most a set's worth of idempotents. It also follows that Th is again a regular epimorphism, for the same reasoning shows that Th is the coequalizer of $Ts \cdot Th$ and the identity.

In the regular category C, we may construct the image factorization of the canonical map $q\colon B \to \prod_h C$

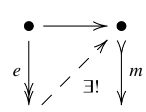
$$B \overset{p}{\twoheadrightarrow} I \overset{i}{\rightarrowtail} \textstyle\prod_h C.$$

We will show that I admits the structure of a T-algebra so that p and i are T-algebra maps, and furthermore that p exhibits this T-algebra as the coequalizer of f and g. Because $U^T\colon \mathsf{C}^T \to \mathsf{C}$ creates products, we can lift q to a map of T-coalgebras $q\colon (B,\beta) \to (\prod_h C, \gamma') := \prod_h (C, \gamma)$. This gives the following commutative diagram in C:

(5.5.14)



As p is a regular epimorphism in C, Tp is as well. We obtain a second regular epimorphism r from the image factorization of Ti. In a regular category, whenever we are presented with a commutative square



[9] In Set, this follows from the axiom of choice.

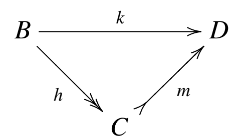
whose left-hand side is a regular epimorphism and whose right-hand side is a monomorphism, there exists a unique diagonal morphism, as displayed, that makes both diagrams commute. Applying this "lifting property" to (5.5.14), we obtain a morphism $k\colon J \to I$ that commutes with the rest of the diagram. Defining $\iota := kr$, we have a candidate algebra structure morphism for I with the property that $p\colon (B,\beta) \to (I,\iota)$ and $j\colon (I,\iota) \to \prod_h (C,\gamma)$ are both maps of algebras.

To prove that ι is an algebra structure map, note that

$$\iota \cdot \eta_I \cdot p = \iota \cdot Tp \cdot \eta_B = p \cdot \beta \cdot \eta_B = p,$$

which implies that $\iota \cdot \eta_I = 1_I$ after canceling the regular epimorphism p. The proof that $\iota \cdot T\iota = \iota \cdot \mu_I$ follows similarly from the corresponding property for (B,β) upon canceling the regular epimorphism $T^2 p$. Now we know that $p\colon (B,\beta) \to (I,\iota)$ is a map of T-algebras as claimed. It remains only to show that it is the coequalizer of f and g.

First note that $ipf = qf = qg = ipg$ implies $pf = pg$ upon canceling the monomorphism i. Next suppose $k\colon (B,\beta) \to (D,\delta)$ is a map of T-algebras so that $kf = kg$. Consider the image factorization



of the underlying map h. The argument just give shows more generally that C lifts to a T-algebra (C,γ) so that h and m are algebra maps. As h is a regular epimorphism, we have a factorization $h = \pi_h \cdot i \cdot p$ in C^T, which shows that the map k factors through the map i. This factorization is unique because p is an epimorphism in C and thus also in C^T.

We defer to [**Bor94**, 4.3.5] for the proof that C^T is again a regular category. □

Exercises.

Exercise 5.5.1. Show that the inclusion of any reflective subcategory is monadic and conclude that the reflective subcategory inherits any limits contained in the larger category, completing the unfinished business of Proposition 4.4.11.

CHAPTER 6

All Concepts are Kan Extensions

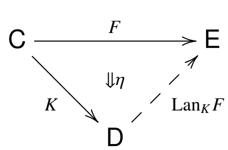
> The notion of Kan extensions subsumes all the other fundamental concepts of category theory.
>
> Saunders Mac Lane, "Categories for the Working Mathematician"

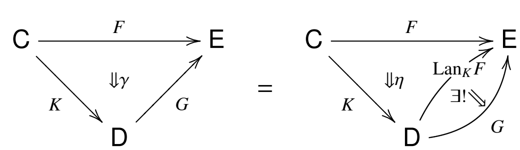
Extension problems are pervasive in mathematics. For instance, a fixed positive real number, such as 2, has well-defined natural number powers, defined to be repeated products. The resulting function $2^{-}\colon \mathbb{N} \to \mathbb{R}_{>0}$ can be extended to a homomorphism from the additive group of integers to the multiplicative group of non-negative reals by declaring $2^{-n} := (\frac{1}{2})^n$ for $n \in \mathbb{N}$. The resulting function $2^{-}\colon \mathbb{Z} \to \mathbb{R}_{>0}$ can be extended further to an additive homomorphism defined on the rationals by declaring $2^{\frac{1}{n}} := \sqrt[n]{2}$ for $n \in \mathbb{N}$. Finally, $2^{-}\colon \mathbb{Q} \to \mathbb{R}_{>0}$ can be extended to a function on the reals, though this final extension is not given by some explicitly described arithmetic formula, but rather by taking advantage of the fact that $\mathbb{R}_{>0}$ has limits of all bounded increasing sequences.

The construction of $2^{-}\colon \mathbb{R} \to \mathbb{R}_{>0}$ as an order-preserving function extending $2^{-}\colon \mathbb{Q} \to \mathbb{R}_{>0}$ is a special case of a general solution to a categorically-defined extension problem. Given a pair of functors $K\colon \mathsf{C} \to \mathsf{D}$, $F\colon \mathsf{C} \to \mathsf{E}$, it may or may not be possible to extend F along K. Obstructions can take several forms: two arrows in C with distinct images in E might be identified in D, or two objects might have empty hom-sets in C and E but not in D. In general, it is more reasonable to ask for a best approximation to an extension taking the form of a universal natural transformation pointing either from or to F. The resulting categorical notion, quite simple to define, is surprisingly ubiquitous throughout mathematics, as we shall soon discover.

6.1. Kan extensions

Definition 6.1.1. Given functors $F\colon \mathsf{C} \to \mathsf{E}$, $K\colon \mathsf{C} \to \mathsf{D}$, a **left Kan extension** of F along K is a functor $\mathrm{Lan}_K F\colon \mathsf{D} \to \mathsf{E}$ together with a natural transformation $\eta\colon F \Rightarrow \mathrm{Lan}_K F \cdot K$ such that for any other such pair $(G\colon \mathsf{D} \to \mathsf{E}, \gamma\colon F \Rightarrow GK)$, γ factors uniquely through η as illustrated.[1]

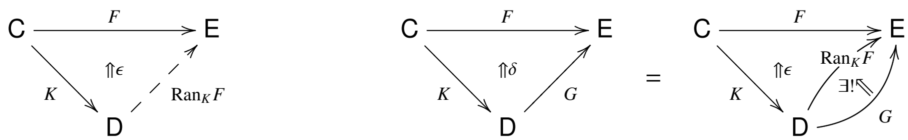




Dually, a **right Kan extension** of F along K is a functor $\mathrm{Ran}_K F\colon \mathsf{D} \to \mathsf{E}$ together with a natural transformation $\epsilon\colon \mathrm{Ran}_K F \cdot K \Rightarrow F$ such that for any $(G\colon \mathsf{D} \to \mathsf{E}, \delta\colon GK \Rightarrow F)$, δ

[1] Writing α for the natural transformation $\mathrm{Lan}_K F \Rightarrow G$, the right-hand **pasting diagrams** express the equality $\gamma = \alpha K \cdot \eta$, i.e., that γ factors as $F \xRightarrow{\eta} \mathrm{Lan}_K F \cdot K \xRightarrow{\alpha K} GK$.

factors uniquely through ϵ as illustrated.



The intuition is clearest when the functor $K\colon \mathsf{C} \to \mathsf{D}$ of Definition 6.1.1 embedds C as a full subcategory of D; assuming certain (co)limits exist, when K is fully faithful, the left and right Kan extensions do in fact extend the functor F along K; see 6.2.3. However, extensions of a functor do not necessarily provide Kan extensions; see Exercise 6.1.1.

Example 6.1.2. The Yoneda lemma says that for any $a \in A$, the representable functor $A(a,-)$ is a left Kan extension of the terminal object $\mathbb{1} \to \mathsf{Set}$ along $a\colon \mathbb{1} \to A$.

Derived functors in homological algebra or algebraic topology are defined using Kan extensions.

Example 6.1.3. In good situations, the composite of a functor $F\colon \mathsf{C} \to \mathsf{D}$ between categories equipped with a subcategory of "weak equivalences" and the localization functor $\mathsf{D} \to \mathsf{HoD}$, that formally inverts the specified class of weak equivalences, admits a right or left Kan extension along the localization functor $\mathsf{C} \to \mathsf{HoC}$, called the **total left derived functor** or **total right derived functor**, respectively.

