Category Theory has developed rapidly. This book aims to present those ideas and methods which can now be effectively used by Mathematicians working in a variety of other fields of Mathematical research. This occurs at several levels. On the first level, categories provide a convenient conceptual language, based on the notions of category, functor, natural transformation, contravariance, and functor category. These notions are presented, with appropriate examples, in Chapters I and II. Next comes the fundamental idea of an adjoint pair of functors. This appears in many substantially equivalent forms: That of universal construction, that of direct and inverse limit, and that of pairs of functors with a natural isomorphism between corresponding sets of arrows. All these forms, with their interrelations, are examined in Chapters III to V. The slogan is "Adjoint functors arise everywhere".

Alternatively, the fundamental notion of category theory is that of a monoid – a set with a binary operation of multiplication which is associative and which has a unit; a category itself can be regarded as a sort of generalized monoid. Chapters VI and VII explore this notion and its generalizations. Its close connection to pairs of adjoint functors illuminates the ideas of universal algebra and culminates in Beck's theorem characterizing categories of algebras; on the other hand, categories with a monoidal structure (given by a tensor product) lead inter alia to the study of more convenient categories of topological spaces.

Since a category consists of arrows, our subject could also be described as learning how to live without elements, using arrows instead. This line of thought, present from the start, comes to a focus in Chapter VIII, which covers the elementary theory of abelian categories and the means to prove all the diagram lemmas without ever chasing an element around a diagram.

Finally, the basic notions of category theory are assembled in the last two chapters: More exigent properties of limits, especially of filtered limits, a calculus of "ends", and the notion of Kan extensions. This is the deeper form of the basic constructions of adjoints. We end with the observations that all concepts of category theory are Kan extensions (§ 7 of Chapter X).
I have had many opportunities to lecture on the materials of these chapters: At Chicago; at Boulder, in a series of Colloquium lectures to the American Mathematical Society; at St. Andrews, thanks to the Edinburgh Mathematical Society; at Zurich, thanks to Beno Eckmann and the Forschungsinstitut für Mathematik; at London, thanks to A. Fröhlich and Kings and Queens Colleges; at Heidelberg, thanks to H. Seifert and Albrecht Dold; at Canberra, thanks to Neumann, Neumann, and a Fulbright grant; at Bowdoin, thanks to Dan Christie and the National Science Foundation; at Tulane, thanks to Paul Mostert and the Ford Foundation, and again at Chicago, thanks ultimately to Robert Maynard Hutchins and Marshall Harvey Stone.

Many colleagues have helped my studies. I have profited much from a succession of visitors to Chicago (made possible by effective support from the Air Force Office of Scientific Research, the Office of Naval Research, and the National Science Foundation): M. André, J. Bénabou, E. Dubuc, F. W. Lawvere, and F. E. J. Linton. I have had good counsel from Michael Barr, John Gray, Myles Tierney, and Fritz Ulmer, and sage advice from Brian Abrahamson, Ronald Brown, W. H. Cockcroft, and Paul Halmos. Daniel Feigin and Geoffrey Phillips both managed to bring some of my lectures into effective written form. My old friend, A. H. Clifford, and others at Tulane were of great assistance. John MacDonald and Ross Street gave pertinent advice on several chapters; Spencer Dickson, S. A. Huq, and Miguel La Plaza gave a critical reading of other material. Peter May's trenchant advice vitally improved the emphasis and arrangement, and Max Kelly's eagle eye caught many soft spots in the final manuscript. I am grateful to Dorothy Mac Lane and Tere Shuman for typing, to Dorothy Mac Lane for preparing the index and to M. K. Kwong for careful proof reading — but the errors which remain, and the choice of emphasis and arrangement, are mine.

Dune Acres, March 27, 1971
Saunders Mac Lane

Table of Contents

Introduction ........................................... 1

I. Categories, Functors and Natural Transformations .... 7
   1. Axioms for Categories .......................... 7
   2. Categories .................................... 10
   3. Functors ..................................... 13
   4. Natural Transformations ...................... 16
   5. Monics, Epis, and Zeros ...................... 19
   6. Foundations .................................. 21
   7. Large Categories ................................ 24
   8. Horn-sets .................................... 27

II. Constructions on Categories .......................... 31
   1. Duality ....................................... 31
   2. Contravariance and Opposites .................. 33
   3. Products of Categories ....................... 36
   4. Functor Categories ............................ 40
   5. The Category of All Categories ............... 42
   6. Comma Categories ............................. 46
   7. Graphs and Free Categories ................... 48
   8. Quotient Categories ............................ 51

III. Universals and Limits ................................ 55
   1. Universal Arrows ................................ 55
   2. The Yoneda Lemma ............................. 59
   3. Coproducts and Colimits ...................... 62
   4. Products and Limits ............................ 68
   5. Categories with Finite Products ............... 72
   6. Groups in Categories ........................... 75
### Table of Contents

#### VIII. Adoints
1. Adjunctions .................................. 77
2. Examples of Adoints ....................... 77
3. Reflective Subcategories .................. 88
4. Equivalence of Categories ................. 90
5. Adajoints for Preorders ................... 93
6. Cartesian Closed Categories .............. 95
7. Transformations of Adajoints ............. 97
8. Composition of Adajoints ................. 101

#### V. Limits
1. Creation of Limits .......................... 105
2. Limits by Products and Equalizers ...... 108
3. Limits with Parameters .................... 111
4. Preservation of Limits ..................... 112
5. Adajoints on Limits ....................... 114
6. Freyd's Adjoint Functor Theorem ......... 116
7. Subobjects and Generators ............... 122
8. The Special Adjoint Functor Theorem .... 124
9. Adajoints in Topology ..................... 128

#### VI. Monads and Algebras
1. Monads in a Category ...................... 133
2. Algebras for a Monad ...................... 135
3. The Comparison with Algebras .......... 138
4. Words and Free Semigroups ................ 140
5. Free Algebras for a Monad ............... 143
6. Split Coequalizers ......................... 145
7. Beck's Theorem ............................ 147
8. Algebras are T-algebras ................... 152
9. Compact Hausdorff Spaces ................ 153

#### VII. Monoids
1. Monoidal Categories ....................... 157
2. Coherence ................................. 161
3. Monoids .................................. 166
4. Actions ................................... 170
5. The Simplicial Category ................... 171
6. Monads and Homology ..................... 176
7. Closed Categories ........................ 180
8. Compactly Generated Spaces ............. 181
9. Loops and Suspensions ................... 184

#### VIII. Abelian Categories
1. Kernels and Cokernels ..................... 187
2. Additive Categories ....................... 187
3. Abelian Categories ....................... 190
4. Diagram Lemmas ........................... 198

#### IX. Special Limits
1. Filtered Limits ............................ 207
2. Interchange of Limits ..................... 210
3. Final Functors .............................. 213
4. Diagonal Naturality ....................... 214
5. Ends ...................................... 218
6. Coends .................................... 222
7. Ends with Parameters ..................... 224
8. Iterated Ends and Limits ................. 226

#### X. Kan Extensions
1. Adjoint and Limits ......................... 229
2. Weak Universality ......................... 231
3. The Kan Extension ......................... 232
4. Kan Extensions as Coends ................. 236
5. Pointwise Kan Extensions ................. 239
6. Density ................................... 241
7. All Concepts are Kan Extensions ........ 244

#### Table of Terminology

#### Bibliography

#### Index
Introduction

Category theory starts with the observation that many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows. Each arrow \( f: X \to Y \) represents a function; that is, a set \( X \), a set \( Y \), and a rule \( x \mapsto f(x) \) which assigns to each element \( x \in X \) an element \( f(x) \in Y \); whenever possible we write \( f(x) \) and not \( f(x) \), omitting unnecessary parentheses. A typical diagram of sets and functions is

\[
\begin{array}{ccc}
W & \xrightarrow{f} & X \\
& \swarrow_f \downarrow & \searrow_{g} \downarrow \\
X \times Y & \xrightarrow{h} & Y;
\end{array}
\]

it is commutative when \( h = g \circ f \), where \( g \circ f \) is the usual composite function \( g \circ f : X \to Z \), defined by \( x \mapsto g(f(x)) \). The same diagrams apply in other mathematical contexts; thus in the “category” of all topological spaces, the letters \( X \), \( Y \), and \( Z \) represent topological spaces while \( f \), \( g \), and \( h \) stand for continuous maps. Again, in the “category” of all groups, \( X \), \( Y \), and \( Z \) stand for groups, \( f \), \( g \), and \( h \) for homomorphisms.

Many properties of mathematical constructions may be represented by universal properties of diagrams. Consider the cartesian product \( X \times Y \) of two sets, consisting as usual of all ordered pairs \( \langle x, y \rangle \) of elements \( x \in X \) and \( y \in Y \). The projections \( \langle x, y \rangle \mapsto x \), \( \langle x, y \rangle \mapsto y \) of the product on its “axes” \( X \) and \( Y \) are functions \( p : X \times Y \to X \), \( q : X \times Y \to Y \). Any function \( h : W \to X \times Y \) from a third set \( W \) is uniquely determined by its composites \( p \circ h \) and \( q \circ h \). Conversely, given \( W \) and two functions \( f \) and \( g \) as in the diagram below, there is a unique function \( h \) which makes the diagram commute; namely, \( h \circ w = \langle f \circ w, g \circ w \rangle \):

\[
\begin{array}{ccc}
W & \xrightarrow{f} & X \\
& \swarrow_f \downarrow & \searrow_{g} \downarrow \\
X \times Y & \xrightarrow{h} & Y.
\end{array}
\]

Thus, given \( X \) and \( Y \), \( \langle p, q \rangle \) is “universal” among pairs of functions from some set to \( X \) and \( Y \), because any other such pair \( \langle f, g \rangle \) factors uniquely (via \( h \)) through the pair \( \langle p, q \rangle \). This property describes the cartesian product \( X \times Y \) uniquely (up to a bijection); the same diagram, read in the category of topological spaces or of groups, describes uniquely the cartesian product of spaces or of the direct product of groups.

Adjointness is another expression for these universal properties. If we write \( \text{hom}(W, X) \) for the set of all functions \( f : W \to X \) and \( \text{hom}(U, V) \) for the set of all pairs of functions \( f: U \to X \), \( g : V \to Y \), the correspondence \( h \mapsto \langle ph, qh \rangle = \langle f, g \rangle \) indicated in the diagram above is a bijection

\[
\text{hom}(W, X \times Y) \cong \text{hom}(\langle W, W \rangle, \langle X, Y \rangle).
\]

This bijection is “natural” in the sense (to be made more precise later) that it is defined in “the same way” for all sets \( W \) and for all pairs of sets \( \langle X, Y \rangle \) (and it is likewise “natural” when interpreted for topological spaces or for groups). This natural bijection involves two constructions on sets: The construction \( W \to W' \), which sends each set to the diagonal pair \( \delta W = \langle W, W \rangle \), and the construction \( \langle X, Y \rangle \to X \times Y \) which sends each pair of sets to its cartesian product. Given the bijection above, we say that the construction \( X \times Y \) is a \textit{right adjoint} to the construction \( \delta \), and that \( \delta \) is left adjoint to the product. Adjoints, as we shall see, occur throughout mathematics.

The construction “cartesian product” is called a “functor” because it applies suitably to sets \( X \) and \( Y \) and to the functions between them; two functions \( k : X \to X' \) and \( f : Y \to Y' \) have a function \( k \times f \) as their cartesian product:

\[
k \times f : X \times Y \to X' \times Y', \quad \langle x, y \rangle \mapsto \langle k(x), f(y) \rangle.
\]

Observe also that the one-point set \( 1 = \{0\} \) serves as an identity under the operation “cartesian product”, in view of the bijections

\[
X \xrightarrow{1} X \quad X \xleftarrow{1} X
\]

given by \( \lambda(x) = x \), \( q(x, 0) = x \).

The notion of a monoid (a semigroup with identity) plays a central role in category theory. A monoid \( M \) may be described as a set \( M \) together with two functions

\[
\mu : M \times M \to M, \quad \eta : 1 \to M
\]

such that the following two diagrams commute

\[
\begin{array}{ccc}
M \times M \times M & \xrightarrow{1 \times \mu} & M \times M \\
\downarrow_{\mu \times 1} & & \downarrow_{\mu} \\
M & = & M
\end{array}
\]

\[
\begin{array}{ccc}
M \times M & \xrightarrow{\eta} & M \\
\downarrow & & \downarrow \\
M & = & M
\end{array}
\]

\[
\begin{array}{ccc}
M \times M & \xrightarrow{\eta \times 1} & M \times M \\
\downarrow & & \downarrow \\
M & = & M
\end{array}
\]

\[
\begin{array}{ccc}
M \times M & \xrightarrow{1 \times \eta} & M \times M \\
\downarrow & & \downarrow \\
M & = & M
\end{array}
\]
here 1 in 1 × μ is the identity function M → M, and 1 in 1 × M is the one-
point set 1 = {0}, while λ and ϑ are the bijections of (1) above. To say
that these diagrams commute means that the following composites are
equal:
\[ \mu \cdot (1 \times \mu) = \mu \cdot (\mu \times 1), \quad \mu \cdot (\eta \times 1) = \lambda, \quad \mu \cdot (1 \times \eta) = \varrho. \]
These diagrams may be rewritten with elements, writing the function μ
(say) as a product \( \mu(x, y) = xy \) for \( x, y \in M \) and replacing the function η
on the one-point set 1 = {0} by its (only) value, an element \( \eta(0) = u \in M \).
The diagrams above then become
\[ \langle x, y, z \rangle \mapsto \langle x, y z \rangle \quad \langle 0, x \rangle \mapsto \langle u, x \rangle \quad \langle x, u \rangle \mapsto \langle x, 0 \rangle \]
\[ \langle x y, z \rangle \mapsto \langle x y z = x(y z), \quad x = u x, \quad u x = x. \]
They are exactly the familiar axioms on a monoid, that the multiplication
be associative and have an element \( u \) as left and right identity. This
indicates, conversely, how algebraic identities may be expressed by
commutative diagrams. The same process applies to other identities;
for example, one may describe a group as a monoid \( M \) equipped with
a function \( \zeta : M → M \) (of course, the function \( x → x^{-1} \)) such that the
following diagram commutes
\[ M \xrightarrow{\delta} M \times M \xrightarrow{\cdot} M \xrightarrow{\cdot} M \]
\[ 1 \xrightarrow{\eta} M \xrightarrow{\mu} M \]
\[ \langle x, y \rangle \mapsto \langle x, y x \rangle \mapsto \langle x, y x x^{-1} \rangle. \]

Here \( \delta : M → M \times M \) is the diagonal function \( x → \langle x, x \rangle \) for \( x \in M \),
while the unnamed vertical arrow \( M → 1 = \{0\} \) is the evident (and unique)
function from \( M \) to the one-point set. As indicated just to the right,
this diagram does state that \( \zeta \) assigns to each element \( x \in M \) an element
\( x^{-1} \) which is a right inverse to \( x \).