We can reexpress the universal property that defines Kan extensions as a representation for an appropriate Set-valued functor. A left Kan extension of $F\colon \mathsf{C} \to \mathsf{E}$ along $K\colon \mathsf{C} \to \mathsf{D}$ is a representation for the functor

$$\mathsf{E}^{\mathsf{C}}(F, - \circ K)\colon \mathsf{E}^{\mathsf{D}} \to \mathsf{Set}$$

that sends a functor $\mathsf{D} \to \mathsf{E}$ to the set of natural transformations from F to its restriction along K. By the Yoneda lemma, any pair (G,γ) as in Definition 6.1.1 defines a natural transformation

$$\mathsf{E}^{\mathsf{D}}(G,-) \xRightarrow{\gamma} \mathsf{E}^{\mathsf{C}}(F, - \circ K).$$

The universal property of the pair $(\mathrm{Lan}_K F, \eta)$ is equivalent to the assertion that the corresponding map

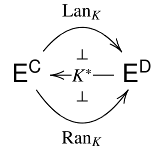
$$\mathsf{E}^{\mathsf{D}}(\mathrm{Lan}_K F,-) \xRightarrow{\eta} \mathsf{E}^{\mathsf{C}}(F, - \circ K)$$

is a natural isomorphism, i.e., that $(\mathrm{Lan}_K F, \eta)$ represents this functor.

Extending this discussion it follows that if, for fixed K, the left and right Kan extensions of any functor $\mathsf{C} \to \mathsf{E}$ exist, then these define left and right adjoints to the precomposition functor $K^*\colon \mathsf{E}^{\mathsf{D}} \to \mathsf{E}^{\mathsf{C}}$.

(6.1.4)

$$\mathsf{E}^{\mathsf{D}}(\mathrm{Lan}_K F, G) \cong \mathsf{E}^{\mathsf{C}}(F, GK)$$



$$\mathsf{E}^{\mathsf{C}}(GK, F) \cong \mathsf{E}^{\mathsf{D}}(G, \mathrm{Ran}_K F)$$

The 2-cells η are the components of the unit for $\mathrm{Lan}_K \dashv K^*$ and the 2-cells ϵ are the components of the counit for $K^* \dashv \mathrm{Ran}_K$. The universal properties of Definition 6.1.1 are precisely those required to define the value at a particular object $F \in \mathsf{E}^{\mathsf{C}}$ of a left and right adjoint to a specified functor, in this case K^*.

Conversely, by uniqueness of adjoints, the objects in the image of any left or right adjoint to a precomposition functor are Kan extensions. This observation leads to several immediate examples.

Example 6.1.5. A small category with a single object and only invertible arrows is precisely a (discrete) group. The objects of the functor category Vect_k^G are G-representations over a fixed field k; arrows are G-equivariant linear maps. If H is a subgroup of G, restriction $\mathsf{Vect}_k^G \to \mathsf{Vect}_k^H$ of a G-representation to an H-representation is simply precomposition by the inclusion functor $i \colon H \hookrightarrow G$. This functor has a left adjoint, induction, which is left Kan extension along i. The right adjoint, coinduction, is right Kan extension along i.

$$\mathsf{Vect}_k^G \xrightarrow{\text{res}} \mathsf{Vect}_k^H, \qquad \text{ind}_H^G \dashv \text{res} \dashv \text{coin}_H^G \tag{6.1.6}$$

The reader unfamiliar with the construction of induced representations need not remain in suspense for very long; see Theorem 6.2.1 and Example 6.2.5. Similar remarks apply for G-sets, G-spaces, based G-spaces, or indeed G-objects in any category—although in the general case these adjoints might not exist.

Remark 6.1.7. This example can be enriched: extension of scalars, taking an R-module M to the S-module $M \otimes_R S$, is the Ab-enriched left Kan extension along an Ab-functor $R \to S$ between one-object Ab-categories, more commonly called a ring homomorphism.

Example 6.1.8. Let $\mathbb{\Delta}$ be the category of finite non-empty ordinals and order preserving maps. **Presheaves** on $\mathbb{\Delta}$, i.e., contravariant Set-valued functors, are called **simplicial sets**. $\mathbb{\Delta}$ is also called the **simplex category**. The ordinal $n + 1 = \{0, 1, \ldots, n\}$ is associated with the topological n-simplex and, with this interpretation in mind, is typically denoted by "$[n]$."

Write $\mathbb{\Delta}_{\leq n}$ for the full subcategory spanned by the first $n+1$-ordinals. Restriction along the inclusion functor $i_n \colon \mathbb{\Delta}_{\leq n} \hookrightarrow \mathbb{\Delta}$ is called n-truncation. This functor has both left and right Kan extensions:

$$\mathsf{Set}^{\mathbb{\Delta}^{\text{op}}} \xrightarrow{i_n^*} \mathsf{Set}^{\mathbb{\Delta}_{\leq n}^{\text{op}}}, \qquad \text{Lan}_{i_n} \dashv i_n^* \dashv \text{Ran}_{i_n}$$

The composite comonad on $\mathsf{Set}^{\mathbb{\Delta}^{\text{op}}}$ is sk_n, the functor that maps a simplicial set to its n-**skeleton**. The composite monad on $\mathsf{Set}^{\mathbb{\Delta}^{\text{op}}}$ is cosk_n, the functor that maps a simplicial set to its n-**coskeleton**. Furthermore, sk_n is left adjoint to cosk_n, as is the case for any comonad and monad arising in this way.

Example 6.1.9. The category $\mathbb{\Delta}$ is a full subcategory containing all but the initial object of the category $\mathbb{\Delta}_+$ of finite ordinals and order preserving maps. Presheaves on $\mathbb{\Delta}_+$ are called **augmented simplicial sets**. Left Kan extension defines a left adjoint to restriction

$$\mathsf{Set}^{\mathbb{\Delta}_+^{\text{op}}} \xrightarrow{\text{res}} \mathsf{Set}^{\mathbb{\Delta}^{\text{op}}}, \qquad \pi_0 \dashv \text{res} \dashv \text{triv}$$

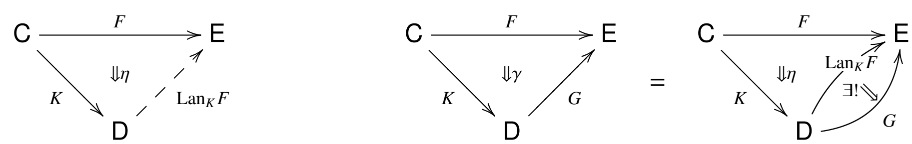
that augments a simplicial set X with its set $\pi_0 X$ of path components. Right Kan extension assigns a simplicial set the trivial augmentation built from the one-point set.

Exercises.

EXERCISE 6.1.1. Construct a toy example to illustrate that if F factors through K along some functor H, it is not necessarily the case that $(H, 1_F)$ is the left Kan extension of F along K.

6.2. A formula for Kan extensions

Importantly, if the target category E has certain limits and colimits, then right and left Kan extensions for any pair of functors exist and furthermore can be computed by a particular (co)limit formula. In the case of left Kan extensions we are seeking to define a functor $\mathrm{Lan}_K F\colon \mathsf{D} \to \mathsf{E}$ that is the "closest approximation to an extension of F along K from the left," i.e., up to a natural transformation η



which is universal among natural transformations $\gamma\colon F \Rightarrow GK$. So, to define the value of $\mathrm{Lan}_K F(d)$ at an object $d \in \mathsf{D}$, we should consider all possible approximations to d that come from the category C. This leads us to consider the comma category $K \downarrow d$ whose objects are morphisms $Kc \to d$ from the image of a specified $c \in \mathsf{C}$ and whose morphisms are morphisms $c \to c'$ in C that give rise to commutative triangles in D. Recall that $K \downarrow d$ is the category of elements of the functor $\mathsf{D}(K-, d)\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$, and as such comes with a canonical projection functor $\Pi^d_K\colon K \downarrow d \to \mathsf{C}$. The approximation to the value of $\mathrm{Lan}_K F(d)$ associated to an object of $K \downarrow d$ is given by the composite $K \downarrow d \xrightarrow{\Pi^d_K} \mathsf{C} \xrightarrow{F} \mathsf{E}$ of the projection functor with F.

THEOREM 6.2.1. *Given functors $F\colon \mathsf{C} \to \mathsf{E}$ and $K\colon \mathsf{C} \to \mathsf{D}$ with C small and D locally small, if for every $d \in \mathsf{D}$ the colimit*

$$\mathrm{Lan}_K F(d) = \mathrm{colim}(K \downarrow d \xrightarrow{\Pi^d_K} \mathsf{C} \xrightarrow{F} \mathsf{E}) \tag{6.2.2}$$

exists then they define the left Kan extension $\mathrm{Lan}_K F\colon \mathsf{D} \to \mathsf{E}$ and the unit transformation $\eta\colon F \Rightarrow \mathrm{Lan}_K F \cdot K$ can be extracted from the colimit cone. Dually, if for every $d \in \mathsf{D}$ the limit

$$\mathrm{Ran}_K F(d) = \lim(d \downarrow K \xrightarrow{\Pi^K_d} \mathsf{C} \xrightarrow{F} \mathsf{E})$$

exists then they define the right Kan extension $\mathrm{Ran}_K F\colon \mathsf{D} \to \mathsf{E}$ and the counit transformation $\epsilon\colon \mathrm{Ran}_K F \cdot K \Rightarrow F$ can be extracted from the limit cone.

PROOF. For now, see [**ML98a**, X.4.1-2]. □

When E has sufficient limits or colimits so that Theorem 6.2.1 applies, if K is fully faithful then the Kan extension indeed defines an extension along K.