This definition of a group by arrows \( \mu, \eta, \zeta \) in such commutative
diagrams makes no explicit mention of group elements, so applies
to other circumstances. If the letter \( M \) stands for a topological space
(not just a set) and the arrows are continuous maps (not just functions),
then the conditions (3) and (4) define a topological group -- for they
specify that \( M \) is a topological space with a binary operation \( \mu \) of multi-
plication which is continuous (simultaneously in its arguments) and
which has a continuous right inverse, all satisfying the usual group
axioms. Again, if the letter \( M \) stands for a differentiable manifold (of
class \( C^k \)) while 1 is the one-point manifold and the arrows \( \mu, \eta, \zeta \)
are smooth mappings of manifolds, then the diagrams (3) and (4) become
the definition of a Lie group. Thus groups, topological groups, and Lie
groups can all be described as "diagrammatic" groups in the respective
categories of sets, of topological spaces, and of differentiable manifolds.

This definition of a group in a category depended (for the inverse
in (4)) on the diagonal map \( \delta : M → M \times M \) to the cartesian square
\( M \times M \). The definition of a monoid is more general, because the cartesian
product \( × \) in \( M \times M \) may be replaced by any other operation \( \square \) on two
objects which is associative and which has a unit 1 in the sense prescribed
by the isomorphisms (1). We can then speak of a monoid in the system
\( (C, \square, 1) \), where \( C \) is the category, \( \square \) is such an operation, and 1 is its
unit. Consider, for example, a monoid \( M \) in \( (Ab, \otimes, \mathbb{Z}) \), where \( Ab \) is
the category of abelian groups, \( × \) is replaced by the usual tensor product
of abelian groups, and 1 is replaced by \( \mathbb{Z} \), the usual group of integers;
then (1) is replaced by the familiar isomorphism
\[ \mathbb{Z} \otimes X \cong X \otimes \mathbb{Z}, \quad X \text{ an abelian group.} \]

Then a monoid \( M \) in \( (Ab, \otimes, \mathbb{Z}) \) is, we claim, simply a ring. For the given
morphism \( \mu : M \otimes M → M \) is, by the definition of \( \otimes \), just a function
\( M \times M → M \), call it multiplication, which is bilinear; i.e., distributive
over addition on the left and on the right, while the morphism \( \eta : 1 → M \)
of abelian groups is completely determined by picking out one element
of \( M \); namely, the image \( u \) of the generator 1 of \( \mathbb{Z} \). The commutative
diagrams (3) now assert that the multiplication \( \mu \) in the abelian group \( M \)
is associative and has \( u \) as left and right unit. In other words, that \( M \)
is indeed a ring (with identity = unit).

The (homo)-morphisms of an algebraic system can also be described
by diagrams. If \( \langle M, \mu, \eta, \zeta \rangle \) and \( \langle M', \mu', \eta', \zeta' \rangle \) are two monoids, each described
by diagrams as above, then a morphism of the first to the second may
be defined as a function \( f : M → M' \) such that the following diagrams commute
\[ \begin{array}{ccc}
M & \xrightarrow{f} & M' \\
\downarrow f & & \downarrow f \\
M & \xrightarrow{f \circ f} & M'
\end{array} \]
In terms of elements, this asserts that \( f(xy) = (fx)(fy) \) and \( fu = u' \),
with \( u \) and \( u' \) the unit elements; thus a homomorphism is, as usual, just
a function preserving composition and units. If \( M \) and \( M' \) are monoids
in \( (Ab, \otimes, \mathbb{Z}) \); that is, rings, then a homomorphism \( f \) as here defined is
just a morphism of rings (preserving the units).
Finally, an action of a monoid \( \langle M, \mu, \eta \rangle \) on a set \( S \) is defined to be a function \( v : M \times S \to S \) such that the following two diagrams commute:

\[
\begin{array}{ccc}
M \times M \times S & \xrightarrow{1 \times v} & M \times S \\
\downarrow{\mu \times 1} & & \downarrow{v} \\
M \times S & \xrightarrow{v} & S
\end{array}
\]

If we write \( v(x, s) = x \cdot s \) to denote the result of the action of the monoid element \( x \) on the element \( s \in S \), these diagrams state just that

\[ x \cdot (y \cdot s) = (x y) \cdot s, \quad u \cdot s = s \]

for all \( x, y \in M \) and all \( s \in S \). These are the usual conditions for the action of a monoid on a set, familiar especially in the case of a group acting on a set as a group of transformations. If we shift from the category of sets to the category of topological spaces, we get the usual continuous action of a topological monoid \( M \) on a topological space \( S \).

If we take \( \langle M, \mu, \eta \rangle \) to be a monoid in \( (\text{Ab}, \otimes, \mathbb{Z}) \), then an action of \( M \) on an object \( S \) of \( \text{Ab} \) is just a left module \( S \) over the ring \( M \).

---

I. Categories, Functors, and Natural Transformations

1. Axioms for Categories

First we describe categories directly by means of axioms, without using any set theory, and calling them "metacategories". Actually, we begin with a simpler notion, a (meta)graph.

A metagraph consists of objects \( a, b, c, \ldots \), arrows \( f, g, h, \ldots \), and two operations, as follows:

- **Domain**, which assigns to each arrow \( f \) an object \( a = \text{dom} f \);
- **Codomain**, which assigns to each arrow \( f \) an object \( b = \text{cod} f \).

These operations on \( f \) are best indicated by displaying \( f \) as an actual arrow starting at its domain (or "source") and ending at its codomain (or "target"):

\[ f : a \to b \quad \text{or} \quad a \xrightarrow{f} b. \]

A finite graph may be readily exhibited: Thus \( \begin{array}{ccc} b & \xrightarrow{f} & c \\
a & \xrightarrow{e} & d \end{array} \)

A metacategory is a metagraph with two additional operations:

- **Identity**, which assigns to each object \( a \) an arrow \( \text{id}_a : a \to a \);
- **Composition**, which assigns to each pair \( \langle g, f \rangle \) of arrows with \( \text{dom} g = \text{cod} f \) an arrow \( g \circ f \), called their composite, with \( g \circ f : \text{dom} f \to \text{cod} g \). This operation may be pictured by the diagram

\[
\begin{array}{ccc}
b & \xrightarrow{f} & c \\
a & \xrightarrow{g} & b \\
\end{array}
\]

which exhibits all domains and codomains involved. These operations in a metacategory are subject to the two following axioms:

- **Associativity.** For given objects and arrows in the configuration

\[ a \xrightarrow{f} b \xrightarrow{g} c \xrightarrow{h} d \]

one always has the equality

\[ k \circ (g \circ f) = (k \circ g) \circ f. \]  \quad (1)

---
This axiom asserts that the associative law holds for the operation of composition whenever it makes sense (i.e., whenever the composites on either side of (1) are defined). This equation is represented pictorially by the statement that the following diagram is commutative

\[
\begin{array}{ccc}
  a & \xrightarrow{f} & b \\
  \downarrow{g} & & \downarrow{h} \\
  c & \xrightarrow{k} & d
\end{array}
\]

\[
\begin{array}{ccc}
  \textbf{Unit law. For all arrows } f: a \rightarrow b \text{ and } g: b \rightarrow c \text{ composition with the identity arrow } 1_b, \text{ gives }
  \\
  1_b \circ f = f \quad \text{and} \quad g \circ 1_b = g.
\end{array}
\]

This axiom asserts that the identity arrow \( 1_b \) of each object \( b \) acts as an identity for the operation of composition, whenever this makes sense. The Eqs. (2) may be represented pictorially by the statement that the following diagram is commutative:

\[
\begin{array}{ccc}
  a & \xrightarrow{f} & b \\
  \downarrow{g} & & \downarrow{1_b} \\
  b & \xrightarrow{1_b} & c
\end{array}
\]

We use many such diagrams consisting of vertices (labelled by objects of a category) and edges (labelled by arrows of the same category). Such a diagram is commutative when, for each pair of vertices \( c \) and \( c' \), any two paths formed from directed edges leading from \( c \) to \( c' \) yield, by composition of labels, equal arrows from \( c \) to \( c' \). A considerable part of the effectiveness of categorical methods rests on the fact that such diagrams in each situation vividly represent the actions of the arrows at hand.

If \( b \) is any object of a metacategory \( C \), the corresponding identity arrow \( 1_b \) is uniquely determined by the properties (2). For this reason, it is sometimes convenient to identify the identity arrow \( 1_b \) with the object \( b \) itself, writing \( b : b \rightarrow b \). Thus \( 1_b = b = \text{id}_b \), as may be convenient.

A metacategory is to be any interpretation which satisfies all these axioms. An example is the metacategory of sets, which has objects all sets and arrows all functions, with the usual identity functions and the usual composition of functions. Here "function" means a function with specified domain and specified codomain. Thus a function \( f: X \rightarrow Y \) consists of a set \( X \), its domain, a set \( Y \), its codomain, and a rule \( x \mapsto f(x) \) (i.e., a suitable set of ordered pairs \( \langle x, f(x) \rangle \) which assigns, to each element \( x \in X \), an element \( f(x) \in Y \). These values will be written as \( f(x), f_x \), or \( f(x) \), as may be convenient. For example, for any set \( S \), the assignment \( s \mapsto s \) for all \( s \in S \) describes the identity function \( 1_S: S \rightarrow S \); if \( S \) is a subset of \( Y \), the assignment \( s \mapsto s \) also describes the inclusion or insertion function \( S \rightarrow Y \); these functions are different unless \( S = Y \). Given functions \( f: X \rightarrow Y \) and \( g: Y \rightarrow Z \), the composite function \( g \circ f: X \rightarrow Z \) is defined by \( (g \circ f)(x) = g(f(x)) \) for all \( x \in X \). Observe that \( g \circ f \) will mean first apply \( f \), then \( g \) -- in keeping with the practice of writing each function \( f \) to the left of its argument. Note, however, that many authors use the opposite convention.

To summarize, the metacategory of all sets has as objects, all sets, as arrows, all functions with the usual composition. The metacategory of all groups is described similarly: Objects are all groups \( G, H, K \); arrows are all those functions \( f \) from the set \( G \) to the set \( H \) for which \( f: G \rightarrow H \) is a homomorphism of groups. There are many other metacategories: All topological spaces with continuous functions as arrows; all compact Hausdorff spaces with the same arrows; all ringed spaces with their morphisms, etc. The arrows of any metacategory are often called its morphisms.

Since the objects of a metacategory correspond exactly to its identity arrows, it is technically possible to dispense altogether with the objects and deal only with arrows. The data for an arrows-only metacategory \( C \) consist of arrows, certain ordered pairs \( \langle g, f \rangle \), called the composable pairs of arrows, and an operation assigning to each composable pair \( \langle g, f \rangle \) an arrow \( g \circ f \), called their composite. We say \( g \circ f \) is defined if \( \langle g, f \rangle \) is a composable pair.

With these data one defines an identity of \( C \) to be an arrow \( u \) such that \( f \circ u = f \) whenever the composite \( f \circ u \) is defined and \( u \circ g = g \) whenever \( u \circ g \) is defined. The data are then required to satisfy the following three axioms:

(i) The composite \((k \circ g) \circ f \) is defined if and only if the composite \( k \circ (g \circ f) \) is defined. It is defined whenever both \( k \circ g \) and \( g \circ f \) are defined.

(ii) The triple composite \( k \circ g \circ f \) is defined whenever both composites \( k \circ g \) and \( g \circ f \) are defined.

(iii) For each arrow \( g \) of \( C \) there exist identity arrows \( u \) and \( u' \) of \( C \) such that \( u' \circ g = g \) and \( g \circ u = g \). The data are then required to satisfy the following three axioms:

In view of the explicit definition given above for identity arrows, the last axiom is a quite powerful one; it implies that \( u' \) and \( u \) are unique in (ii), and it gives for each arrow \( g \) a codomain \( u' \) and a domain \( u \). These axioms are equivalent to the preceding ones. More explicitly, given a metacategory of objects and arrows, its arrows, with the given composition, satisfy the "arrows-only" axioms; conversely, an arrows-only metacategory satisfies the objects-and-arrows axioms when the identity arrows, defined as above, are taken as the objects (Proof as exercise).
2. Categories

A category (as distinguished from a metacategory) will mean any interpretation of the category axioms within set theory. Here are the details. A graph (also called a “diagram scheme”) is a set \( O \) of objects, a set \( A \) of arrows, and two functions

\[
\frac{\text{dom}}{\text{cod}} : \quad A \rightarrow O .
\]

In this graph, the set of composable pairs of arrows is the set

\[
A \times_o A = \{ (g, f) \mid g, f \in A \text{ and } \text{dom} g = \text{cod} f \} ,
\]

called the “product over \( O \).”

A category is a graph with two additional functions

\[
\begin{array}{l}
O \xrightarrow{\text{id}} A, \\
A \times_o A \xrightarrow{\text{c}} A,
\end{array}
\]

called identity and composition, such that

\[
\text{dom}(\text{id}_a) = a = \text{cod}(\text{id}_a), \quad \text{dom}(g \circ f) = \text{dom} f, \quad \text{cod}(g \circ f) = \text{cod} g
\]

for all objects \( a \in O \) and all composable pairs of arrows \( (g, f) \in A \times_o A \), and such that the associativity and unit axioms (1.1) and (1.2) hold.

In treating a category \( C \), we usually drop the letters \( A \) and \( O \), and write

\[
c \in C \quad f \in C
\]

for “\( c \) is an object of \( C \)” and “\( f \) is an arrow of \( C \),” respectively. We also write

\[
\text{hom}(b, c) = \{ f \mid f \in C, \text{dom} f = b, \text{cod} f = c \}
\]

for the set of arrows from \( b \) to \( c \). Categories can be defined directly in terms of composition acting on these “\( \text{hom} \)-sets” (§ 8 below); we do not follow this custom because we put the emphasis not on sets but on categories, in the general sense described in the introduction. For the moment, we consider examples.

0. The empty category (no objects, no arrows);
1. The category \( \mathcal{D} \) with one object and one (identity) arrow;
2. The category \( \mathcal{D} \rightarrow \mathcal{D} \) with two objects \( a, b \), and just one arrow \( a \rightarrow b \) not the identity;
Categories, Functors, and Natural Transformations

0 is the empty set, while the first infinite ordinal is \( \omega = \{0, 1, 2, \ldots \} \).
Each ordinal \( n \) is linearly ordered, and hence is a category (a preorder).
For example, the categories 1, 2, and 3 listed above are the preorders belonging to the (linearly ordered) ordinal numbers 1, 2, and 3. Another example is the linear order \( \omega \). As a category, it consists of the arrows
\[
0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow \cdots,
\]
all their composites, and the identity arrows for each object.

\( D \) is the category with objects all finite ordinals and arrows \( f : m \rightarrow n \) all order-preserving functions \( i \leq j \) in \( m \) implies \( f(i) \leq f(j) \) in \( n \). This category \( D \), sometimes called the simplicial category, plays a central role (Chapter VII).