COROLLARY 6.2.3. *If E is complete and K is fully faithful, then the counit defines a natural isomorphism $\mathrm{Ran}_K F \cdot K \cong F$. Dually, if E is cocomplete and K is fully faithful, then the unit defines a natural isomorphism $F \cong \mathrm{Lan}_K F \cdot F$.*

PROOF. Consider the comma category $K \downarrow Kc$ for some $c \in \mathsf{C}$. If K is fully faithful, then this is isomorphic to $\mathsf{C} \downarrow c$ and this isomorphism commutes with the projection functors to C. We have

$$\mathrm{Lan}_K F(Kc) := \mathrm{colim}(\mathsf{C} \downarrow c \xrightarrow{\Pi} \mathsf{C} \xrightarrow{F} \mathsf{E},$$

but in this case the indexing category $\mathsf{C} \downarrow c$ has a terminal object 1_c. Therefore, by Exercise 3.1.6, the colimit can be computed by evaluating at the terminal object, and so $\mathrm{Lan}_K F(Kc) \cong Fc$. □

EXAMPLE 6.2.4. Consider the partial order $\mathbb{Q}$ of the rationals and $\mathbb{R}_{>0}$ of the non-negative reals. We have a functor $2^{-} \colon \mathbb{Q} \to \mathbb{R}_{>0}$ defined as in the introduction to this chapter. We can extend its target to $\bar{R}_{\geq 0} := \mathbb{R}_{\geq 0} \cup \{\infty\}$, which is a complete and cocomplete poset. Now Theorem 6.2.1 tells us how to define the left Kan extension along $\mathbb{Q} \hookrightarrow \mathbb{R}$. The value of $\mathrm{Lan} 2^x$ is the supremum (the colimit) of all 2^q with $q \in \mathbb{Q}$ and $q \leq x$. This is the usual definition of 2^x, so we conclude that the exponential function $2^{-} \colon \mathbb{R} \to \mathbb{R}_{>0}$ is the left Kan extension of $2^{-} \colon \mathbb{Q} \to \mathbb{R}_{>0}$ along the inclusion $\mathbb{Q} \hookrightarrow \mathbb{R}$.

In this case, the exponential function $2^{-} \colon \mathbb{R} \to \mathbb{R}_{>0}$ is also the right Kan extension of $2^{-} \colon \mathbb{Q} \to \mathbb{R}_{>0}$ along the inclusion $\mathbb{Q} \hookrightarrow \mathbb{R}$; this is because 2^x is also the infimum of 2^q for all $q \in \mathbb{Q}$ with $x \leq q$. It won't commonly be the case that left and right Kan extensions agree.

EXAMPLE 6.2.5. Let us return to Example 6.1.5. In the category Vect_k, finite products and finite coproducts coincide: these are just direct sums of vector spaces. If V is an H-representation and H is a finite index subgroup of G, then the end and coend formulas of Theorem 6.2.1 and its dual both produce the direct sum of copies of V indexed by left cosets of H in G. Thus, for finite index subgroups, the left and right adjoints of (6.1.6) are the same; i.e., induction from a finite index subgroup is both left and right adjoint to restriction.

EXAMPLE 6.2.6. We can use Theorem 6.2.1 to understand the functors sk_n and cosk_n of Example 6.1.8. If $m > n$ and $k \leq n$, each map in $\mathbb{\Delta}^{\mathrm{op}}([k],[m]) = \mathbb{\Delta}([m],[k])$ factors uniquely as a non-identity epimorphism followed by a monomorphism.[2] It follows that every simplex in $\mathrm{sk}_n X$ above dimension m is degenerate; indeed $\mathrm{sk}_n X$ is obtained from the n-truncation of X by freely adding back the necessary degenerate simplices.

Now we use the adjunction $\mathrm{sk}_n \dashv \mathrm{cosk}_n$ to build some intuition for the n-coskeleton. Suppose $X \cong \mathrm{cosk}_n X$. By adjunction an $(n+1)$-simplex corresponds to a map $\mathrm{sk}_n \Delta^{n+1} = \partial\Delta^{n+1} \to X$. In words, each $(n+1)$-sphere in an n-coskeletal simplicial set has a unique filler. Indeed, any m-sphere in an n-coskeletal simplicial set, with $m > n$, has a unique filler. More precisely, an m-simplex is uniquely determined by the data of its faces of dimension n and below.

For any small category J, we can form a new category $\mathsf{J}^{\triangleright}$, the **cone** on J, via the following colimit diagram in Cat:

$$\begin{array}{ccc} \mathsf{J} & \xrightarrow{!} & \mathbb{1} \\ {\scriptstyle 1_{\mathsf{J}} \times 1} \downarrow & & \downarrow \\ \mathsf{J} \times \mathbb{2} & \longrightarrow & \mathsf{J}^{\triangleright} \end{array}$$

This is reminiscent of how cones are constructed on topological spaces, with the category $\mathbb{2} = 0 \to 1$ playing the role of the interval. The category $\mathsf{J}^{\triangleright}$ has one new object, a formally

[2]This is the content of the Eilenberg-Zilber lemma [**GZ67**, II.3.1, pp. 26-27].

adjoined terminal object t that serves as the nadir of a new cone under the inclusion $\mathsf{J} \hookrightarrow \mathsf{J}^{\triangleright}$. There are no additional new objects or morphisms.

The informal intuition that a colimit of a diagram should be the "closest" extension of the diagram to a cone under it is formalized by the following result.

Proposition 6.2.7. *A category* C *admits all colimits of diagrams indexed by a small category* J *if and only if the restriction functor* $\mathsf{C}^{\mathsf{J}^{\triangleright}} \to \mathsf{C}^{\mathsf{J}}$ *admits a left adjoint, defined by left Kan extension:*

$$\mathsf{C}^{\mathsf{J}^{\triangleright}} \underset{\text{res}}{\overset{\text{colim}}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathsf{C}^{\mathsf{J}}$$

Proof. By construction, there is a fully faithful inclusion $\mathsf{J} \hookrightarrow \mathsf{J}^{\triangleright}$. Consider a diagram $F \colon \mathsf{J} \to \mathsf{C}$. For $j \in \mathsf{J} \subset \mathsf{J}^{\triangleright}$, the colimit (6.2.2) defining the left Kan always exists: it is simply Fj. For the cone point $t \in \mathsf{J}^{\triangleright}$, the comma category $\mathsf{J} \downarrow t$ is isomorphic to J, by construction. So if $\mathrm{colim}(\mathsf{J} \xrightarrow{F} \mathsf{C})$ exists, then Theorem 6.2.1 tells us that these values define the left Kan extension functor, which we call $\mathrm{colim} \colon \mathsf{C}^{\mathsf{J}} \to \mathsf{C}^{\mathsf{J}^{\triangleright}}$ in this context. □

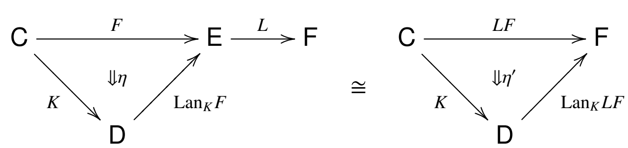
Exercises.

Exercise 6.2.1. Proposition 6.2.7 expresses a stronger universal property of the colimit cone than is usual stated. What is it?

Exercise 6.2.2. Directed graphs are functors from the category with two objects E, V and a pair of maps $s, t \colon E \rightrightarrows V$ to Set. A natural transformation between two such functors is a graph morphism. The forgetful functor $\mathsf{DirGraph} \to \mathsf{Set}$ that maps a graph to its set of vertices is given by restricting along the functor from the terminal category $\mathbb{1}$ that picks out the object V. Use Theorem 6.2.1 to compute left and right adjoints to this forgetful functor.

6.3. Pointwise Kan extensions

A functor $L \colon \mathsf{E} \to \mathsf{F}$ **preserves** $(\mathrm{Lan}_K F, \eta)$ if the whiskered composite $(L\mathrm{Lan}_K F, L\eta)$ is the left Kan extension of LF along K.



Example 6.3.1. The forgetful functor $U \colon \mathsf{Top} \to \mathsf{Set}$ has both left and right adjoints, and hence preserves both limits and colimits. It follows from Theorem 6.2.1 and that U preserves the left and right Kan extensions of Example 6.1.5.

Example 6.3.2. The forgetful functor $U \colon \mathsf{Vect}_k \to \mathsf{Set}$ preserves limits but not colimits because the underlying set of a direct sum is not simply the coproduct of the underlying sets of vectors. Hence, it follows from 6.2.1 and 6.1.5 that the underlying set of a G-representation induced from an H-representation is not equal to the G-set induced from the underlying H-set.

Even when we cannot appeal to the formula presented in 6.2.1:

Lemma 6.3.3. *Left adjoints preserve left Kan extensions.*

Proof. Suppose given a left Kan extension $(\mathrm{Lan}_K F, \eta)$ with codomain E and suppose further that $L \colon \mathsf{E} \to \mathsf{F}$ has a right adjoint R with unit ι and counit ν. Then given $H \colon \mathsf{D} \to \mathsf{F}$

there are natural isomorphisms

$$\mathsf{F}^{\mathsf{D}}(L\mathrm{Lan}_K F, H) \cong \mathsf{E}^{\mathsf{D}}(\mathrm{Lan}_K F, RH) \cong \mathsf{E}^{\mathsf{C}}(F, RHK) \cong \mathsf{F}^{\mathsf{C}}(LF, HK).$$

Taking $H = L\mathrm{Lan}_K F$, these isomorphisms act on the identity natural transformation as follows:

$$1_{L\mathrm{Lan}_K F} \mapsto \iota_{\mathrm{Lan}_K F} \mapsto \iota_{\mathrm{Lan}_K F\cdot K} \cdot \eta \mapsto \nu_{L\mathrm{Lan}_K F\cdot K} \cdot L\iota_{\mathrm{Lan}_K F\cdot K} \cdot L\eta = L\eta.$$

Hence $(L\mathrm{Lan}_K F, L\eta)$ is a left Kan extension of LF along K. □

Unusually for a mathematical object defined by a universal property, generic Kan extensions are rather poorly behaved. This is particularly visible in the context of total derived functors, as introduced in Example 6.1.3. The defining universal property of a total left derived functor as a right Kan extension is insufficient to prove, for instance, that the composite of two total left derived functors is the total left derived functor of the composite. The defining universal property is also insufficient to prove that the total left derived functor of a left adjoint is left adjoint to the total right derived functor of its right adjoint.[3]

If these examples are unfamiliar or unconvincing, we can also rely on expert opinion. For instance, Max Kelly reserves the name "Kan extension" for pairs satisfying the condition we will presently introduce, calling those of our Definition 6.1.1 "weak" and writing "Our present choice of nomenclature is based on our failure to find a single instance where a weak Kan extension plays any mathematical role whatsoever" [**Kel82**, §4]. By the categorical community's consensus, the important Kan extensions are **pointwise** Kan extensions.

Definition 6.3.4. When E is locally small, a right Kan extension is a **pointwise** right Kan extension[4] if it is preserved by all representable functors $\mathsf{E}(e, -)$.

Because covariant representables preserve all limits, it is clear that if a right Kan extension is given by the formula of Theorem 6.2.1, then that Kan extension is pointwise; dually, left Kan extensions computed in this way are pointwise. The surprise is that the converse also holds. This characterization justifies the terminology: a pointwise Kan extension can be computed pointwise as a limit in E.

Theorem 6.3.5. *A right Kan extension of F along K is pointwise if and only if it can be computed by*

$$\mathrm{Ran}_K F(d) = \lim\left(d/K \xrightarrow{\Pi^K_d} \mathsf{C} \xrightarrow{F} \mathsf{E}\right)$$

in which case, in particular, this limit exists.

Proof. If $\mathrm{Ran}_K F$ is pointwise, then by the Yoneda lemma and the defining universal property of right Kan extensions

$$\begin{aligned}\mathsf{E}(e, \mathrm{Ran}_K F(d)) &\cong \mathsf{Set}^{\mathsf{D}}(\mathsf{D}(d,-), \mathsf{E}(e, \mathrm{Ran}_K F)) \cong \mathsf{Set}^{\mathsf{C}}(\mathsf{D}(d, K-), \mathsf{E}(e, F-)) \\ &\cong \mathrm{Cone}(E, F\Pi^K_d),\end{aligned}$$

where the final natural isomorphism is the content of Lemma 6.3.6, whose elementary proof is left to the reader. Hence, this bijection exhibits $\mathrm{Ran}_K F(d)$ as the limit of FU. □

[3]For the derived functors one meets in practice, these properties do in fact hold, but for a more sophisticated reason. See [**Rie14**, §§2.1-2.2] for a complete discussion.

[4]A functor $K : \mathsf{C} \to \mathsf{D}$ is equally a functor $K : \mathsf{C}^{\mathrm{op}} \to \mathsf{D}^{\mathrm{op}}$ but the process of replacing each category by its opposite reverses the direction of any natural transformations; succinctly, "op" is a 2-functor $(-)^{\mathrm{op}} : \mathsf{Cat}^{\mathrm{co}} \to \mathsf{Cat}$. A left Kan extension is **pointwise**, as we are in the process of defining, if the corresponding right Kan extension in the image of this 2-functor is pointwise.

Lemma 6.3.6. *Given functors $F\colon \mathsf{C} \to \mathsf{E}$ and $K\colon \mathsf{C} \to \mathsf{D}$ and an object $d \in \mathsf{D}$, from which we define $\Pi^K_d\colon d\downarrow K \to \mathsf{C}$, there is a natural bijection*

$$\mathrm{Cone}(e, F\Pi^K_d) \cong \mathsf{Set}^{\mathsf{C}}(\mathsf{D}(d, K-), \mathsf{E}(e, F-)).$$

Most commonly, pointwise Kan extensions are found whenever the codomain category is cocomplete (for left Kan extensions) or complete (for right), but this is not the only case. Surprisingly, the most common construction of the total derived functors defined in 6.1.3 produces pointwise Kan extensions, even though homotopy categories have notoriously few limits and colimits [**Rie14**, 2.2.13].

Exercises.

Exercise 6.3.1. Prove Lemma 6.3.6.

6.4. All concepts

In this section we see that trivial special cases of Kan extensions can be used to define the other basic categorical concepts. This justifies Saunders Mac Lane's famous assertion that "The notion of Kan extensions subsumes all the other fundamental concepts of category theory" [**ML98a**, §X.7].

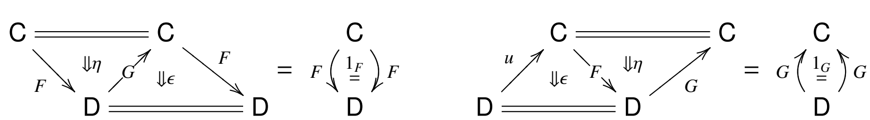
Proposition 6.4.1. *The Kan extension of $F\colon \mathsf{C} \to \mathsf{E}$ along the unique functor $!\colon \mathsf{C} \to \mathbb{1}$ defines the colimit of F in E, each existing if and only if the other does.*

Proof. A functor $G\colon \mathbb{1} \to \mathsf{E}$ picks out an object of E; precomposing with $!\colon \mathsf{C} \to \mathbb{1}$ yields the constant functor $\mathsf{C} \to \mathsf{E}$ at this object. Hence, the universal property (6.1.4) specifies that $\mathrm{Lan}_! F$ represents the set of natural transformations from $F\colon \mathsf{C} \to \mathsf{E}$ to a constant functor, i.e., that $\mathrm{Lan}_! F$ represents cones under F, i.e., that $\mathrm{Lan}_! F$ is the colimit of F. □

Dually $\mathrm{Ran}_! F$ is the limit.

Proposition 6.4.2. *If $F\colon \mathsf{C} \rightleftarrows \mathsf{D}\colon G$ is an adjunction with unit $\eta\colon 1 \Rightarrow GF$ and counit $\epsilon\colon FG \Rightarrow 1$, then (G, η) is a left Kan extension of the identity functor at C along F and (F, ϵ) is a right Kan extension of the identity functor at D along G. Conversely, if (G, η) is a left Kan extension of the identity along F and if F preserves this Kan extension, then $F \dashv G$ with unit η.*

Proof. Left as an exercise for the reader. The author finds a "pasting diagram" style proof, which combines the defining universal property of Definition 6.1.1 with the pasting diagram encoding of the triangle identities of an adjunction:



to be simpler than proving this result through a sequence of diagram chases. □

From the defining universal property, the right Kan extension of a functor F along the identity is (isomorphic to) F. It follows immediately from Definition 6.3.4 that this Kan extension is also pointwise, so by Theorem 6.3.5 we can apply the limit formula to conclude that

$$Fc \cong \lim(c\downarrow\mathsf{C} \xrightarrow{\Pi} \mathsf{C} \xrightarrow{F} \mathsf{E}),$$

in any category E. When E has products and equalizers, we can reexpress this limit formula by saying that there is an equalizer diagram

$$Fc \longrightarrow \prod_{c\to x} Fx \rightrightarrows \prod_{c\to x\to y} Fy$$

This formula is often considered to be a generalization of the Yoneda lemma.

We can recover the classical Yoneda lemma by restricting to the case of functors $F\colon \mathsf{C} \to \mathsf{Set}$. By the defining universal property of the limit, we have

$$\mathsf{Set}(*, Fc) \cong \mathrm{Cone}(*, F\Pi) \cong \mathsf{Set}^{\mathsf{C}}(\mathsf{C}(c,-), \mathsf{Set}(*, F-)) \cong \mathsf{Set}^{\mathsf{C}}(\mathsf{C}(c,-), F),$$

where the second isomorphism is a special case of Lemma 6.3.6. This is the Yoneda lemma.