\( \text{Finord} = \text{Set}_n \) is the category with objects all finite ordinals \( n \) and arrows \( f : m \rightarrow n \) all functions from \( m \) to \( n \). This is essentially the category of all finite sets, using just one finite set \( n \) for each finite cardinal number \( n \).

Large Categories. In addition to the metacategory of all sets – which is not a set – we want an actual category \( \text{Set} \), the category of all small sets. We shall assume that there is a big enough set \( U \), the “universe”, then describe a set \( x \) as “small” if it is a member of the universe, and take \( \text{Set} \) to be the category whose set \( U \) of objects is the set of all small sets, with arrows all functions from one small set to another. With this device (details in §7 below) we construct other familiar large categories, as follows:

- \( \text{Set} \): Objects, all small sets; arrows, all functions between them.
- \( \text{Set}_n \): Objects, small sets each with a selected base-point; arrows, base-point preserving functions.
- \( \text{Ens} \): Category of all sets and functions within a (variable) set \( V \).
- \( \text{Cat} \): Objects, all small categories; arrows, all functors (§3).
- \( \text{Mon} \): Objects, all small monoids; arrows, all morphisms of monoids.
- \( \text{Grp} \): Objects, all small groups; arrows, all morphisms of groups.
- \( \text{Ab} \): Objects, all small (additive) abelian groups, with morphisms of such.
- \( \text{Ring} \): All small rings, with the ring morphisms (preserving units) between them.
- \( \text{CRing} \): All small commutative rings and their morphisms.
- \( \text{R-Mod} \): All small left modules over the ring \( R \), with linear maps.
- \( \text{Mod-R} \): Small right \( R \)-modules.
- \( \text{K-Mod} \): Small modules over the commutative ring \( K \).
- \( \text{Top} \): Small topological spaces and continuous maps.
- \( \text{Toph} \): Topological spaces, with arrows homotopy classes of maps.
- \( \text{Toph}_* \): Spaces with selected base point, base point-preserving maps.

Particular categories (like these) will always appear in bold-face type.

Script capitals are used by many authors to denote categories.

3. Functors

A functor is a morphism of categories. In detail, for categories \( C \) and \( B \) a functor \( T : C \rightarrow B \) with domain \( C \) and codomain \( B \) consists of two suitably related functions: The object functor \( T \), which assigns to each object \( c \) of \( C \) an object \( Tc \) of \( B \) and the arrow function (also written \( T \)) which assigns to each arrow \( f : c \rightarrow c' \) of \( C \) an arrow \( Tf : Tc \rightarrow Tc' \) of \( B \), in such a way that
\[
T(1_c) = 1_{Tc}, \quad T(g \circ f) = Tg \circ Tf,
\]
the latter whenever the composite \( g \circ f \) is defined in \( C \). A functor, like a category, can be described in the “arrows-only” fashion: It is a function \( T \) from arrows \( f \) of \( C \) to arrows \( Tf \) of \( B \), carrying each identity of \( C \) to an identity of \( B \) and each composable pair \( (g, f) \) in \( C \) to a composable pair \( (Tg, Tf) \) in \( B \), with \( Tg \circ Tf = T(g \circ f) \).

A simple example is the power set functor \( \mathcal{P} : \text{Set} \rightarrow \text{Set} \). Its object function assigns to each set \( X \) the usual power set \( \mathcal{P} X \), with elements all subsets \( S \subseteq X \); its arrow function assigns to each \( f : X \rightarrow Y \) that map \( \mathcal{P}f : \mathcal{P}X \rightarrow \mathcal{P}Y \) which sends each \( S \subseteq X \) to its image \( f(S) \subseteq Y \). Since both \( \mathcal{P}(1_X) = 1_{\mathcal{P}X} \) and \( \mathcal{P}(g \circ f) = \mathcal{P}g \circ \mathcal{P}f \), this clearly defines a functor \( \mathcal{P} : \text{Set} \rightarrow \text{Set} \).

Functors were first explicitly recognized in algebraic topology, where they arise naturally when geometric properties are described by means of algebraic invariants. For example, singular homology in a given dimension \( n \) (a natural number) assigns to each topological space \( X \) an abelian group \( H_n(X) \), the \( n \)-th homology group of \( X \), and also to each continuous map \( f : X \rightarrow Y \) of spaces a corresponding homomorphism \( H_n(f) : H_n(X) \rightarrow H_n(Y) \) of groups, and this in such a way that \( H_n \) becomes a functor \( \text{Top} \rightarrow \text{Ab} \). For example, if \( X = I^m \) is the circle, \( H_0(I^m) = \mathbb{Z} \), so the group homomorphism \( H_0(f) : Z \rightarrow Z \) is determined by an integer \( d \) (the image of 1); this integer is the usual “degree” of the continuous map \( f : I^m \rightarrow I^n \). In this case and in general, homotopic maps \( g, h : X \rightarrow Y \) yield the same homomorphism \( H_n(X) \rightarrow H_n(Y) \), so \( H_n \) can actually be regarded as a functor \( \text{Toph} \rightarrow \text{Grp} \), defined on the homotopy category. The Eilenberg-Steenrod axioms for homology start with the axioms that \( H_n \) for each natural number \( n \) is a functor on \( \text{Toph} \), and continue with certain additional properties of these functors. The more recently developed extraordinary homology and cohomology theories are also functors on \( \text{Toph} \). The homotopy groups \( \pi_n(X) \) of a space \( X \) can also be regarded as functors; since they depend on the choice of a base point in \( X \), they are functors \( \text{Toph}_* \rightarrow \text{Grp} \). The leading idea in the use of functors in topology is that \( H_n \) or \( \pi_n \) gives an algebraic picture or image not just of the topological spaces, but also of all the continuous maps between them.
Functors arise naturally in algebra. To any commutative ring \( K \) the set of all non-singular \( n \times n \) matrices with entries in \( K \) is the usual general linear group \( GL_n(K) \); moreover, each homomorphism \( f : K \to K' \) of rings produces in the evident way a homomorphism \( GL_n(f) : GL_n(K) \to GL_n(K') \) of groups. These data define for each natural number \( n \) a functor \( GL_n : CRng \to Grp \). For any group \( G \) the set of all products of commutators \( xyx^{-1}y^{-1} (x, y \in G) \) is a normal subgroup \([G, G] \) of \( G \), called the commutator subgroup. Since any homomorphism \( G \to H \) of groups carries commutators to commutators, the assignment \( G \to [G, G] \) defines an evident functor \( Grp \to Grp \), while \( G \to [G, G] \) defines a functor \( Grp \to Ab \), the factor-commutator functor. Observe, however, that the center \( Z(G) \) of \( G \) (all \( a \in G \) with \( ax = xa \) for all \( x \)) does not naturally define a functor \( Grp \to Grp \), because a homomorphism \( G \to H \) may carry an element in the center of \( G \) to one not in the center of \( H \).

A functor which simply "forgets" some or all of the structure of an algebraic object is commonly called a forgetful functor (or, an underlying functor). Thus the forgetful functor \( U : Grp \to Set \) assigns to each group \( G \) the set \( U(G) \) of its elements ("forgetting" the multiplication and hence the group structure), and assigns to each morphism \( f : G \to G' \) of groups the same function \( f \), regarded just as a function between sets. The forgetful functor \( U : Rng \to Ab \) assigns to each ring \( R \) the additive abelian group \( \mathbb{Z} \) and to each morphism \( f : R \to R' \) of rings the same function, regarded just as a morphism of addition.

Functors may be composed. Explicitly, given functors

\[
C \xrightarrow{F} B \xrightarrow{G} A
\]

between categories \( A, B, \) and \( C \), the composite functions

\[
c \mapsto S(Tc) \quad f \mapsto S(Tf)
\]
on objects \( c \) and arrows \( f \) of \( C \) define a functor \( S : T : C \to A \), called the composite (in that order) of \( S \) with \( T \). This composition is associative. For each category \( B \) there is an identity functor \( I_B : B \to B \), which acts as an identity for this composition. Thus we may consider the metacategory of all categories: its objects are all categories, its arrows are all functors with the composition above. Similarly, we may form the category \( \text{Cat} \) of all small categories — but not the category of all categories.

An isomorphism \( T : C \to B \) of categories is a functor \( T \) from \( C \) to \( B \) which is a bijection, both on objects and on arrows. Alternatively, but equivalently, a functor \( T : C \to B \) is an isomorphism if and only if there is a functor \( S : B \to C \) for which both compositions \( ST = 1_C \) and \( TS = 1_B \) are identity functors; then \( S \) is the two-sided inverse \( S = T^{-1} \).

Certain properties much weaker than isomorphism will be useful.

A functor \( T : C \to B \) is full when to every pair \( c, c' \) of objects of \( C \) and to every arrow \( g : Tc \to Tc' \) of \( B \), there is an arrow \( f : c \to c' \) of \( C \) with \( g = Tf \). Clearly the composite of two full functors is a full functor.

A functor \( T : C \to B \) is faithful (or an embedding) when to every pair \( c, c' \) of objects of \( C \) and to every pair \( f_1, f_2 : c \to c' \) of parallel arrows of \( C \) the equality \( Tf_1 = Tf_2 : Tc \to Tc' \) implies \( f_1 = f_2 \). Again, composites of faithful functors are faithful. For example, the forgetful functor \( Grp \to Set \) is faithful but not full and not a bijection on objects.

These two properties may be visualized in terms of hom-sets (see (2.5)). Given a pair of objects \( c, c' \in C \), the arrow function of \( T : C \to B \) assigns to each \( f : c \to c' \) an arrow \( Tf : Tc \to Tc' \) and so defines a function

\[
T_{c,c'} : \text{hom}(c, c') \to \text{hom}(Tc, Tc'). \quad f \mapsto Tf.
\]

Then \( T \) is full when every such function is surjective, and faithful when every such function is injective. For a functor which is both full and faithful, every such function is a bijection, but this need not mean that the functor itself is an isomorphism of categories, for there may be objects of \( B \) not in the image of \( T \).

A subcategory \( S \) of a category \( C \) is a collection of some of the objects and some of the arrows of \( C \), which includes with each arrow \( f \) both the object dom \( f \) and the object cod \( f \), with each object \( x \) its identity arrow \( 1_x \) and with each pair of composable arrows \( x \to y \to z \) their composite. These conditions insure that these collections of objects and arrows themselves constitute a category \( S \). Moreover, the injection (inclusion) map \( S \to C \) which sends each object and each arrow of \( S \) to itself (in \( C \)) is a functor, the inclusion functor. This inclusion functor is automatically faithful. We say that \( S \) is a full subcategory of \( C \) when the inclusion functor \( S \to C \) is full. A full subcategory, given \( C \), is thus determined by giving just the set of its objects, since the arrows between any two of these objects \( s \to t \) are all morphisms \( s \to t \) in \( C \). For example, the category \( \text{Set} \) of all finite sets is a full subcategory of the category \( \text{Set} \).

Exercises

1. Show how each of the following constructions can be regarded as a functor:
   - The field of quotients of an integral domain: the Lie algebra of a Lie group.
   - Show that functors \( 1 \to C, 2 \to C, \) and \( 3 \to C \) respectively correspond to objects, arrows, and composable pairs of arrows in \( C \).
   - Interpret "functor" in the following special types of categories: (a) A functor between two preorders is a function \( T \) which is monotonic (i.e., \( p \leq q \) implies \( Tp \leq Tq \)). (b) A functor between two groups (one-object categories) is a morphism of groups. (c) If \( G \) is a group, a functor \( G \to \text{Set} \) is a permutation representation of \( G \); while \( G \to \text{Mat}_k \) is a matrix representation of \( G \).
   - Prove that there is no functor \( Grp \to Ab \) sending each group \( G \) to its center (Consider \( S_3 \to S_3 \), the symmetric groups).
   - Find two different functors \( T : Grp \to Grp \) with object function \( T(G) = G \) the identity for every group \( G \).
4. Natural Transformations

Given two functors $S, T : C \to B$, a natural transformation $\tau : S \to T$ is a function which assigns to each object $c$ of $C$ an arrow $\tau_c : Sc \to Tc$ of $B$ in such a way that every arrow $f : c \to c'$ in $C$ yields a diagram

$$
\begin{array}{ccc}
Sc & \xrightarrow{\tau_c} & Tc \\
\downarrow{f} & & \downarrow{\tau_f} \\
Sc' & \xrightarrow{\tau_c'} & Tc'
\end{array}
$$

which is commutative. When this holds, we also say that $\tau_c : Sc \to Tc$ is natural in $c$. If we think of the functor $S$ as giving a picture in $B$ of all the objects and arrows of $C$, then a natural transformation $\tau$ is a set of arrows mapping (or, translating) the picture $S$ to the picture $T$, with all squares (and parallelograms!) like that above commutative.

We call $\tau a, \tau b, \tau c, \ldots$ the components of the natural transformation $\tau$.

A natural transformation is often called a morphism of functors; a natural transformation $\tau$ with every component $\tau c$ invertible in $B$ is called a natural isomorphism or a natural equivalence; in symbols, $\tau : S \cong T$. In this case, the inverses $(\tau c)^{-1}$ in $B$ are the components of a natural isomorphism $\tau^{-1} : T \cong S$.

The determinant is a natural transformation. To be explicit, let $\det_K M$ be the determinant of the $n \times n$ matrix $M$ with entries in the commutative ring $K$, while $K^*$ denotes the group of units (invertible elements) of $K$. Thus $M$ is non-singular when $\det_K M$ is a unit, and $\det_K$ is a morphism $\text{GL}_n K \to K^*$ of groups (an arrow in $\text{Grp}$). Because the determinant is defined by the same formula for all rings $K$, each morphism $f : K \to K'$ of commutative rings leads to a commutative diagram

$$
\begin{array}{ccc}
\text{GL}_n K & \xrightarrow{\det_K} & K^* \\
\downarrow{f^*} & & \downarrow{f^*} \\
\text{GL}_n K' & \xrightarrow{\det_{K'}} & K'^*
\end{array}
$$

This states that the transformation $\det : \text{GL}_n K \to K^*$ is natural between two functors $\text{Crtng} \to \text{Grp}$.