Dually, any $F\colon \mathsf{C} \to \mathsf{E}$ is a pointwise left Kan extension of itself along 1_{C}. The colimit formula of Theorem 6.3.5 applies: in any category, we have

$$Fc \cong \mathrm{colim}(\mathsf{C}\!\downarrow\! c \xrightarrow{\Pi} \mathsf{C} \xrightarrow{F} \mathsf{E},$$

which is appropriately called the **coYoneda lemma**. When E has coproducts and coequalizers, this colimit can be expressed via the following coequalizer diagram.

$$\coprod_{y\to x\to c} Fy \rightrightarrows \coprod_{x\to c} Fx \longrightarrow Fc.$$

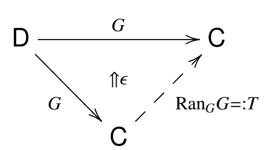
Again, we can conclude something special in the case of a functor $F\colon \mathsf{C} \to \mathsf{Set}$. For sets S and T, we have $\coprod_S T \cong S \times T \cong \coprod_T S$, so this coequalizer is isomorphic to the coequalizer

$$\coprod_{Fy\times\mathsf{C}(y,x)} \mathsf{C}(x,c) \rightrightarrows \coprod_{Fx} \mathsf{C}(x,c) \longrightarrow Fc.$$

where the indexing sets have been swapped. Letting c vary, we conclude that F is canonically a colimit of representable functors, a fact that is frequently called the **density theorem**. Reversing Theorem 3.3.7, the indexing category for this diagram of representables is one whose objects are elements of the set Fx for some $x \in \mathsf{C}$ and whose morphisms are maps $y \to x$ in C that send the chosen element of Fy to the corresponding element of Fx. Thus, the density theorem expresses F as a colimit of a diagram of representable of functors that is indexed by its category of elements.

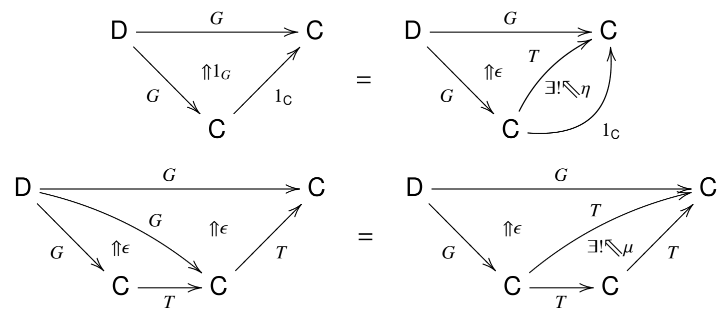
Finally, we turn to monads. We have seen that a right adjoint functor $G\colon \mathsf{D} \to \mathsf{C}$ induces a monad on C. More generally, assuming C has sufficient limits, any functor $G\colon \mathsf{D} \to \mathsf{C}$ induces a monad on C. If G has a left adjoint F it is the monad GF but this construction is more general.

Definition 6.4.3. The **codensity monad** of $G\colon \mathsf{D} \to \mathsf{C}$ is given by the right Kan extension of T along itself, whenever this exists.[5]



[5]In particular, "sufficient limits" in C means those necessary to define $\mathrm{Ran}_G G$ as a pointwise right Kan extension.

The unit and multiplication natural transformations are defined using the universal property of $\epsilon\colon TG \Rightarrow G$ as follows



EXAMPLE 6.4.4. Consider the inclusion $\mathsf{Ring} \hookrightarrow \mathsf{Field}$. We have argued that it has no left (or right) adjoint, but it does have a density monad T defined by

$$TA = \prod_{p \in \mathrm{Spec}(A)} \mathrm{Frac}(A/p)$$

where $\mathrm{Spec}(A)$ is the set of prime ideals $p \subset A$ and $\mathrm{Frac}(A/p)$ is the field of fractions of the quotient ring A/p, which is an integral domain because p is prime. The details for this example and considerably more can be found in [**Lei13**].

Exercises.

EXERCISE 6.4.1. Prove Proposition 6.4.2. As a hint, note an adjunction $F \dashv G$ induces an adjunction

$$\mathsf{E}^{\mathsf{C}} \underset{F^*}{\overset{G^*}{\underset{\perp}{\rightleftarrows}}} \mathsf{E}^{\mathsf{D}}$$

i.e., for any $H\colon \mathsf{C} \to \mathsf{E}$, $K\colon \mathsf{D} \to \mathsf{E}$, $\mathsf{E}^{\mathsf{D}}(HG, K) \cong \mathsf{E}^{\mathsf{C}}(H, KF)$.

EXERCISE 6.4.2. Use Theorem 6.2.1, the Yoneda lemma, and the coYoneda lemma to deduce another form of the density theorem: that the left Kan extension of the Yoneda embedding $\mathsf{C} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$ along itself is the identity functor. This says that the representable functors form a **dense subcategory** of the presheaf category $\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$.

EXERCISE 6.4.3. If $F \dashv G$ show that the codensity monad $\mathrm{Ran}_G G$ of G exists an is equal to the monad induced by the adjunction $F \dashv G$.

Epilogue: Theorems in category theory

The author is told with distressing regularity that “there are no theorems in category theory”—which typically means that the speaker does not know any theorems in category theory. What follows is a modest collection of counterexamples to this claim. Of course this list is by no means exhaustive. Sadly, the majority of the theorems that are personal favorites of the author were excluded because their significance is more difficult to explain.

E.1. Theorems in basic category theory

Significant theorems in basic category theory include, first and foremost, the Yoneda lemma:

THEOREM (2.2.3). *For any functor* $F\colon \mathsf{C} \to \mathsf{Set}$*, whose domain* C *is locally small, there is a bijection*

$$\mathrm{Hom}(\mathsf{C}(c,-),F) \cong Fc$$

that identifies a natural transformation $\alpha\colon \mathsf{C}(c,-) \Rightarrow F$ *with the element* $\alpha_c(1_c) \in Fc$*. Moreover this correspondence is natural in both* c *and* F*.*

A Set-valued functor is **representable** just when it is naturally isomorphic to a represented functor, covariant or contravariant as the case dictates. The idea that representable functors encode universal properties of their representing objects is made precise by the following result:

THEOREM (2.4.4). *A functor* $F\colon \mathsf{C} \to \mathsf{Set}$ *is representable if and only if its category of elements has an initial object. A functor* $F\colon \mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ *is representable if and only if its category of elements has a terminal object.*

To calculate particular limits and colimits, it is very useful to know that they can be built out of simpler limit and colimit constructions, when these exist.

THEOREM (3.3.7). *A locally small category* C *with coproducts and coequalizers has colimits of any shape. Dually a category with products and equalizers has all small limits.*

Fix a cardinal κ. Recall a category is κ-**small** if its set of morphisms is smaller than κ.

THEOREM (3.7.2). *Any* κ*-small category that admits all limits or all colimits of diagrams indexed by* κ*-small categories is equivalent to a poset. In particular, the only small categories that are complete or cocomplete are posets, up to equivalence.*

The following theorem is quite easy to prove, but this does not diminish its significance.

THEOREM (4.4.2, 4.4.3). *Right adjoints preserve limits, and left adjoints preserve colimits.*

This result gives a necessary condition for a functor to admit a left or right adjoint. The adjoint functor theorems supply additional conditions that are sufficient for an adjoint to exist.

THEOREM (4.5.3, 4.5.8). *Let $U: \mathsf{A} \to \mathsf{S}$ be a continuous functor whose domain is complete and whose domain and codomain are locally small. If either*

(i) U satisfies the ***solution set condition****: for every $s \in \mathsf{S}$ there exists a set of morphisms $f_i: s \to Ua_i$ so that any $f: s \to Ua$ factors through some f_i along a morphism $a_i \to a$ in A, or*

(ii) A has a small cogenerating set and every collection of subobjects of a fixed object in A admits an intersection,

then U admits a left adjoint.

The following theorem collects together several properties of monadic functors.

THEOREM (5.4.1, 5.5.6, 5.5.13). *If $U: \mathsf{A} \to \mathsf{S}$ admits a left adjoint and creates coequalizers of U-split pairs then:*

(i) A morphism f in A is an isomorphism if and only if Uf is an isomorphism in S.

(ii) U creates all limits that exist in S.

(iii) U creates all colimits that exist in S and are preserved by the monad on S that is induced from the adjunction.

(iv) If S is locally small, complete and cocomplete, and regular, then A admits all colimits.

The proofs of the adjoint functor theorems also lead to proofs of related representability criteria.

THEOREM (4.5.12,4.5.13). *If $F: \mathsf{C} \to \mathsf{Set}$ preserves limits and C is locally small and complete, then if either*

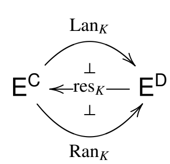
(i) C has a small cogenerating set and has the property that every collection of subobjects of a fixed object has an intersection, or if

(ii) F satisfies the solution set condition: there exists a set S of objects of C so that for any $c \in \mathsf{C}$ and any element $x \in Fc$ there exists an $s \in S$, an element $y \in Fs$, and a morphism $f: s \to c$ so that $Ff(y) = x$.

then F is representable.

The following theorem establishes the existence of and provides a formula for adjoints to restriction functors, given by pointwise Kan extension.