For each group $G$ the projection $p_G : G \to G/[G, G]$ to the factor-commutator group defines a transformation $p$ from the identity functor on $\text{Grp}$ to the factor-commutator functor $\text{Grp} \to \text{Ab} \to \text{Grp}$. Moreover, $p$ is natural, because each group homomorphism $f : G \to H$ defines the evident homomorphism $f'$ for which the following diagram commutes:

$$
\begin{array}{ccc}
G & \xrightarrow{p_G} & G/[G, G] \\
\downarrow{f} & & \downarrow{f'} \\
H & \xrightarrow{p_H} & H/[H, H]
\end{array}
$$

The double character group yields a suggestive example in the category $\text{Ab}$ of all abelian groups $G$. Let $D(G)$ denote the character group of $G$, so that $D(G) = \text{hom}(G, \mathbb{R}/\mathbb{Z})$ is the set of all homomorphisms $\chi : G \to \mathbb{R}/\mathbb{Z}$ with the familiar group structure, where $\mathbb{R}/\mathbb{Z}$ is the additive group of real numbers modulo 1. Each arrow $f : G \to H$ in $\text{Ab}$ determines an arrow $D f : D G \to D H$ (opposite direction!) in $\text{Ab}$, with $(D f) t = t f$ for each $t$; for composable arrows, $D(f \circ g) = D f \circ D g$. Because of this reversal, $D$ is not a functor (it is a “contravariant” functor on $\text{Ab}$ to $\text{Ab}$, see §II.12); however, the twice iterated character group $G \to D(DG)$ and the identity $I(G) = G$ are both functors $\text{Ab} \to \text{Ab}$. For each group $G$ there is a homomorphism

$$
\tau_G : G \to D(DG)
$$

obtained in a familiar way: To each $g \in G$ assign the function $\tau_G g : D G \to \mathbb{R}/\mathbb{Z}$ given for any character $\chi \in D G$ by $t \mapsto t\chi(g)$; thus $(\tau_G g)(t) = t(g)$. One verifies at once that $\tau$ is a natural transformation $\tau : I \to D D$; this statement is just a precise expression for the elementary observation that the definition of $\tau$ depends on no artificial choices of bases, generators, or the like. In case $G$ is finite, $\tau_G$ is an isomorphism; thus, if we restrict all functors to the category $\text{Ab}_F$ of finite abelian groups, $\tau$ is a natural isomorphism.

On the other hand, for each finite abelian group $G$ there is an isomorphism $\sigma_G : G \cong D G$ of $G$ to its character group, but this isomorphism depends on a representation of $G$ as a direct product of cyclic groups and so cannot be natural. More explicitly, we can make $D$ into a co-variant functor $D' : \text{Ab}_F \to \text{Ab}_F$, on the category $\text{Ab}_F$ with objects all finite abelian groups and arrows all isomorphisms $f$ between such groups, setting $D' G = D G$ and $D' f = D f$. Then $\sigma_G : G \to D' G$ is a map $g : l \mapsto D' l$ of functors $\text{Ab}_F \to \text{Ab}_F$, but it is not natural in the sense of our definition.

A parallel example is the familiar natural isomorphism of a finite-dimensional vector space to its double dual.

Another example of naturality arises when we compare the category $\text{Floord}$ of all finite ordinal numbers $n$ with the category $\text{Set}$ of all finite
sets (in some universe \( U \)). Every ordinal \( n = \{0, 1, \ldots, n-1\} \) is a finite set, so the inclusion \( S \) is a functor \( S : \text{Finord} \to \text{Set} \). On the other hand, each finite set \( X \) determines an ordinal number \( n = \# X \), the number of elements in \( X \); we may choose for each \( X \) a bijection \( \theta_X : X \to X \). For any function \( f : X \to Y \) between finite sets we may then define a corresponding function \( \ast f : \# X \to \# Y \) between ordinals by \( \ast f = \theta_Y f \theta_X^{-1} \); this insures that the diagram

\[
\begin{array}{ccc}
X & \xrightarrow{g} & \# X \\
\downarrow f & & \downarrow \ast f \\
Y & \xrightarrow{\theta_Y} & \# Y
\end{array}
\]

will commute, and makes \( \ast \) a functor \( \ast : \text{Set} \to \text{Finord} \). If \( X \) is itself an ordinal number, we may take \( \theta_X \) to be the identity. This insures that the composite functor \( \# \circ S \) is the identity functor \( I \) of \( \text{Finord} \). On the other hand, the composite \( S \circ \# \) is not the identity functor \( I \) : \text{Set} \to \text{Set} \), because it sends each finite set \( X \) to a special finite set – the ordinal number \( n \) with the same number of elements as \( X \). However, the square diagram above does show that \( \theta : I \to \# \) is a natural isomorphism. All told we have \( I \cong S \circ \# \), \( I' = \# \circ S \).

More generally, an equivalence between categories \( C \) and \( D \) is defined to be a pair of functors \( S : C \to D \), \( T : D \to C \) together with natural isomorphisms \( l_C \cong T \circ S \), \( l_D \cong S \circ T \). This example shows that this notion (to be examined in §IV.4) allows us to compare categories which are “alike” but of very different “sizes”.

We shall use many other examples of naturality. As Eilenberg-Mac Lane first observed, “category” has been defined in order to define “functor” and “functor” has been defined in order to define “natural transformation”.

**Exercises**

1. Let \( S \) be a fixed set, and \( X^S \) the set of all functions \( h : S \to X \). Show that \( X \to X^S \) is the object function of a functor \( \text{Set} \to \text{Set} \), and that evaluation \( e_x : X^S \times S \to X \), defined by \( e(h, s) = h(s) \), is a natural transformation.

2. If \( H \) is a fixed group, show that \( G \to H \times G \) defines a functor \( H \times - : \text{Grp} \to \text{Grp} \), and that each morphism \( f : H \to K \) of groups defines a natural transformation \( H \times - \to K \times - \).

3. If \( B \) and \( C \) are groups (regarded as categories with one object each) and \( S, T : B \to C \) are functors (homomorphisms of groups), show that there is a \( Tg = hSg \) if and only if \( S \) and \( T \) are conjugate; i.e., if and only if there is an element \( h \in C \) with \( Tg = hSg h^{-1} \) for all \( g \in B \).

4. For functors \( S, T : C \to P \) where \( C \) is a category and \( P \) a preorder, show that there is a natural transformation \( S \to T \) (which is then unique) if and only if \( Sc \subseteq Tc \) for every object \( c \in C \).

5. Show that every natural transformation \( \tau : S \to T \) defines a function (also called \( \tau \)) which sends each arrow \( f : c \to c' \) of \( C \) to an arrow \( \tau f : Sc \to Tc' \) of \( B \) in such a way that \( Tg \circ \tau f = \tau (g \circ f) = \tau g \circ \tau f \) for each composable pair \( (g, f) \). Conversely, show that every such function \( \tau \) comes from a unique natural transformation with \( \tau e = \tau (1_e) \). (This gives an “arrows only” description of a natural transformation.)

6. Let \( F \) be a field. Show that the category of all finite-dimensional vector spaces over \( F \) (with morphisms all linear transformations) is equivalent to the category \( \text{Matr} \) described in §2.

5. **Monics, Epis, and Zeros**

In categorical treatments many properties ordinarily formulated by means of elements (elements of a set or of a group) are instead formulated in terms of arrows. For example instead of saying that a set \( X \) has just one element, one can say that for any other set \( Y \) there is exactly one function \( Y \to X \). We now formulate a few more instances of such methods of “doing without elements”.

An arrow \( a \to b \) is invertible in \( C \) if there is an arrow \( e' : b \to a \) in \( C \) with \( e'e = 1_b \) and \( e'e' = 1_a \). If such an \( e' \) exists, it is unique, and is written as \( a^{-1} \). By the usual proof, \( (e_1 e_2)^{-1} = e_2^{-1} e_1^{-1} \), provided the composite \( e_1 e_2 \) is defined and both \( e_1 \) and \( e_2 \) are invertible. Two objects \( a \) and \( b \) are isomorphic in the category \( C \) if there is an invertible arrow \( \tau \) (an isomorphism) \( \tau : a \to b \); we write \( a \cong b \). The relation of isomorphism is manifestly reflexive, symmetric, and transitive.

An arrow \( m : a \to b \) is monic in \( C \) when for any two parallel arrows \( f_1, f_2 : d \to a \) the equality \( m \circ f_1 = m \circ f_2 \) implies \( f_1 = f_2 \); in other words, \( m \) is monic if it can always be cancelled on the left (is left cancellable). In \( \text{Set} \) and in \( \text{Grp} \) the monics are precisely the injections (monomorphisms) in the usual sense; i.e., the functions which are one-one into.

An arrow \( h : a \to b \) is epic in \( C \) when for any two arrows \( g_1, g_2 : c \to a \) the equality \( g_1 \circ h = g_2 \circ h \) implies \( g_1 = g_2 \); in other words, \( h \) is epic when it is right cancellable. In \( \text{Set} \) the epic arrows are precisely the surjections (epimorphisms) in the usual sense; i.e., the functions onto.

For an arrow \( h : a \to b \), a right inverse is an arrow \( r : b \to a \) with \( hr = 1_b \). A right inverse (which is usually not unique) is also called a section of \( h \). If \( h \) has a right inverse, it is evidently epic; the converse holds in \( \text{Set} \), but fails in \( \text{Grp} \). Similarly, a left inverse for \( h \) is called a retraction for \( h \), and any arrow with a left inverse is necessarily monic. If \( gh = 1_a \), then \( g \) is a split epic, \( h \) a split monic, and the composite \( f = hg \) is defined
and is an idempotent. Generally, an arrow \( f : b \to b \) is called idempotent when \( f^2 = f \); an idempotent is said to split when there exist arrows \( g \) and \( h \) such that \( f = hg \) and \( gh = 1 \).

An object \( t \) is terminal in \( C \) if to each object \( a \) in \( C \) there is exactly one arrow \( a \to t \). If \( t \) is terminal, the only arrow \( t \to t \) is the identity, and any two terminal objects of \( C \) are isomorphic in \( C \). An object \( s \) is initial in \( C \) if to each object \( a \) there is exactly one arrow \( s \to a \). For example, in the category \( 
abla \), the empty set is an initial object and any one-point set is a terminal object. In \( 
abla \), the group with one element is both initial and terminal.

A null object \( z \) in \( C \) is an object which is both initial and terminal. If \( C \) has a null object, that object is unique up to isomorphism, while for any two objects \( a \) and \( b \) of \( C \) there is a unique arrow \( a \to z \to b \) (the composite through \( z \)) called the zero arrow from \( a \) to \( b \). Any composite with a zero arrow is itself a zero arrow. For example, the categories \( 
abla \) and \( 
abla \) have null objects (namely \( 0 \)), as does \( 
abla \) (namely the one-point set).

A groupoid is a category in which every arrow is invertible. A typical groupoid is the fundamental groupoid \( \pi(X) \) of a topological space \( X \). An object of \( \pi(X) \) is a point \( x \) of \( X \), and an arrow \( x \to x' \) of \( \pi(X) \) is a homotopy class of paths \( f \) from \( x \) to \( x' \). (Such a path \( f \) is a continuous function \( I \to X \), \( I \) the closed interval \( [0, 1] \), with \( f(0) = x, f(1) = x' \), while two paths \( f, g \) with the same end-points \( x \) and \( x' \) are homotopic when there is a continuous function \( F : I \times I \to X \) with \( F(t, 0) = f(t), F(t, 1) = g(t), \) and \( F(0, s) = x, F(1, s) = x' \) for all \( s \) and \( t \) in \( I \).) The composite of paths \( g : x' \to x'' \) and \( f : x \to x' \) is the path \( k \) which is "\( f \) followed by \( g \)" given explicitly by

\[
\begin{align*}
  h(t) = f(2t), & \quad 0 \leq t \leq 1/2, \\
  = g(2t - 1), & \quad 1/2 \leq t \leq 1.
\end{align*}
\]

Composition applies also to homotopy classes, and makes \( \pi(X) \) a category and a groupoid (the inverse of any path is the same path traced in the opposite direction).

Since each arrow in a groupoid \( G \) is invertible, each object \( x \) in \( G \) determines a group \( \text{hom}_G(x, x) \), consisting of all \( g : x \to x \). If there is an arrow \( f : x \to x' \), the groups \( \text{hom}_G(x, x) \) and \( \text{hom}_G(x', x') \) are isomorphic, under \( g \mapsto f g f^{-1} \) (i.e., under conjugation). A groupoid is said to be connected if there is an arrow joining any two of its objects. One may readily show that a connected groupoid is determined up to isomorphism by a group (one of the groups \( \text{hom}_G(x, x) \)) and by a set (the set of all objects). In this way, the fundamental groupoid \( \pi(X) \) of a path-connected space \( X \) is determined by the set of points in the space and a group \( \text{hom}_G(x, x) \) of \( X \) — the fundamental group of \( X \).

---

### Exercises

1. Find a category with an arrow which is both epi and monic, but not invertible (e.g., dense subset of a topological space).
2. Prove that the composite of monics is monic, and likewise for epis.
3. If a composite \( g \circ f \) is monic, so is \( f \). Is this true of \( g \)?
4. Show that the inclusion \( \mathbb{Z} \to \mathbb{Q} \) is epi in the category \( \mathbb{R} \).
5. In \( \mathbb{R} \) prove that every arrow is surjective (Hint: If \( \varphi : G \to H \) has image \( M \) not \( H \), use the factor group \( H/M \) if \( M \) has index 2. Otherwise, let \( \text{Perm} \) be the group of all permutations of the set \( H \), choose three different cosets \( M, Mu, Mv \) of \( M \), define \( \sigma \in \text{Perm} \) by \( \sigma(xu) = xu, \sigma(xv) = xv \) for \( x \in M \), and \( e \) otherwise the identity. Let \( \varphi : H \to \text{Perm} \) send each \( h \) to left multiplication \( \varphi_h \) of \( h \), while \( \varphi_h(e) = e \cdot \varphi_h(e) \).
6. In \( \mathbb{S} \), show that all idempotents split.
7. An arrow \( f : a \to b \) in a category \( C \) is regular when there exists an arrow \( g : b \to a \) such that \( gf = f \). Show that \( f \) is regular if it has either a left or right inverse, and prove that every arrow in \( \mathbb{S} \) is regular.
8. Consider the category with objects \((X, e, t)\), where \( X \) is a set, \( e \in X \), and \( t : X \to X \), and with arrows \( f : (X, e, t) \to (X', e', t') \) the functions \( f \) on \( X \) with \( e' = e \) and \( f t = t' f \). Prove that this category has an initial object in which \( X \) is the set of natural numbers, \( e = 0 \), and \( t \) is the successor function.
9. If the functor \( T : C \to B \) is faithful and \( TF \) is monic, show \( f \) monic.

### 6. Foundations

One of the main objectives of category theory is to discuss properties of totalities of Mathematical objects such as the "set" of all groups or the "set" of all homomorphisms between any two groups. Now it is the custom to regard a group as a set with certain added structure, so we are here proposing to consider a set of all sets with some given structure. This amounts to applying a comprehension principle: Given a property \( \varphi(x) \) of sets \( x \), form the set \( \{ x \mid \varphi(x) \} \) of all sets \( x \) with this property. However, such a principle cannot be adopted in this generality, since it would lead to some of the famous paradoxical sets, such as the set of all sets not members of themselves.

For this reason, the standard practise in naive set theory, with the usual membership relation \( \in \), is to restrict the application of the comprehension principle. One allows the formation from given sets \( u, v \) of the set \( \{ u, v \} \) (the set with exactly \( u \) and \( v \) as elements), of the ordered pair \( \langle u, v \rangle \), of an infinite set (the set \( \omega = \{ 0, 1, 2, \ldots \} \) of all finite ordinals, and of

- The Cartesian Product \( u \times v = \langle x, y \rangle \} | x \in u \text{ and } y \in v \),
- The Power Set \( \mathcal{P}(u) = \{ v \mid v \subseteq u \} \),
- The Union (of a set of sets) \( \bigcup x = \{ y \mid y \in x \text{ for some } x \} \).
Finally, given a property \( \varphi(x) \) (technically, a property expressed in terms of \( x \), the membership relation, and the usual logical connectives, including "for all sets \( \text{and} \)" and "there exists a set \( \text{a} \)" and given a set \( u \) one allows

Comprehension for elements of \( u \): \( \{ x | x \in u \text{ and } \varphi(x) \} \).

In words: One allows the set of all those \( x \) with a given property \( \varphi \) which are members of an already given set \( u \).

To this practise, we add one more assumption: The existence of a universe. A universe is defined to be a set \( U \) with the following somewhat redundant properties:

(i) \( x \in u \in U \Rightarrow x \subset U \),
(ii) \( u \in U \) and \( v \in U \) imply \( \{ u, v \}, \langle u, v \rangle, \text{and } u \times v \in U \),
(iii) \( \omega \in U \) (here \( \omega = \{ 0, 1, 2, \ldots \} \) is the set of all finite ordinals),
(iv) \( v \in U \) (here \( v = \{ 0, 1, 2, \ldots \} \) is the set of all infinite ordinals),
(v) \( f: a \to b \) is a surjective function with \( a \in U \) and \( b \subset U \), then \( b \in U \).

These closure properties for \( U \) insure that any of the standard operations of set theory applied to elements of \( U \) will always produce elements of \( U \); in particular, \( \omega \subset U \) provides that \( U \) also contains all the usual sets of real numbers and related infinite sets. We can then regard "ordinary" mathematics as carried out exclusively within \( U \) (i.e., on elements of \( U \)) while \( U \) itself and sets formed from \( U \) are to be used for the construction of the desired large categories.

Now hold the universe \( U \) fixed, and call a set \( u \in U \) a small set. Thus the universe \( U \) is the set of all small sets. Similarly, call a function \( f: u \to v \) small when \( u \) and \( v \) are small sets. This implies that \( f \) itself can be regarded as a small set -- say, as the ordered triple \( \langle u, G_f, v \rangle \), with \( G_f \subset U \times U \) the usual set of all \( \langle x, y \rangle \) with \( x \in u \) and \( y = x \). The limited comprehension principle thus allows the construction of the set \( A \) of all those \( x \) which are small functions, since these functions are all elements of \( U \). We can now define the category Set of all small sets to be that category in which \( U \) (the set of all small sets) is the set of objects and \( A \) (the set of all small functions) is the set of arrows. Henceforth Set will always denote this category.

A small group is similarly a small set with a group structure; i.e., is an ordered pair \( \langle u, m \rangle \), where \( u \) is a small set and \( m: u \times u \to u \) a function (binary operation on \( u \)) satisfying the usual group axioms. Since any small group is an element of \( U \), we may form the set of all small groups and the set of all homomorphisms between two small groups. They constitute the category Grp of all small groups.

The same process will construct the category of all small Mathematical objects of other types. For example, a category is small if the set of its arrows and the set of its objects are both small sets; we will soon form the category Cat of all small categories. Observe, however, that Set is not a small category, because the set \( U \) of its objects is not a small set (otherwise \( U \in U \), and this is contrary to the axiom of regularity, which asserts that there are no infinite chains \( \ldots \in x_{-1} \in x_{-2} \in \cdots \in x_0 \)). Similarly, Grp is not small.

This description of the foundations may be put in axiomatic form. We are assuming the standard Zermelo-Fraenkel axioms for set theory, plus the existence of a set \( U \) which is a universe. The Zermelo-Fraenkel axioms (on a membership relation \( \in \)) are: Extensionality (sets with the same elements are equal), existence of the null set, existence of the sets \( \{ u, v \}, \langle u, v \rangle, \text{and } u \times v \) for all sets \( u, v \), and \( x \) the axiom of infinity, the axiom of choice, the axiom of regularity, and the replacement axiom:

Replacement. Let \( a \) be a set and \( \varphi(x, y) \) a property which is functional for \( x \) in \( a \), in the sense that \( \varphi(x, y) \) and \( \varphi(x, y') \) for \( x \in a \) imply \( y = y' \), and that for each \( x \in a \) there exists a \( y \) with \( \varphi(x, y) \). Then there exists a set consisting of all those \( y \) such that \( \varphi(x, y) \) holds for \( x \in a \).

Briefly speaking, the replacement axiom states that the image of a set \( a \) under a "function" \( \varphi \) is a set. It can be shown that the replacement axiom implies the comprehension axiom, as stated above. Moreover, our conditions defining a universe \( U \) imply that all the sets \( x \in U \) (all the small sets) do satisfy the Zermelo-Fraenkel axioms -- for example, condition (v) in the definition of a universe corresponds to replacement. We shall see that our assumption of one universe suffices for the purposes of category theory.

Some authors assume instead sets and "classes," using, for these concepts, the Gödel-Bernays axioms. To explain this, define a class \( C \) to be any subset \( C \subset U \) of the universe. Since \( x \in u \in U \) implies \( x \in U \), every element of \( U \) is also a subset of \( U \), therefore every small set is also a class; but conversely, some classes (such as \( U \) itself) are not small sets. These latter are called the proper classes. Together, the small sets and the classes satisfy the standard Gödel-Bernays axioms (see Gödel [1940]).

A large category is one in which both the set of objects and the set of arrows are classes (proper or otherwise). Using only small sets and all classes one can describe many of the needed categories in particular, our categories Set, Grp, etc. are proper classes, hence are large categories in this sense. Initially, category theory was restricted to the study of small and large categories (and based on the Gödel-Bernays axioms). However, we will have many occasions to form categories which are not classes. One such is the category Cls of all classes: Its objects are all classes; its arrows all functions \( f: C \to C \) between classes. Then the set of objects of \( Cls \) is the set \( B(U) \) of all subsets of \( U \); it is not a class; in fact, its cardinal number is larger than the cardinal of the universe \( U \). Another useful category is Cat, the category of all large categories. It is not a class.

In the sequel we shall drop the notation \( U \) for the chosen universe and speak simply of small sets, of classes, and of sets, observing that the
"sets" include the small sets and the classes, as well as many other sets
such as $\mathcal{P}(U), \mathcal{P}(\mathcal{P}(U)), \{U\}$, and the like. Note, in particular, that $\{U\}$ is
a set which has only one element (namely, the universe $U$). It is thus
intuitively very "small", but it is not a small set in our sense: $\{U\} \in U$
would imply $U \in U$, a contradiction to the axiom of regularity. Thus
"small set" for us means a member of the universe, and not a set with a
small cardinal number.

Our foundation by means of one universe does provide, within set
theory, an accurate way of discussing the category of all small sets and all
small groups, but it does not provide sets to represent certain meta-
categories, such as the metacategory of all sets or that of all groups.

Grothendieck uses an alternative device. He assumes that for every set $X$
there is a universe $U$ with $X \in U$. This stronger assumption evidently
provides for each universe $U$ a category of all those groups which are
members of $U$. However, this does not provide any category of all
groups. For this reason, there has been considerable discussion of a
foundation for category theory (and for all of Mathematics) not based
on set theory. This is why we initially gave the definition of a category $C$
in a set-free form, simply by regarding the axioms as first-order axioms
on undefined terms "object of $C$", "arrow of $C$", "composite", "identity",
"domain", and "codomain". In this style, Lawvere [1964] has given
axioms for the elementary (i.e., first-order) theory of the category of all
sets, as an alternative to the usual axioms on membership.

Exercises

1. Given a universe $U$ and a function $f : I \to U$ with domain $i \in U$ and with every
   value $f_i$ an element of $U$, for $i \in I$, prove that the usual cartesian product $\Pi_i f_i$
is an element of $U$.

2. (a) Given a universe $U$ and a function $f : I \to U$ with domain $i \in U$, show that
   the usual union $\cup f_i$ is a set of $U$.

   (b) Show that this one closure property of $U$ may replace condition (i) and the
   condition $x \in U$ implies $\cup x \in U$ in the definition of a universe.

7. Large Categories

In many relevant examples, a category consists of all (small) Mathematical
objects with a given structure, with arrows all the functions which
preserve that structure. We list useful such examples.

Ab, the category of all small abelian groups, has objects all small
(additive) abelian groups $A, B, \ldots$ and arrows all homomorphisms
$f : A \to B$ of abelian groups, with the usual composition. In this category,
an arrow is monic if and only if it is a monomorphism (one-one into).

Also, an epimorphism (a homomorphism onto) is clearly epi. Conversely,
a homomorphism $f : A \to B$ which is epi as an arrow must be onto
as a function. For, otherwise, the quotient group $B/fA$ is nonzero, so
there are then two different morphisms $B \to B/fA$, the projection $p$
and the zero morphism $0$, which have $p f = 0 = 0 f$, a contradiction to
the assumption that $f$ is epi. In Ab, the zero group is both initial and terminal.

A small ring $R$ is a small set with binary operations of addition and
multiplication which satisfy the usual axioms for a ring — including the
existence of a two-sided identity (= unit 1) for multiplication. Ring
will denote the category of all small rings: the objects are the small
rings $R$, the arrows $f : R \to S$ the (homomorphisms of rings — where
a morphism of rings is assumed to carry the unit of $R$ to that of $S$.
In this category the zero ring is terminal, and the ring $\mathbb{Z}$
of integers is initial since $\mathbb{Z} \to R$ is the unique arrow carrying 1 to $R$.
The monic arrows are precisely the monomorphisms of rings.
Every epimorphism of rings is epi as an arrow, but the inclusion $\mathbb{Z} \to \mathbb{Q}$
of $\mathbb{Z}$ in the field $\mathbb{Q}$ of rational numbers is epi, but not an epimorphism.

If $R$ is any small ring, the category $R$-Mod has objects all small
left $R$-modules $A, B, \ldots$ and arrows $f : A \to B$ all morphisms of $R$-
modules (left $R$-linear maps). In this category monics are epimorphisms,
epis are epimorphisms, and the zero module is initial and terminal.
If $F$ is a field the category $F$-Mod, also written Vect, is that of all vector
spaces (linear spaces) over $F$. By $Mod R$ we denote the category of all
small right $R$-modules. If $R$ and $S$ are two rings, $R$-Mod-$S$ is the category
of all small $R$-$S$-bimodules (left $R$, right $S$-modules $A$ with $r(a) = (ra)s$
for all $r \in R, a \in A$, and $s \in S$). One may similarly construct categories of
small algebraic objects of any given type.

The category Top of topological spaces has as objects all small
topological spaces $X, Y, \ldots$ and as morphisms all continuous maps
$f : X \to Y$. Again, the monics are the injections and the epis the surjections.
The one-point space is terminal, and the empty space is initial. Similarly,
one may form the category of all small Hausdorff spaces or of all small
compact Hausdorff spaces.

The category Top has as objects all small topological spaces $X, Y, \ldots$,
while a morphism $\alpha : X \to Y$ is a homotopy class of continuous maps
$f : X \to Y$; in other words, two homotopic maps $f \simeq g : X \to Y$
determine the same morphism from $X$ to $Y$. The composition of morphisms is
the usual composition of homotopy classes of maps. In this category, the
homotopy class of an injection need not be a monic, as one may see, for
example, for the injection of a circle into a disc (as the bounding circle
of that disc). This category Top, which arises naturally in homotopy
theory, shows that an arrow in a category need not be the same thing
as a function. There are a number of other categories which are useful
in homotopy theory: For example, the categories of $CW$-complexes,
of simplicial sets, of compactly generated spaces (see § VII.8), and of Kan complexes.

\textbf{Set}_* will denote the category of small pointed sets (often called “based” sets). By a pointed set is meant a nonvoid set \( P \) with a selected element, written \( \ast \) or \( * \), and called the “base point” of \( P \). A map \( f: P \to Q \) of pointed sets is a function on the set \( P \) to the set \( Q \) which carries base point to base point; i.e., which satisfies \( f(\ast) = \ast \). The pointed sets with these maps as morphisms constitute the category \( \text{Set}_* \). In this category the set \( \{\ast\} \) with just one point (the base point) is both an initial and a terminal object. A morphism \( f \) is monic in \( \text{Set}_* \) if and only if it has a left inverse, epi if and only if it has a right inverse, and invertible if and only if it is both monic and epic.

Similarly, \( \text{Toph}_* \) denotes the category of small pointed topological spaces: the objects are spaces \( X \) with a designated base point \( \ast \); the morphisms are continuous maps \( f: X \to Y \) which send the base point of \( X \) to that of \( Y \). Again, \( \text{Toph}_* \) is the category with objects pointed spaces and morphisms homotopy classes of continuous base-point-preserving maps (where also the homotopies are to preserve base points). Both categories arise in homotopy theory, where the choice of a base point is always needed in defining the fundamental group or higher homotopy groups of a space.

Binary relations can be regarded as the arrows of a category \( \text{Rel} \). The objects are all small sets \( X, Y, \ldots \), and the arrows \( R: X \to Y \) are the binary relations on \( X \) to \( Y \); that is, the subsets \( RX \times Y \). If \( S: Y \to Z \) is another such relation, the composite relation \( S \circ R: X \to Z \) is defined to be

\[ S \circ R = \{ (x, z) \mid \text{for some } y \in Y, \langle x, y \rangle \in R \text{ and } \langle y, z \rangle \in S \}. \]

The identity arrow \( X \to X \) is the identity relation on \( X \), consisting of all \( \langle x, x \rangle \) for \( x \in X \). The axioms for a category evidently hold. This category \( \text{Rel} \) contains \( \text{Set} \) as a subcategory on the same objects, where each function \( f: X \to Y \) is interpreted as the relation consisting of all pairs \( \langle x, f(x) \rangle \) for \( x \in X \). But \( \text{Rel} \) has added structure: For each \( R: X \to Y \) there is a converse relation \( R': Y \to X \) consisting of all pairs \( \langle y, x \rangle \) with \( \langle x, y \rangle \in R \).

A concrete category is a pair \( (C, U) \) where \( C \) is a category and \( U \) a faithful functor \( U: C \to \text{Set} \). Since \( U \) is faithful, we may identify each arrow \( f \) of \( C \) with the function \( Uf \). In these terms, a concrete category may be described as a category \( C \) in which each object \( c \) comes equipped with an “underlying” set \( UC \), each arrow \( f: b \to c \) is an actual function \( Uf: UB \to UC \), and composition of arrows is composition of functions. Many of the explicit large categories described above are concrete categories in this sense, relative to the evident forgetful functor \( U \), but this is not so for \( \text{Toph} \) or for \( \text{Rel} \). For the applications, the notion of category is simpler (and more “abstract”) than that of concrete category.