THEOREM (6.2.1). *Consider a functor $K: \mathsf{C} \to \mathsf{D}$ where C is small and D is locally small. Whenever E is cocomplete, then the restriction functor along K admits a left adjoint, defined by a particular colimit formula, and whenever E is complete, then the restriction functor along K admits a right adjoint, defined by a dual limit formula.*



Other candidates for inclusion on this list include theorems, such as 4.2.4 or 6.3.5, that prove that two a priori distinct categorical definitions are equivalent.

E.2. Coherence for monoidal categories

There are a number of coherence theorems in the categorical literature, the first and perhaps most famous being the following theorem of Mac Lane [**ML63**].

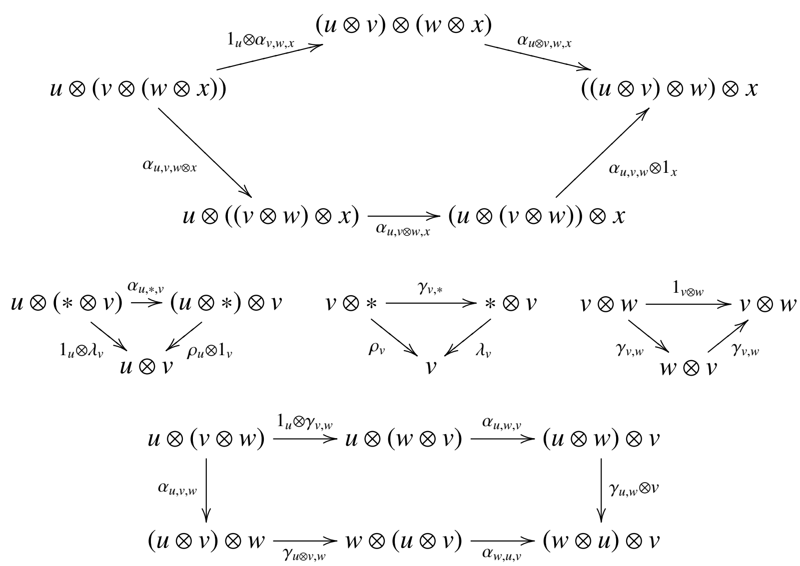
The data of a **symmetric monoidal category** $(\mathsf{V}, \otimes, *)$ consists of a category V, a bifunctor $- \otimes -\colon \mathsf{V} \otimes \mathsf{V} \to \mathsf{V}$ called the **monoidal product**, and a **unit object** $* \in \mathsf{V}$ together with specified natural isomorphisms

$$\text{(E.2.1)} \qquad v \otimes w \underset{\gamma}{\cong} w \otimes v \qquad u \otimes (v \otimes w) \underset{\alpha}{\cong} (u \otimes v) \otimes w \qquad * \otimes v \underset{\lambda}{\cong} v \underset{\rho}{\cong} v \otimes *$$

expressing symmetry, associativity, and unit conditions on the monoidal product. A **monoidal category** is defined analogously, except that the first symmetry natural isomorphism is omitted.

Examples include all categories with finite products, such as $(\mathsf{Set}, \times, *)$, $(\mathsf{Top}, \times, *)$, or $(\mathsf{Cat}, \times, \mathbb{1})$, where the monoidal unit in each case is the terminal object. The category of modules over a ring admits a number of monoidal structures: $(\mathsf{Mod}_R, \times, *)$, $(\mathsf{Mod}_R, \otimes_{\mathbb{Z}}, \mathbb{Z})$, and $(\mathsf{Mod}_R, \otimes_R, R)$, the last only in the case where R is commutative. The category of unbounded chain complexes of R-modules, with R commutative, is also a monoidal category $(\mathsf{Ch}_R, \otimes_R, R)$, where "$\otimes_R$" denotes the usual graded tensor product of chain complexes and R is the chain complex with the R-module R in degree zero and zeros elsewhere. In this case, the symmetry natural isomorphism $\otimes_R$, extending this bifunctor on Mod_R, can be defined in multiple ways depending on whether or not a sign is introduced in odd degrees.

The natural isomorphisms (E.2.1) defining a symmetric monoidal category are required to satisfy the following coherence diagrams, which we express in terms of components at objects $u, v, w, x \in \mathsf{V}$:



The coherence theorem for monoidal categories, first proven by Mac Lane and then improved by Max Kelly, tells us that once these conditions are satisfied, then all diagrams whose edges are comprised of composites of these natural isomorphisms, parenthesized in any order, commute; see [**ML98a**] for a precise statement.

Theorem E.2.2 (Mac Lane, Kelly). *Any diagram in a symmetric monoidal category that is comprised of associators α, unitors γ or ρ, or symmetors γ that is "formal" commutes.*

As Mac Lane writes in the introduction to [**ML63**]:

> The usual associative law $a(bc) = (ab)c$ is known to imply the "general associative law," which states that any two iterated products of

the same factors in the same order are equal, irrespective of the arrangement of parentheses.

The upshot of Theorem E.2.2 is that any two functors $\mathsf{V}^{\times n} \to \mathsf{V}$ that are built iteratively from the bifunctor $\otimes$ and the unit object $*$, permuting and parenthesizing the inputs in some manner, are connected by a unique natural isomorphism that is built out of the given natural isomorphism α, λ, ρ, and γ. In practice, this naturality means that we need not concern ourselves with particular parenthesizations or orderings when defining the n-ary tensor product, and, hence, the structural isomorphisms (E.2.1) are seldom emphasized.

E.3. The universal property of the unit interval

The **unit interval** is the topological space $I = [0,1] \subset \mathbb{R}$ regarded as a subspace of the real line, with the standard Euclidean metric topology. It is used to define the **fundamental groupoid** $\Pi_1(X)$ of paths in a topological space X. A **path** is simply a continuous function $p\colon I \to X$. The path has two endpoints $p(0), p(1) \in X$ which are defined by evaluating at the endpoints $0, 1 \in I$. If $q\colon I \to X$ is a second path with the property that $p(1) = q(0)$, then there exists a composite path $p * q\colon I \to X$ defined by the composite continuous function

$$I \xrightarrow[\cong]{\delta} I \vee I \xrightarrow{p \vee q} X.$$

Here $I \vee I$ is the space formed by gluing two copies of I together by identifying the point 1 in the left-hand copy with the point 0 in the right-hand copy; we might think of $I \vee I$ as the space $[0,2] \subset \mathbb{R}$. The map $\delta\colon I \to I \vee I$ is the homeomorphism $t \mapsto 2t$. Note that this map sends the endpoints of the domain I to the endpoints of the fattened interval $I \vee I$. Thus $(p * q)(0) = p(0)$ and $(p * q)(1) = q(1)$; that is, $p * q$ is a path in X from the starting point of the path p to the ending point of the path q.

Now the fundamental groupoid $\Pi_1(X)$ is the category whose objects are points of X and whose morphisms are endpoint-preserving homotopy classes of paths in X, with composition defined using the structure on the unit interval just described.

By a theorem of Peter Freyd, refined by Tom Leinster, the structure on the unit interval that is used to define composition of paths also describes the universal property of this topological space. By a **bipointed space** we mean a topological space I with two distinct closed points $0 \neq 1 \in I$, one designated as the "left" and the other as the "right." A map of bipointed spaces is a continuous function that sends the left point to the left point and the right point to the right point. For any bipointed space (X, x_0, x_1), we can define $X \vee X$ to be the pushout

$$\begin{array}{ccc} * & \xrightarrow{x_0} & X \\ {\scriptstyle x_1}\downarrow & & \downarrow \\ X & \longrightarrow & X \vee X \end{array}$$

Theorem E.3.1 (Freyd-Leinster). *The unit interval is the initial bipointed space equipped with a map of bipointed spaces $X \to X \vee X$. That is, for any other bipointed space (X, x_0, x_1), there is a unique bipointed continuous map $p\colon I \to X$ so that*

$$\begin{array}{ccc} I & \xrightarrow[\cong]{\delta} & I \vee I \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle p \vee p} \\ X & \longrightarrow & X \vee X \end{array}$$

commutes.

Put more concisely, Theorem E.3.1 says that (I, i_0, i_1) is the **initial algebra** for the endofunctor $X \mapsto X \vee X$ on the category of bipointed spaces. See [**Lei11**] for a discussion and proof.

E.4. A characterization of Grothendieck toposes

Given a small category C, a **presheaf** is simply a contravariant Set-valued functor on C. A **Grothendieck topos** is a reflective full subcategory E of a presheaf category

$$\mathsf{E} \underset{\hookrightarrow}{\overset{L}{\underset{\perp}{\leftarrow}}} \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

with the property left adjoint preserves finite limits. Objects in E are **sheaves** on a small **site**. A typical example might take $\mathsf{C} = O(X)$ to be the poset of open sets for a topological space X. A presheaf $P\colon O(X)^{\mathrm{op}} \to \mathsf{Set}$ assigns a set $P(U)$ to each open set $U \subset X$ so that this assignment is functorial with respect to restrictions along inclusion $V \subset U \subset X$ of open subsets. A presheaf P is a **sheaf** if and only if the diagram of restriction maps

$$P(U) \longrightarrow \prod_\alpha P(U_\alpha) \rightrightarrows \prod_{\alpha,\beta} P(U_\alpha \cap U_\beta)$$

is an equalizer for every open cover $U = \cup_\alpha U_\alpha$ of an open subset $U \subset X$.