\section{Hom-sets}

For objects \( a \) and \( b \) in the category \( C \) the hom-set

\[ \text{hom}_C(a, b) = \{ f \mid f \text{ is an arrow } f: a \to b \text{ in } C \} \]

consists of all arrows of the category with domain \( a \) and codomain \( b \). The notation for this set is frequently and variously abbreviated as \( \text{hom}_C(a, b) = C(a, b) = \text{hom}(a, b) = (a, b) \).

A category may be defined in terms of hom-sets as follows. A category is given by the following data:

(i) A set of objects \( a, b, c, \ldots \).

(ii) A function which assigns to each ordered pair \( \langle a, b \rangle \) of objects a set \( \text{hom}(a, b) \).

(iii) For each ordered triple \( \langle a, b, c \rangle \) of objects a function

\[ \text{hom}(b, c) \times \text{hom}(a, b) \to \text{hom}(a, c), \]

called composition, and written \( \langle g, f \rangle \mapsto g \circ f \) for \( g \in \text{hom}(b, c), f \in \text{hom}(a, b) \).

(iv) For each object \( b \), an element \( 1_b \in \text{hom}(b, b) \), called the identity of \( b \).

These data are required to satisfy the familiar associativity and unit axioms (1.1) and (1.2), plus an added “disjointness” axiom:

(v) If \( \langle a, b \rangle \in \text{hom}(a, b) \cap \text{hom}(a', b) \), then \( \text{hom}(a, b) \cap \text{hom}(a', b) = \emptyset \), where \( \emptyset \) is the empty set.

In particular, the associativity axiom may be restated as the requirement that the following diagram, with each arrow given in the evident way by composition, be a commutative diagram:

\[
\begin{array}{ccc}
\text{hom}(c, d) \times \text{hom}(b, c) \times \text{hom}(a, b) & \to & \text{hom}(b, d) \times \text{hom}(a, b) \\
\downarrow & & \downarrow \\
\text{hom}(d, c) \times \text{hom}(a, c) & \to & \text{hom}(a, d)
\end{array}
\]

This definition of a category is equivalent to the original definition of § 2. Axiom (v) above requires that “distinct” hom-sets be disjoint; it is included to insure that each arrow have a definite domain and a definite codomain. Should this axiom fail in an example, it can be readily reinstated by adjusting the hom-sets so that they do become disjoint. For example, we can replace each original set \( \text{hom}(a, b) \) by the set \( \{ a \} \times \text{hom}(a, b) \times \{ b \} \); this amounts to “labeling” each \( f \in \text{hom}(a, b) \) with its domain \( a \) and codomain \( b \). Some authors omit this axiom (v).
A functor $T: C \to B$ may be described in terms of hom-sets as the (usual) object function $T$ together with a collection of functions

$$T_{c,c'}: C(c, c') \to B(Tc, Tc')$$

(namely, the functions $f \mapsto Tf$, for $f \in C(c, c')$) such that each $T_{c,c'}1_c = 1_{Tc}$ and such that every diagram

$$
\begin{array}{c}
C(c', c') \times C(c, c') \\
\downarrow T_{c',c'} \times T_{c,c'} \\
B(Tc', Tc') \times B(Tc, Tc') \\
\downarrow T_{c',c'}
\end{array}
\xrightarrow{T_{c,c'}}
\begin{array}{c}
C(c, c') \\
\downarrow T_{c,c'}
\end{array}
\to
\begin{array}{c}
B(Tc, Tc') \\
\downarrow T_{c,c'}
\end{array}
$$

with horizontal arrows the composition in $B$ and $C$, is commutative.

We leave the reader to describe a natural transformation $\tau: S \to T$ in terms of functions $C(c, c') \to B(Sc, Tc')$.

In many relevant examples, the hom-sets of a category themselves have some structure; for instance, in the category of vector spaces $V, W, \ldots$ over a fixed field, each $\text{hom}(V, W)$ is itself a vector space (of all linear transformations $V \to W$). The simplest such case is that in which the hom-sets are abelian groups. Formally, define an $Ab$-category (also called a preadditive category) to be a category $A$ in which each hom-set $A(a, b)$ is an additive abelian group and for which composition is bilinear: For arrows $f, f': a \to b$ and $g, g': b \to c$,

$$(g + g')(f + f') = g \circ f + g \circ f' + g' \circ f + g' \circ f' .$$

Thus $Ab$, $R\text{-Mod}$, $\text{Mod}-R$ and the like are all $Ab$-categories.

Because the composition $\langle g, f \rangle \mapsto g \circ f$ is bilinear,

$$A(b, c) \times A(a, b) \to A(a, c) ,$$

it can also be written (using the tensor product $\otimes = \otimes R$) as a linear map

$$A(b, c) \otimes A(a, b) \to A(a, c) ,$$

and the $Ab$-category $A$ may be described completely in these terms (without assuming ahead of time that it is a category). Thus an $Ab$-category is given by the data

(i) A set of objects $a, b, c, \ldots$;
(ii) A function which assigns to each ordered pair of objects $\langle b, c \rangle$ an abelian group $A(b, c)$;

(iii) For each ordered triple of objects $\langle a, b, c \rangle$ a morphism

$$A(b, c) \otimes A(a, b) \to A(a, c)$$

of abelian groups called composition, and written $g \otimes f \mapsto g \circ f$;

(iv) For each object, a morphism $Z \to A(a, a)$. (Here $Z$ is the additive group of integers; this morphism is completely determined by the image of $1 \in Z$, which may be written as $1_a$.)

These data are required to satisfy the associative and unit laws for composition, stated as in (1.1) and (1.2), or by diagrams. The definition of $Ab$-category is just like the definition of category by hom-sets: Set is replaced by $Ab$, cartesian product $\times$ of sets by tensor product in $Ab$, and the one-point set $\ast$ is replaced by $Z$. There is an evident generalization to categories $A$ which have hom-objects $A(b, c)$ in a category like $Ab$ which is equipped with a multiplication like $\otimes$ and a unit like $Z$ for this multiplication.

If $A$ and $B$ are $Ab$-categories, a functor $T: A \to B$ is said to be additive when every function $T: A(a, a') \to B(Ta, Ta')$ is a homomorphism of abelian groups; that is, when $T(f + f') = Tf + Tf'$ for all parallel pairs $f$ and $f'$. Clearly, the composite of additive functors is additive. $Ab\text{-cat}$ will denote the category of all small $Ab$-categories, with arrows additive functors.

Notes.

These notes, like those at the end of later chapters, are informal remarks on the background and prospects of our subject, with references to the bibliography (for example, H. Pétard [1980b] refers to the second article by Pétard listed for the year 1980).

The fundamental idea of representing a function by an arrow first appeared in topology about 1940, probably in papers or lectures by W. Hurewicz on relative homotopy groups; cf. [1941].

His initiative immediately attracted the attention of R. H. Fox (see Fox [1943]) and N. E. Steenrod, whose [1941] paper used arrows and (implicitly) functors; see also Hurewicz-Steenrod [1941]. The arrow $f: X \to Y$ rapidly displaced the occasional notation $f(X) \subset Y$ for a function. It expressed well a central interest of topology. Thus a notation (the arrow) led to a concept (category).

Commutative diagrams were probably also first used by Hurewicz.

Categories, functors, and natural transformations themselves were discovered by Eilenberg-Mac Lane [1942] in their study of limits (via natural transformations) for universal coefficient theorems in Čech cohomology. In this paper commutative diagrams appeared in print (probably for the first time). Thus Ext was one of the first functors considered. A direct treatment of categories in their own right appeared in Eilenberg-Mac Lane [1945]. Now the discovery of ideas as general as these is chiefly the willingness to make a brash or speculative abstraction, in this case supported by the pleasure of purloining words from the philosophers: "Category"
from Aristotle and Kant, "Functor" from Carnap (Logische Syntax der Sprache), and "natural transformation" from then current informal parlance. Initially, categories were used chiefly as a language, notably and effectively in the Eilenberg-Steenrod axioms for homology and cohomology theories. With recent increasing use, the question of proper foundations has come to the fore. Here experts are still not in agreement; our present assumption of "one universe" is an adequate stopgap, not a forecast of the future.

Category theory asks of every type of Mathematical object: "What are the morphisms?"; it suggests that these morphisms should be described at the same time as the objects. Categorists, however, ordinarily name their large categories by the common name of the objects; thus Set, Cat. Only Ehresmann [1965] and his school have the courage to name each category by the common name of its arrows: our Cat is their category of functors.

II. Constructions on Categories

1. Duality

Categorical duality is the process "Reverse all arrows". An exact description of this process will be made on an axiomatic basis in this section and on a set-theoretical basis in the next section. Hence for this section a category will not be described by sets (of objects and of arrows) and functions (domain, codomain, composition) but by axioms as in § I.1.

The elementary theory of an abstract category (ETAC) consists of certain statements \( \Sigma \) which involve letters \( a, b, c, \ldots \) for objects and letters \( f, g, h, \ldots \) for arrows. These statements are the ones built up from the atomic statements which involve the usual undefined terms of category theory; thus, atomic statements are "\( a \) is the domain of \( f \)" , "\( b \) is the codomain of \( f \)" , "\( i \) is the identity arrow of \( a \)" , and "\( g \) can be composed with \( f \) and \( h \) is the composite" , "\( a = b \)" and "\( f = g \)". These atomic statements can also be written as equations in the familiar way: "\( a = \text{dom } f \)" , "\( h = g \cdot f \)". A statement \( \Sigma \) is defined to be any phrase (well formed formula) built up from the types of atomic statements listed above in the usual fashion by means of the ordinary propositional connectives (and, or, not, implies, if and only if) and the usual quantifiers ("for all \( a \)", "for all \( f \)", "there exists an \( a \) ...", "there exists an \( f \) ... "). Thus "\( f: a \rightarrow b \)" is the abbreviation we have adopted for the statement, "\( a \) is the domain of \( f \) and \( b \) is the codomain of \( f \)".

A sentence is a statement with all variables quantified (i.e., all variables are "bound", none being "free"). For example, "for all \( f \) there exist \( a \) and \( b \) with \( f: a \rightarrow b \)" is a sentence (one which in fact is an axiom, true in every category). The axioms of ETAC (as given in § I.1) are certain such sentences.

The dual of any statement \( \Sigma \) of ETAC is formed by making the following replacements throughout in \( \Sigma \): "domain" by "codomain", "codomain" by "domain" and "\( h \) is the composite of \( g \) with \( f \)" by "\( h \) is the composite of \( f \) with \( g \)"; arrows and composites are reversed. Logic (and, or, ...) is unchanged. This gives the following table (a more extensive table appears in Exercise IV.3.1).
Statement \( \Sigma \) | Dual statement \( \Sigma^* \)  
--- | ---  
\( f: a \rightarrow b \) | \( f: b \rightarrow a \)  
\( a = \text{dom } f \) | \( a = \text{cod } f \)  
\( i = 1_a \) | \( i = 1_b \)  
\( h = g \cdot f \) | \( h = f \cdot g \)  
\( f \) is monic | \( f \) is epic  
\( u \) is a right inverse of \( h \) | \( u \) is a left inverse of \( h \)  
\( f \) is invertible | \( f \) is invertible  
\( t \) is a terminal object | \( t \) is an initial object.

Note that the dual of the dual is the original statement (\( \Sigma^{**} = \Sigma \)). If a statement involves a diagram, the dual statement involves that diagram with all arrows reversed.

The dual of each of the axioms for a category is also an axiom. Hence in any proof of a theorem about an arbitrary category from the axioms, replacing each statement by its dual gives a valid proof (of the dual conclusion). This is the duality principle: If a statement \( \Sigma \) of the elementary theory of an abstract category is a consequence of the axioms, so is the dual statement \( \Sigma^* \). For example, we noted the (elementary) theorem that a terminal object of a category, if it exists, is unique up to isomorphism. Therefore we have the dual theorem: An initial object, if it exists, is unique up to isomorphism. For more complicated theorems, the duality principle is a handy way to have (at once) the dual theorem. No proof of the dual theorem need be given. We usually leave even the formulation of the dual theorem to the reader.

The duality principle also applies to statements involving several categories and functors between them. The simplest (and typical) case is the elementary theory of one functor; i.e., of two categories \( C \) and \( B \) and a functor \( T: C \rightarrow B \). For this theory, the atomic statements are those listed above for the category \( C \), a corresponding list for the category \( B \), as well as the statements “\( Tc = c' \)” or “\( Tf = f' \)” giving the values of the object and arrow functions of \( T \) on objects \( c \) and arrows \( f \) of \( C \). The axioms include the axioms for a category for \( C \) and for \( B \) also and the statements \( T(gf) = (Tg)(Tf) \) and \( T(1_c) = 1_{Tc} \) which assert that \( T \) is a functor. The dual of a statement is formed by simultaneously dualizing the atomic parts referring to \( C \) and to \( B \) (i.e., reversing arrows in \( C \) and in \( B \)). Since the statement that \( T \) is a functor is self-dual, the duality principle above is still true.

We emphasize that duality for a statement involving several categories and functors between them reverses the arrows in each category but does not reverse the functors.

---

2. Contravariance and Opposites

To each category \( C \) we also associate the opposite category \( C^\circ \). The objects of \( C^\circ \) are the objects of \( C \), the arrows of \( C^\circ \) are arrows \( f^\circ \) in one-one correspondence \( f \rightarrow f^\circ \) with the arrows of \( C \). For each arrow \( f: a \rightarrow b \) of \( C \), the domain and codomain of the corresponding \( f^\circ \) are as in \( f^\circ: b \rightarrow a \) (the direction is reversed). The composite \( f^\circ g^\circ \) is defined in \( C^\circ \) exactly when the composite \( gf \) is defined in \( C \). This clearly makes \( C^\circ \) a category. Moreover, if \( f \) is the codomain of \( f^\circ \) is monic if and only if \( f \) is epic, and so on. Indeed, this process translates any statement \( \Sigma \) about \( C \) into the dual statement \( \Sigma^* \) about \( C^\circ \). In detail, an evident induction on the construction of \( \Sigma \) from atomic statements proves that if \( \Sigma \) is any statement with free variables \( f, g, \ldots \) in the elementary theory of an abstract category, then \( \Sigma \) is true for arrows \( f, g, \ldots \) of a category \( C \) if and only if if the dual statement \( \Sigma^* \) is true for the arrows \( f^\circ, g^\circ, \ldots \) of the opposite category \( C^\circ \). In particular, a sentence \( \Sigma \) is true in \( C^\circ \) if and only if if the dual sentence \( \Sigma^* \) is true in \( C \). This observation allows us to interpret the dual of a property \( \Sigma \) as the original property applied to the opposite category (some authors call \( C^\circ \) the “dual” category, and write it \( C^\circ = C^* \)).

If \( T: C \rightarrow B \) is a functor, its object function \( c \mapsto Tc \) and its mapping function \( f \mapsto Tf \), rewritten as \( f^\circ \mapsto (Tf)^\circ \), together define a functor from \( C^\circ \) to \( B^\circ \), which we denote as \( T^\circ: C^\circ \rightarrow B^\circ \). The assignments \( C \mapsto C^\circ \) and \( T \mapsto T^\circ \) define a (covariant) functor \( \text{Cat} \rightarrow \text{Cat} \).

Consider a functor \( S: C^\circ \rightarrow B \). By the definition of a functor, it assigns to each object \( c \in C^\circ \) an object \( Sc \) of \( B \) and to each arrow \( f^\circ: b \rightarrow a \) of \( C^\circ \) an arrow \( Sf^\circ: Sa \rightarrow Sb \) of \( B \), with \( S(f^\circ g^\circ) = (Sf^\circ)(Sg^\circ) \) whenever \( f^\circ g^\circ \) is defined. The functor \( S \) so described may be expressed directly in terms of the original category \( C \) if we write \( Sf \) for \( Sf^\circ \) then \( S \) is a contravariant functor on \( C \) to \( B \), which assigns to each object \( c \in C \) an object \( Sc \in B \) and to each arrow \( f: a \rightarrow b \) an arrow \( Sf: Sa \rightarrow Sb \) (in the opposite direction), all in such a way that

\[
S(1_c) = 1_{Sc}, \quad S(fg) = (Sf)(Sg),
\]

the latter whenever the composite \( fg \) is defined in \( C \). Note that the arrow function \( S \) of a contravariant functor inverts the order of composition. Specific examples of contravariant functors may be conveniently presented in this form; i.e., as functions \( S \) inverting composition. An example is the contravariant power-set functor \( \mathbb{P} \) on \( \text{Set} \) to \( \text{Set} \). For each set \( X \), \( \mathbb{P}X = \{ S \mid S \subseteq X \} \) is the set of all subsets of \( X \); for each function \( f: X \rightarrow Y \), \( \mathbb{P}f: \mathbb{P}Y \rightarrow \mathbb{P}X \) sends each subset \( T \subseteq Y \) to its inverse image \( f^{-1}T \subseteq X \). Another example is the familiar process which assigns to each vector space \( V \) its dual (conjugate) vector space \( V^\circ \) and to each linear transformation \( f: V \rightarrow W \) its dual \( f^\circ: W^\circ \rightarrow V^\circ \); these assignments describe a
contravariant functor on the category of all vector spaces (over a fixed field) to itself.

To contrast, a functor \( T: C \to B \) as previously defined, in § I.3, is called a covariant functor on \( C \) to \( B \). For general discussions it is much more convenient to represent a contravariant functor \( S \) on \( C \) to \( B \) as a covariant functor \( S^\circ: C^\circ \to B \), or sometimes as a covariant functor \( S^\circ: C \to B^\circ \). In this book an arrow between (symbols for) categories will always denote a covariant functor \( T: C \to B \) or \( S: C^\circ \to B \) between the designated categories.

Hom-sets provide an important example of co- and contravariant functors. Suppose that \( C \) is a category with small hom-sets, so that each \( \text{hom}(a, b) = \{ f: a \to b \text{ in } C \} \) is a small set, hence an object of the category \( \text{Set} \) of all small sets. Thus we have for each object \( a \in C \) the covariant \( \text{hom-} \text{-functor} \)

\[ C(a, -) = \text{hom}(a, -): C \to \text{Set} \tag{2} \]

its object function sends each object \( b \) to the set \( \text{hom}(a, b) \); its arrow function sends each arrow \( k: b \to b' \) to the function

\[ \text{hom}(a, k): \text{hom}(a, b) \to \text{hom}(a, b') \tag{3} \]

defined by the assignment \( f \mapsto k \circ f \) for each \( f: a \to b \). To simplify the notation, this function \( \text{hom}(a, k) \) is sometimes written \( k_* \) and called “composition with \( k \) on the left”, or “the map induced by \( k \).”

The contravariant \( \text{hom-} \text{-functor}, \) for each object \( b \in C \), will be written covariantly, as

\[ C(-, b) = \text{hom}(-, b): C^\circ \to \text{Set} \tag{4} \]

it sends each object \( a \) to the set \( \text{hom}(a, b) \), and each arrow \( g: a \to a' \) of \( C \) to the function

\[ \text{hom}(g, b): \text{hom}(a', b) \to \text{hom}(a, b) \tag{5} \]

defined by \( f \mapsto f \circ g \). Omitting the object \( b \), this function \( \text{hom}(g, b) \) is sometimes written simply as \( g^* \) and called “composition with \( g \) on the right”. Thus, for each \( f: a' \to b \),

\[ k_* f = k \circ f, \quad g^* f = f \circ g. \]

For two arrows \( g: a \to a' \) and \( k: b \to b' \) the diagram

\[ \begin{array}{ccc}
\text{hom}(a', b') & \xrightarrow{g^*} & \text{hom}(a, b) \\
\downarrow k_* & & \downarrow k_* \\
\text{hom}(a', b) & \xrightarrow{f^*} & \text{hom}(a, b')
\end{array} \tag{6} \]

in \( \text{Set} \) is commutative, because both paths send \( f \in \text{hom}(a', b) \) to \( kfg \).

These \( \text{hom-} \text{-functors have been defined only for a category } C \text{ with small hom-sets. The familiar large categories } \text{Grp}, \text{Set}, \text{Top}, \text{etc. do have this property. To include categories without this property, we can proceed as follows: Given a category } C \text{, take a set } V \text{ large enough to include all subsets of the set of arrows of } C \text{ (for example, } V \text{ could be the power set of the set of arrows of } C \text{). Let } \text{Ens} = \text{Set}, \text{ be the category with objects all sets } X \in V, \text{ arrows all functions } f: X \to Y \text{ between two such sets and composition the usual composition of functions. Then each hom-set } C(a, b) = \text{hom}(a, b) \text{ is an object of this category } \text{Ens}, \text{ so the above procedure defines two hom-} \text{-functors}

\[ C(a, -): C \to \text{Ens}, \quad C(-, b): C^\circ \to \text{Ens}. \tag{7} \]

In particular, when \( V \) is the universe of all small sets, \( \text{Ens} = \text{Set} \); in general, \( \text{Ens} \) is a (variable) category of sets which acts as a receiving category for the hom-functors of a category or categories of interest.

There are many other examples of contravariant functors. For \( X \) a topological space, the set \( \text{Open}(X) \) of all open subsets \( U \) of \( X \), when ordered by inclusion, is a partial order and hence a category; there is an arrow \( V \to U \) precisely when \( V \subset U \). Let \( \mathcal{C}(U) \) denote the set of all continuous real-valued functions \( h: U \to \mathbb{R} \); the assignment \( h \mapsto h|_V \) restricting each \( h \) to the subset \( V \) is a function \( \mathcal{C}(U) \to \mathcal{C}(V) \) for each \( V \subset U \). This makes \( \mathcal{C} \) a contravariant functor on \( \text{Open}(X) \) to \( \text{Set} \). This functor is called the sheaf of germs of continuous functions on \( X \). On a smooth manifold, the sheaf of germs of \( C^\infty \)-differentiable functions is constructed in similar fashion.

\( \text{Mod-R} \) is a contravariant functor from rings \( R \) to categories. Specifically, if \( g: R \to S \) is any morphism of (small) rings, each right \( S \)-module \( B \) becomes a right \( R \)-module \( B_0 = (\text{Mod}_R)B \) by “pull-back” along \( g \). Each \( r \in R \) acts on \( b \in B \) by \( b \cdot r = \cdot (g(r)) \). Clearly \( \text{Mod}_R \) is a functor \( \text{Mod-S} \to \text{Mod-R} \), and \( \text{Mod}(g_1, g_2) = (\text{Mod}_R(g_2), \text{Mod}_R(g_1)) \), so \( \text{Mod} \) itself can be regarded as a contravariant functor on \( \text{Rng} \to \text{Cat} \), the category of all large categories.

One may also form the category \( \text{Mod} \) of all (right) modules over all rings. An object of \( \text{Mod} \) is a pair \( (R, A) \), where \( R \) is a small ring and \( A \) a small right \( R \)-module. A morphism \( (R, A) \to (S, B) \) is a pair \( (\phi, f) \), where \( \phi: R \to S \) is a morphism of rings and \( f: A \to (\text{Mod}_R)B \) is a morphism of right \( R \)-modules. With the evident composition, this yields a category \( \text{Mod} \). A projection functor \( \text{Mod-} \to \text{Rng} \) is given by \( (R, A) \to R \). Further study of the relation of this functor to the previous functor \( \text{Rng-} \to \text{Cat} \) leads to the theory of fibered categories. (\( \text{Mod} \) is fibered over \( \text{Rng} \), the fiber over each \( R \) being the category \( \text{Mod-R} \)).
3. Products of Categories

From two given categories $B$ and $C$ we construct a new category $B \times C$, called the product of $B$ and $C$, as follows. An object of $B \times C$ is a pair $\langle b, c \rangle$ of objects $b$ of $B$ and $c$ of $C$; an arrow $\langle b, c \rangle \rightarrow \langle b', c' \rangle$ of $B \times C$ is a pair $\langle f, g \rangle$ of arrows $f : b \rightarrow b'$ and $g : c \rightarrow c'$, and the composite of two such arrows

$$\langle b, c \rangle \rightarrow \langle b', c' \rangle \rightarrow \langle b'', c'' \rangle$$

is defined in terms of the composites in $B$ and $C$ by

$$\langle f', g' \rangle \circ \langle f, g \rangle = \langle f' \circ f, g' \circ g \rangle.$$  

(1)

Functors

$$P : B \times C \rightarrow B, \quad Q : B \times C \rightarrow C,$$

called the projections of the product, are defined on (objects and) arrows by

$$P \langle f, g \rangle = f, \quad Q \langle f, g \rangle = g.$$  

They have the following property: Given any category $D$ and two functors

$$B \times D \rightarrow D, C,$$

there is a unique functor $F : D \rightarrow B \times C$ with $PF = R, QF = T$; explicitly, these two conditions require that $Fh$, for any arrow $h$ in $D$, must be $\langle Rh, Th \rangle$; conversely, this value for $Fh$ does make $F$ a functor with the required properties. The construction of $F$ may be visualized by the following commutative diagram of functors:

$$\begin{array}{ccc}
D & \xrightarrow{R} & B \\
\downarrow{F} \quad \qquad \quad \downarrow{P} \quad \qquad \quad \downarrow{Q} \quad \qquad \quad \downarrow{C} \\
B \times D & \rightarrow & B \times C \\
\end{array}$$  

(2)

This property of the product category states that the projections $P$ and $Q$ are "universal" among pairs of functors to $B$ and $C$. It is exactly like a similar property of the projections from the (cartesian) product of two sets, two groups, or two spaces. The general properties of such products in any category will be considered in Chapter III.

Two functors $U : B \rightarrow B'$ and $V : C \rightarrow C$ have a product $U \times V : B \times C \rightarrow B' \times C$ which may be defined explicitly on objects and arrows as

$$\langle U \times V \rangle \langle h, c \rangle = \langle Uh, Vc \rangle, \quad \langle U \times V \rangle \langle f, g \rangle = \langle Uf, Vg \rangle.$$
Applying the functor $S$ to this equation gives

$$S(b', g) S(f, c) = S(f, c') S(b, g);$$

as a commutative diagram

$$
\begin{array}{ccc}
S(b, c) & \xrightarrow{S(f, g)} & S(b, c') \\
S(f, c) & \xrightarrow{S(b', g)} & S(f, c').
\end{array}
$$

This is just condition (4) rewritten, so that condition (4) is necessary. Conversely, given all $L_\ast$ and $M_\ast$, this condition defines $S(f, g)$ for every pair $f, g$; it may be verified that this definition does yield a bifunctor $S$ with the required properties.

One may also form products of three or more categories, or combine the construction of product categories and opposite categories. There is an evident isomorphism $(B \times C)^{op} \cong C^{op} \times B^{op}$. A functor $B^{op} \times C \to D$ is often called a bifunctor, contravariant in $B$ and covariant in $C$, with values in $D$. For example, if $C$ is a category with small hom-sets, the hom-sets define such a bifunctor

$$
\text{hom} : C^{op} \times C \to \text{Set}.
$$

Indeed, the commutative diagram (6) of §2 shows exactly that the co- and contravariant hom-functors

$$
\text{hom}(\cdot, \cdot) : C^{op} \to \text{Set}, \quad \text{hom}(\cdot, \cdot) : C \to \text{Set}
$$
do satisfy the condition (4) of the theorem, necessary to make hom a bifunctor.

Next consider natural transformations between bifunctors $S, S' : B \times C \to D$. Let $\alpha$ be a function which assigns to each pair of objects $b \in B, c \in C$ an arrow

$$
\alpha(b, c) : S(b, c) \to S'(b, c)
$$

in $D$. Call $\alpha$ natural in $b$ if for each $c \in C$ the components $\alpha(b, c)$ for all $b$ define

$$a(\cdot, c) : \alpha(\cdot, c) \to S'(\cdot, c),$$

a natural transformation of functors $B \to D$. The reader may readily prove the useful result:

**Proposition 2.** For bifunctors $S, S'$, the function $\alpha$ displayed in (5) is a natural transformation $\alpha : S \to S'$ (i.e., of bifunctors) if and only if $\alpha(b, c)$ is natural in $b$ for each $c \in C$ and natural in $c$ for each $b \in B$.

Such natural transformations appear in the fundamental definition of adjoint functors (Chapter IV). A functor $F : X \to C$ is the left adjoint

of a functor $G : C \to X$ (opposite direction) when there is a bijection

$$\text{hom}_C(Fx, c) \cong \text{hom}_X(x, Gc)$$
natural in $x \in X$ and $c \in C$. Here $\text{hom}_C(F \cdot, -)$ is a bifunctor, the composite

$$X^{op} \times C \xrightarrow{F^{op} \times 1D} C^{op} \times C \xrightarrow{\text{hom}_C} \text{Set},$$

and $\text{hom}_X(-, G \cdot)$ similarly (at least when $X$ and $C$ have small hom-sets).