Giraud's theorem states that Grothendieck toposes can be completely characterized by a combination of "exactness" and size conditions.

Theorem E.4.1 (Giraud). *A category E is a Grothendieck topos if and only if it satisfies the following conditions:*

- *(i) E is locally small.*
- *(ii) E has finite limits.*
- *(iii) E has all small coproducts and they are disjoint and universal*
- *(iv) Equivalence relations in E have universal coequalizers.*
- *(v) Every equivalence relation in E is effective, and every epimorphism in E is a coequalizer.*
- *(vi) E has a set of generators.*

Some explanation is in order. A coproduct $\coprod_\alpha A_\alpha$ is **disjoint** if each inclusion $A_\alpha \to \coprod_\alpha A_\alpha$ is a monomorphism and if the pullback of any two distinct inclusions is an initial object. A colimit cone $(A_\alpha \to A)_\alpha$ is **universal** if the pullbacks of these maps along any $f\colon B \to A$ defines a colimit cone $(B \times_A A_\alpha \to A)_\alpha$. An effective equivalence relation is one that arises, as in Example 3.4.5 as a kernel pair of some morphism. Finally, a set of **generators** for E is a set of objects $\mathcal{G} \subset \mathrm{ob}\,\mathsf{E}$ that is jointly separating: for any $f, g\colon B \rightrightarrows A$ with $f \neq g$ there is some $h\colon G \to B$, with $G \in \mathcal{G}$, so that $fh \neq gh$.

A proof of Theorem E.4.1 can be found in [**Joh14**, §0.4], which also presents a very clear exposition of the ideas involved.

E.5. Embeddings of abelian categories

Properties of the category Ab of abelian groups or Mod_R of modules over a unital ring are abstracted in the notion of an **abelian category**. Abelian category were introduced by Buchsbaum [**Buc55**] and developed by Grothendieck [**Gro57**] with the aim of unifying the "coefficients" of the various cohomology theories then under development.

Definition E.5.1. A category A is **abelian** if

- it has a **zero object** $0 \in \mathsf{A}$, that is both initial and terminal,
- it has all binary products and binary coproducts,

- it has all **kernels** and **cokernels**, defined respectively to be equalizers and coequalizers of a map $f\colon A \to B$ with the zero map $A \to 0 \to B$,
- all monomorphisms and epimorphisms arise as kernels or cokernels, respectively.

In an abelian category, finite products and finite coproducts coincide; these are often called **biproducts** or **direct sums**. These axioms imply that the hom-sets in A canonically inherit the structure of an abelian group, with the zero map in each hom-set serving as the identity, and moreover that composition is bilinear. This gives A the structure of a category **enriched over** the monoidal category $(\mathsf{Ab}, \otimes_{\mathbb{Z}}, \mathbb{Z})$.

The **image** of a morphism is defined to be the kernel of its cokernel, or equivalently, the cokernel of its kernel; these objects are always isomorphic. This permits the definition of an **exact sequence** in A, a sequence of composable morphisms

$$\cdots \xrightarrow{f_{n+2}} A_{n+1} \xrightarrow{f_{n+1}} A_n \xrightarrow{f_n} A_{n-1} \xrightarrow{f_{n-1}} \cdots$$

so that $\ker f_n = \mathrm{im} f_{n+1}$. A functor is **exact** if it preserves exact sequences.

THEOREM E.5.2 (Freyd-Mitchell). *If A is a small abelian category, then there is a ring*[6] *R and an exact, fully faithful functor $\mathsf{A} \hookrightarrow \mathsf{Mod}_R$, which embeds A as a full subcategory.*

A proof appears as the very last result of the book [**Fre03**]. Unfortunately, a plurality of the material presented in the preceding 150 pages appears to be necessary. No substantial simplification of this argument is known.

To conclude, we invite the reader to reflect upon the following lovely quote from the introduction to P.T. Johnstone's *Topos Theory* [**Joh14**, pp. xii-xiii] which describes the significance of the Freyd-Mitchell embedding theorem and, by analogy, the broader project of category theory:

> Incidentally, the Freyd-Mitchell embedding theorem is frequently regarded as a culmination rather than a starting point; this is because of what seems to me a misinterpretation (or at least an inversion) of its true significance. It is commonly thought of as saying "If you want to prove something about an abelian category, you might as well assume it is a category of modules"; whereas I believe its true import is "If you want to prove something about categories of modules, you might as well work in a general abelian category"—for the embedding theorem ensures that your result will be true in this generality, and by forgetting the explicit structure of module categories you will be forced to concentrate on the essential aspects of this problem. As an example, compare the module-theoretic proof of the Snake Lemma in [**HS97**] with the abelian-category proof in [**ML98a**].

[6] As per our conventions elsewhere, R is unital and associative but not necessarily commutative.

Bibliography

[Awo96] S Awodey. Structure in mathematics and logic: a categorical perspective. *Philos. Math. (3)*, 4(3):209–237, 1996.

[Awo10] Steve Awodey. *Category theory*, volume 52 of *Oxford Logic Guides*. Oxford University Press, Oxford, second edition, 2010.

[Bae06] John Baez. Quantum quandaries: a category-theoretic perspective. In *The structural foundations of quantum gravity*, pages 240–265. Oxford Univ. Press, Oxford, 2006.

[Bor94] Francis Borceux. *Handbook of categorical algebra. 2*, volume 51 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1994.

[Buc55] D. A. Buchsbaum. Exact categories and duality. *Trans. Amer. Math. Soc.*, 80:1–34, 1955.

[EM42a] Samuel Eilenberg and Saunders MacLane. Group extensions and homology. *Ann. of Math. (2)*, 43(757–831), 1942.

[EM42b] Samuel Eilenberg and Saunders MacLane. Natural isomorphisms in group theory. *Proc. Nat. Acad. Sci. U. S. A.*, 28:537–543, 1942.

[EM45] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Trans. Amer. Math. Soc.*, 58:231–294, 1945.

[ES52] Samuel Eilenberg and Norman Steenrod. *Foundations of algebraic topology*. Princeton University Press, Princeton, New Jersey, 1952.

[Fre03] Peter J. Freyd. Abelian categories. *Repr. Theory Appl. Categ.*, (3):1–190, 2003.

[Fre04] Peter Freyd. Homotopy is not concrete. *Reprints in Theory and Applications of Categories*, (6):1–10, 2004.

[FS90] P.J. Freyd and A. Scedrov. *Categories, Allegories*. North-Holland Mathematical Library (Book 39). North Holland, 1990.

[Gro57] Alexander Grothendieck. Sur quelques points d'algèbre homologique. *Tôhoku Math. J. (2)*, 9:119–221, 1957.

[Gro58] Alexander Grothendieck. A general theory of fibre spaces with structure sheaf. National Science Foundation Research Project on Geometry of Function Space, Report No. 4, Second Edition, May 1958.

[Gro60] Alexander Grothendieck. Technique de descente et théorèmes d'existence en géométrie algébriques. ii. le théorème d'existence en théorie formelle des modules. *Séminaire Bourbaki*, 5(195):195, 1958-1960.

[GZ67] P. Gabriel and M. Zisman. *Calculus of fractions and homotopy theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 35. Springer-Verlag New York, Inc., New York, 1967.

[HS97] P. J. Hilton and U. Stammbach. *A course in homological algebra*, volume 4 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[Joh14] P. T. Johnstone. *Topos theory*. reprint of London Mathematical Society Monographs, Vol. 10. Dover Publications, 2014.

[Joy81] André Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42(1):1–82, 1981.

[Kan58] Daniel M. Kan. Adjoint functors. *Trans. Amer. Math. Soc.*, 87:294–329, 1958.

[Kel82] G. M. Kelly. *Basic concepts of enriched category theory*. Reprints in Theory and Applications of Categories. 1982. Reprint of the 1982 original [Cambridge Univ. Press, Cambridge; MR0651714].

[Lan84] Serge Lang. *Algebra*. Addison-Wesley Publishing Company, Advanced Book Program, Reading, MA, second edition, 1984.

[Lan02] Serge Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002.

[Lei11] Tom Leinster. A general theory of self-similarity. *Adv. Math.*, 226(4):2935–3017, 2011.

[Lei13] Tom Leinster. Codensity and the ultrafilter monad. *Theory Appl. Categ.*, 28:No. 13, 332–370, 2013.

[ML63] Saunders Mac Lane. Natural associativity and commutativity. *Rice Univ. Studies*, 49(4):28–46, 1963.

[ML88] Saunders Mac Lane. Concepts and categories in perspective. In *A century of mathematics in America, Part I*, volume 1 of *Hist. Math.*, pages 323–365. Amer. Math. Soc., Providence, RI, 1988.

[ML98a] S. Mac Lane. *Categories for the working mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.

[ML98b] Saunders Mac Lane. The yoneda lemma. *Mathematica Japonica*, 47:156, 1998.

[ML05] Saunders Mac Lane. *Saunders Mac Lane—a mathematical autobiography*. A K Peters, Ltd., Wellesley, MA, 2005.

[Poo14] Bjorn Poonen. Why all rings should have a 1. arXiv:1404.0135 [math.RA], 2014.

[Rei88] Miles Reid. *Undergraduate algebraic geometry*, volume 12 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1988.

[Rie14] Emily Riehl. *Categorical homotopy theory*, volume 24 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2014.

[Seg74] Graeme Segal. Categories and cohomology theories. *Topology*, 13:293–312, 1974.

[Shu08] Michael Shulman. Set theory for category theory. arXiv:0810.1279, 2008.

[Sim11] Harold Simmons. *An introduction to category theory*. Cambridge University Press, Cambridge, 2011.

[Smi] Peter Smith. The galois connection between syntax and semantics. http://www.logicmatters.net/resources/pdfs/Galois.pdf.

Glossary of Notation

Δ, 57
C^T, 106
$\cdot$, 6
$\dashv$, 83
$\int F$, 51
$\lrcorner$, 60
$\mapsto$, 2
$\rightrightarrows$, 2
$\rightsquigarrow$, 2
$\simeq$, 24
$\to$, 2
$\ulcorner$, 62

$\mathbb{A}^n$, 49

$\mathsf{C}(X, Y)$, 9
C/c, 52
c/C, 52
cosk_n, 127
$\mathsf{C} \times \mathsf{D}$, 17

$\mathbb{\Delta}$, 127
$\mathbb{\Delta}_{\leq n}$, 127
$\mathbb{\Delta}_{+}$, 127
D^n, 57

$\mathrm{hom}(X, Y)$, 9

$\mathbb{1}$, 71

$\mathrm{Lan}_K F$, 125

O_G, 18

$\mathrm{Ran}_K F$, 125

sk_n, 127
skC, 28
S^n, 57
$\mathrm{Sym}(G)$, 47

Vect_k^G, 127
$\mathsf{Vect}_{\Bbbk}^{\mathrm{fd}}$, 19

Index

abelian category, 139
abelianization, 94
adjoint
  mutually left or right, 85
adjoint functor, 82
adjunction, 81, 87
  as Kan extension, 132
  counit, 87
  monadic, 84
  morphism, 89
  triangle identity, 87
  two-variable, 86
  unit, 86
affine space, 49, 105
algebra, 31
  category of, 106
  free, 107
algebraic theory
  model of, 117
apex, 58
arrow, 8
automorphism group, 19
axiom
  in a first-order language, 85

based object, 52
basepoint, 7
bifunctor, 17
bilinear, 49

categorification, 23
category, 6
  abelian, 139
  cardinality of, 76
  comma, 55
  complete, 63
  concrete, 7, 36
  connected, 27
  discrete, 8
  Eilenberg-Moore, 106
  essentially small, 29
  free, 84
  indexing, 31
  isomorphism of, 17
  Kleisli, 108
  locally small, 9
  monoidal, 137
  of algebras, 106
  of categories, 17
  of elements, 51, 71
  opposite, 10
  product of, 17
  regular, 73, 121
  skeletal, 28
  slice, 52
  small, 9
  symmetric monoidal, 136
Cayley's theorem, 47
cell complex, 57
center, 19
chain rule, 14
coalgebra, 108
coarse, 59
cocone, 58
codomain, 6
coequalizer, 61
  split, 113
cokernel, 62
colimit, 57, 58
  absolute, 113
  as Kan extension, 132
comma category, 55
commutative
  cube, 33
  rectangle, 32
  square, 32
commutative square, 20
commutator subgroup, 19, 94
commute, 30
complete, 63
complete lattice, 77
composite morphism, 6
concrete category, 36
cone, 58
  legs of, 58
  on a category, 129
  summit of, 58
continuous
  functor, 96
contractible, 53
coproduct, 61
cotensor, 74

counit
  of an adjunction, 87
coYoneda lemma, 133
create, 67
currying, 41

decategorification, 23
dense subcategory, 134
density theorem, 133
derivative
  total, 14
derived functor
  total, 126
diagram, 30, 31, 57, 85
  commutative, 30, 31
  shape of, 32, 57
  small, 76
direct limit, 61
directed graph, 130
discrete, 8
discrete dynamical system, 53
domain, 6
dual, 11
dual basis, 19
dual vector space, 15
  double, 19
duality, 131

Eilenberg-Moore category, 106
Eilenberg-Zilber
  lemma, 129
elements
  category of, 51
embedding, 25
  full, 25
endofunctor, 13
epic, 36
epimorphism, 36
  split, 37
equalizer, 59
equivalence relation, 72, 73
equivariant, 21
essential image, 29, 95
essentially small, 29
essentially surjective, 25
evil, 29
extension of scalars, 127

faithful, 25
fiber, 61
fiber product, 60
fiber space, 73
fine, 59
fixed point, 63
free
  algebra, 107
free action, 43, 48
free group, 13
free product, 120
full, 25
  embedding, 25
function
  partial, 18
functor, 13
  adjoint, 82
  bi-, 17
  constant, 57
  continuous, 96
  contravariant, 14
  covariant, 14
  faithful, 25
  forgetful, 13
  full, 25
  morphism of, 18
  representable, 39
  represented, 16
  two-sided represented, 17, 39
functoriality, 13, 15
fundamental group, 13, 27
fundamental groupoid, 10, 27, 138

G-equivariant, 21
G-object, 127
G-representation, 16
G-set, 16
Galois connection, 83
Galois extension, 17
Galois group, 18
Galois theory
  fundamental theorem of, 18
generating set, 98
generator, 98
generators, 112
global, 25
gluing, 57
Grothendieck topos, 73, 139
group
  free, 40, 88
  opposite, 11
  presentation, 112
  torsion sub-, 22
group action, 16
  free, 43, 48
  transitive, 48
group extension, 5
groupoid, 10
  contractible, 53
  maximal, 10
  maximal sub-, 10

hom-set, 9
horizontal composition, 34

idempotent, 65
  split, 66
identity morphism, 6
image, 121
  essential, 95

image factorization, 73
induced representation, 127, 129, 130
infimum, 13, 70
initial
  jointly, 97
  weakly, 97
initial object, 35, 61
intersection, 99
inverse
  left, 9, 37
  right, 9, 37
inverse limit, 61
isomorphic
  representably, 48
isomorphism, 9
  natural, 20

Jacobian matrix, 14

Kan extension, 125–126
  pointwise, 131
  preservation of, 130
kernel pair, 72
Kleisli category, 108

lattice
  complete, 77
left inverse, 9
limit, 57, 58
  as Kan extension, 132
  of a sequence, 79
linear functional, 15
local, 25
local object, 95
localization, 95
locally small, 9
loop, 41

map, 8
middle four interchange, 38
model, 85
monad, 101
  codensity, 133
  double dual, 103
  free R-module, 103
  free abelian group, 103
  free commutative monoid, 104
  free group, 103
  free monoid, 102
  free vector space, 103
  idempotent, 105
  list, 102
  maybe, 102
monadicity theorem, 112
monic, 36
monoid, 8, 10, 30, 108
  topological, 30
monoidal
  product, 136
monoidal category, 137
monomorphism, 36
  split, 37
morphism
  composable pair, 6
  composite, 6
  identity, 6
  parallel pair, 2

n-(co)skeleton, 127, 129
natural isomorphism, 20
natural transformation, 20
  components of, 20
  constant, 57
no cloning theorem, 22

object
  local, 95
opposite
  group, 11
orbit, 28
orbit category, 18
orbit-stabilizer theorem, 29

p-adic integers, 61
parallel pair, 2, 59
pasting diagram, 125
path, 41, 138
pointed set, 107
poset, 8
power, 74
preorder, 8
presentation
  for a group, 112
preserve, 67
product, 57, 59
  fiber, 60
pullback, 60
pushout, 61

quotient, 57

recursion, 54
reflect, 67
reflective subcategory, 94
regular category, 73, 121
relation, 72
representable
  vs represented, 39
representable functor, 39
representation, 39
represented functor, 16
  two-sided, 17, 39
restriction of scalars, 84
retraction, 9, 37
right inverse, 9
ring, 31

section, 9, 37
Sierpinski space, 41

simplicial set, 127
augmented, 127
skeleton, 28
slice category, 52
small, 9, 76
essentially, 29
locally, 9
solution set, 97, 136
space
fiber, 73
species, 75
split
epimorphism, 37
monomorphism, 37
split coequalizer, 113
stabilizer, 28
Stone-Čech compactification, 94
subcategory, 10
subobject, 57, 99
subobject classifier, 54
summit, 58
supremum, 12, 70
symmetric monoidal category, 136

tensor product, 49
terminal object, 35, 59
the, 48
torsion subgroup, 22
torsor, 49
torus, 62
transitive action, 48
translation groupoid, 28, 53
triangle identity, 87

ultrafilter, 103
union, 62
unit
of an adjunction, 86
unit interval, 138
unit object, 136
universal object, 53
universal property, 48

vector space
dual, 15
vertical composition, 34

walking
isomorphism, 71
whiskering, 35

Yoneda embedding, 47
density of, 134
Yoneda lemma, 44, 133, 135
corollary, 47