The product category can be visualized in the case $C \times 2$, where 2 is the category with one non-identity arrow $0 \to 1$; explicitly $C \times 2$ consists of two copies $C \times 0$ and $C \times 1$ of $C$ with arrows joining the first to the second, as in the figure ("diagonal" arrows omitted) for $C = 3$:

$$
\begin{array}{ccc}
C \times 1 & \xrightarrow{f} & C \times 0 \\
\downarrow & & \downarrow \\
C \times 1 & \xrightarrow{g} & C \times 0
\end{array}
$$

Here the functors $T_0, T_1 : C \to C \times 2$ ("bottom" and "top", respectively) are defined for each arrow $f$ of $C$ by $T_0 f = \langle f, 0 \rangle$ and $T_1 f = \langle f, 1 \rangle$. If $\downarrow$ denotes the unique non-identity arrow $0 \to 1$ of 2, then we may define a transformation between $T_0, T_1 : C \to C \times 2$ by

$$
\mu : T_0 \to T_1, \quad \mu c = \langle c, \downarrow \rangle,
$$

for any object $c$. It maps "bottom" to "top" and is clearly natural. We call $\mu$ the universal natural transformation from $C$ for the following reason. Given any natural transformation $\tau : S \to T$ between $S, T : C \to B$, there is a unique functor $F : C \times 2 \to B$ with $F \mu c = \tau c$ for any object $c$. Specifically, $F$ is, when $f : c \to c'$,

$$
F\langle f, 0 \rangle = S f, \quad F\langle f, 1 \rangle = T f, \quad F\langle f, \downarrow \rangle = T f \circ \tau c = \tau c' \circ S f.
$$

It may be readily verified that these assignments do define a bifunctor $F : C \times 2 \to B$, and that $F \mu = \tau$.

**Exercises**

1. Show that the product of categories includes the following known special cases:
   - The product of monoids (categories with one object), of groups, of sets (discrete categories).
2. Show that the product of two preorders is a preorder.
3. If \( \{ C_i | i \in I \} \) is a family of categories indexed by a set \( I \), describe the product \( C = \Pi_{i \in I} C_i \), its projections \( P_i : C \rightarrow C_i \), and establish the universal property of these projections.

4. Describe the opposite of the category \( \text{Mat}_k \) of \( \S 1.2 \).

5. Show that the ring of continuous real-valued functions on a topological space is the object function of a contravariant functor on \( \text{Top} \) to \( \text{Ring} \).

4. Functor Categories

Given categories \( C \) and \( B \), we consider all functors \( R, S, T, \ldots : C \rightarrow B \).

If \( \sigma : R \rightarrow S \) and \( \tau : S \rightarrow T \) are two natural transformations, their components for each \( c \in C \) define composite arrows \( (\tau \circ \sigma)_c = \tau_c \circ \sigma_c \) which are the components of a transformation \( \tau \circ \sigma : R \rightarrow T \). To show \( \tau \circ \sigma \) natural, take any \( f : c \rightarrow c' \) in \( C \) and consider the diagram

\[
\begin{array}{ccc}
Rc & \rightarrow & Rc' \\
\downarrow \sigma_c & & \downarrow \sigma_{c'} \\
Sc & \rightarrow & Sc' \\
\downarrow \alpha & & \downarrow \alpha' \\
Tc & \rightarrow & Tc' \\
\end{array}
\]

Since \( \sigma \) and \( \tau \) are natural, both small squares are commutative. Hence the rectangle commutes, so \( \tau \circ \sigma \) is natural.

This composition of transformations is associative; moreover it has for each functor \( T \) an identity, the natural transformation \( 1_T : T \rightarrow T \) with components \( 1_{Tc} = 1_{Te} \). Hence, given the categories \( B \) and \( C \), we may construct formally a functor category \( B^C = \text{Funct}(C, B) \) with objects the functors \( T : C \rightarrow B \) and morphisms the natural transformations between two such functors. It is often suggestive to write

\[
\text{Nat}(S, T) = B^C(S, T) = \{ \tau : S \rightarrow T \text{ natural} \}
\]

for the “hom-set” of this category. It need not be a small set.

Functor categories will be used extensively. For example, if \( B \) and \( C \) are sets (categories with all arrows identities), then \( B^C \) is also a set; namely, the familiar “function-set” consisting of all functions \( C \rightarrow B \).

In particular, for \( B = \{ 0, 1 \} \) a two-point set, \( \{ 0, 1 \}^C \) is (isomorphic to) the set of all subsets of \( C \) (the “power set” \( \mathcal{P} C \)). For any category \( B \), \( B^1 \) is isomorphic to \( B \), while \( B^2 \) is called the category of arrows of \( B \); its objects are arrows \( f : a \rightarrow b \) of \( B \), and its arrows \( f \rightarrow f' \) are those pairs

\[
\langle h, k \rangle \text{ of arrows in } B \text{ for which the square}
\]

\[
\begin{array}{ccc}
a & \rightarrow & a' \\
\downarrow f & & \downarrow f' \\
b & \rightarrow & b'
\end{array}
\]

commutes. If \( M \) is a monoid (category with one object) \( \text{Set}^M \) is the category with objects the actions of \( M \) (on some set) and arrows the morphisms of such actions. An object of the functor category \( \text{Grp}^M \) is a group with operators \( M \).

If \( K \) is a commutative ring and \( G \) a group, then the functor category \( (K-\text{Mod})^G \) is the category of (\( K \)-linear) representations of \( G \). Specifically, each functor \( T : G \rightarrow K-\text{Mod} \) is determined by a \( K \)-module \( V \) (the image of the single object of the category \( G \)) and a morphism \( T : G \rightarrow \text{Aut}(V) \) of groups (a representation of \( G \) by linear transformations \( V \rightarrow V \)). If \( T' \) is a second such representation, a natural transformation \( \sigma : T \rightarrow T' \) is given by a single arrow \( \sigma : V \rightarrow V' \) (its component at the single object of \( G \)) such that the diagram

\[
\begin{array}{ccc}
V & \rightarrow & V' \\
\downarrow \sigma & & \downarrow \sigma' \\
V & \rightarrow & V'
\end{array}
\]

commutes for every \( g \in G \). In representation theory, such a \( \sigma \) is called an intertwining operator. Thus \( (K-\text{Mod})^G \) is the category with objects the representations of \( G \) and morphisms the intertwining operators.

Next we consider the “size” of functor categories. Since every category \( B \) or \( C \) is essentially a set of morphisms, within our set theory one can always form the set of all functors \( C \rightarrow B \) and the set of all natural transformations between two such functors. Hence the functor category \( B^C \) always exists, but it can be “larger” than \( B \) and \( C \). Recall that a small set is an element of the (fixed) universe, and a large set (a class) is a subset of the universe. We will show:

- If \( B \) and \( C \) are both small categories, so is \( B^C \);
- If the category \( B \) is a class and \( C \) is small, \( B^C \) is large;
- If \( B \) has small hom-sets and \( C \) is small, \( B^C \) has small hom-sets.

The first is evident. Next consider the second. Since \( B \) and \( C \) are fixed, each functor \( T : C \rightarrow B \) is determined by its arrow function \( T \) : \text{Arr} C \rightarrow \text{Arr} B \). But \( C \) is small, so \( \text{Arr} C \) is a small set, as is therefore the image of \( \text{Arr} C \) under \( T \) and therefore the representation of \( T \) as a set of ordered pairs. With each functor \( T \) represented by a small set in this way, one may form the set of all these small sets. This set (the set of objects
of $B^c$ will be a subset of the universe, hence a class. Similarly, the set of all arrows in $B^c$ is a class. Therefore $B^c$ is large, as asserted.

Consider the third assertion. Given $S$ and $T$, each natural transformation $\tau: S \Rightarrow T$ is determined by the usual function $c \mapsto \tau_c$,

$$\tau: \text{Obj} C \rightarrow \bigcup B(S_c, T_c).$$

where the union is taken over the small set of all objects $c \in C$. But $B$ has small hom-sets, so each $B(-,-)$ is small, as is their union. Therefore the set of all these functions $\tau$ is a small set, so $B^c$ has small hom-sets.

When the category $C$ is large, the functor category $B^c$ need not be a subset of the universe. For example, if $B = \{0, 1\}$ is the set with just two elements, while $C$ is the set $U$, then a functor $U \rightarrow B$ is just a function on $U$ to a set with two elements. The possible such functions correspond (as characteristic functions) to the possible subsets of $U$. Therefore the set of objects in $\{0, 1\}^U$ is equivalent to the set $\mathcal{P}(U)$ of all subsets of $U$, and this set has a larger cardinal number than $U$.

Exercises
1. For $R$ a ring, describe $R$-Mod as a full subcategory of the functor category $\text{Ab}^R$.
2. Describe $B^c$, for $X$ a finite set (a finite discrete category).
3. Let $\mathbb{N}$ be the discrete category of natural numbers. Describe the functor category $\text{Ab}^\mathbb{N}$ (commonly known as the category of graded abelian groups).
4. If $P$ and $Q$ are preorders, describe the functor category $Q^P$ and show that it is a preorder.
5. If $\text{Fin}$ is the category of all finite sets and $G$ is a finite group, describe $\text{Fin}^G$ (the category of all permutation representations of $G$).
6. Let $\mathbb{M}$ be the infinite cyclic monoid (elements $1, m, m^2, \ldots$). In the functor categories $\text{Matr}(\mathbb{M})^\mathbb{N}$ and $\text{Matr}(\mathbb{M})^\mathbb{M}$ show that objects are matrices and isomorphic objects (matrices) are exactly equivalent and similar matrices, respectively, in the usual sense of linear algebra.
7. Given categories $B, C$, and the functor category $B^c$, show that each functor $H: C \rightarrow B^c$ determines two functors $S, T: C \rightarrow B$ and a natural transformation $\tau: S \Rightarrow T$, and show that this assignment $H \mapsto \langle S, T, \tau \rangle$ is a bijection.
8. Relate the functor $H$ of Exercise 7 to $F$ of (3.6).

5. The Category of All Categories
We have defined a "vertical" composite $\tau \cdot \sigma$.

\[
\begin{array}{ccc}
C & \xrightarrow{\tau} & B \\
\downarrow \sigma & & \downarrow \tau \\
C' & \xrightarrow{\sigma} & B' \\
\end{array}
\]

of two natural transformations. There is another "horizontal" composition for natural transformations. Given functors and natural trans-

The Category of All Categories

\[
\begin{array}{ccc}
C & \xrightarrow{S} & B \\
\downarrow I & & \downarrow I \\
T & \xrightarrow{T} & A \\
\end{array}
\]

one may form first the composite functors $S \circ S$ and $T \Rightarrow T$. $C \rightarrow A$ and then construct a square

\[
\begin{array}{ccc}
S & \xrightarrow{S c} & T \\
\downarrow T c & & \downarrow T c \\
S T & \xrightarrow{T c} & T T \\
\end{array}
\]

which is commutative because of the naturality of $\tau$ for the arrows $\tau_c$ of $B$. Now define $(\tau \circ \tau)_c$ to be the diagonal of this square,

$$\tau \circ \tau: S \circ S \Rightarrow T \circ T,$$

To show $\tau \circ \tau: S \circ S \Rightarrow T \circ T$ natural, form

\[
\begin{array}{ccc}
S & \xrightarrow{S c} & T \\
\downarrow T c & & \downarrow T c \\
S T & \xrightarrow{T c} & T T \\
\end{array}
\]

for any arrow $f$ of $C$. Horizontally, the composites by definition are $(\tau \circ \tau)_c = (\tau \circ \tau)_B$; the left-hand square commutes because $\tau$ is natural and $S$ is a functor, while the right-hand square commutes because $\tau$ is natural and $T f: T c \Rightarrow T b$ is an arrow. The commutativity of the outside of the diagram states that $\tau \circ \tau$ is natural.

This composition $\langle \tau, \tau \rangle \rightarrow \tau \circ \tau$ is readily shown to be associative. It moreover has identities. If $I_b: B \Rightarrow B$ is the identity functor for the category $B$ and $1_b: I_b \Rightarrow I_b$ the identity natural transformation of that functor to itself, one has $1_b = \tau \Rightarrow \tau$ and $1_b \Rightarrow \tau = \tau$. Thus $1_b$ is the identity for the composition $\Rightarrow$; it is also the identity for the composition $\cdot$. It is useful to let the symbol $S$ for a functor also denote the identity transformation $S \Rightarrow S$. With this notation in the situation above we have composite natural transformations

$$S \circ \tau: S \circ S \Rightarrow S \circ T,$$

\[
\begin{array}{ccc}
S & \xrightarrow{S c} & T \\
\downarrow T c & & \downarrow T c \\
S T & \xrightarrow{T c} & T T \\
\end{array}
\]

The definition (2) can then be rewritten, using also the vertical composition, as

$$\tau \circ \tau = (T \circ \tau) \cdot (\tau \circ T) = (T \circ \tau) \cdot (S \circ \tau).$$

There is a more general rule. Given three categories and four transformations

\[
\begin{array}{ccc}
C & \xrightarrow{S} & B \\
\downarrow I & & \downarrow I \\
D & \xrightarrow{S} & E \\
\end{array}
\]

$$\tau$$
the “vertical” composites under • and the “horizontal” composites under •
are related by the identity (interchange law)
\[(τ ancestral σ) • (τ • σ) = (τ ancestral σ) • (τ • σ).
\] (5)
The reader may enjoy writing down the evident diagrams needed to prove
this fact.
These results may be summarized as follows (considering only small
categories):

**Theorem 1.** The collection of all natural transformations is the set
of arrows of two different categories under two different operations of
composition, • and •, which satisfy the interchange law (5). Moreover, any
arrow (transformation) which is an identity for the composition • is also
an identity for the composition •. 

Note that the objects for the horizontal composition • are the categories,
for the vertical composition, the functors. In using these compositions,
the symbol • for the “horizontal” composition is often omitted (as it is
usually in writing composition of arrows in a category), while the solid
dot designating “vertical” composition is retained. Observe that objects
and arrows of C may be written as functors C : 1 → C or f : I → C; then
symbols such as σ • c = σ c have their accepted meaning in a situation
such as

\[1 → C \xleftarrow{c} B. \]

By a double category (Ehresmann) is meant a set which (like the set
of all natural transformations) is the set of arrows for two different compositions
which together satisfy (5). A 2-category (short for two-dimensional
category) is a double category in which every identity arrow for the first
composition is also an identity for the second composition. For example,
the category of all commutative squares in Set is a double category
(under the evident horizontal and vertical compositions) but not a
2-category. There are also n-categories for higher n.

Two (partially defined) binary operations • and • are said to satisfy the
interchange law when (5) holds wherever the compositions on either
side are defined. Here some other examples. If C is a category and
• : C × C → C is a functor (for example, a tensor product), while σ, σ', τ,
and τ' are arrows of C such that the composites σ • σ and τ • τ are defined,
then the interchange law (5) holds; indeed, it is precisely the requirement
that the functor • preserve composition •. If σ, σ', τ, and τ' are rectangular
matrices such that the usual matrix products σ • σ and τ • τ are defined,
while τ • σ denotes the matrix

\[
\begin{pmatrix}
τ & 0 \\
0 & σ
\end{pmatrix}
\]

with blocks τ and σ along the diagonal, zeroes elsewhere, then (5) holds.

The functor category \(B^C\) is itself a functor of the categories B and C,
covariant in B and contravariant in C. Specifically, if we consider only
the category \(\text{Cat}\) of all small categories, it is a functor \(\text{Cat}^\text{op} × \text{Cat} → \text{Cat}\;
the object function sends a pair of categories \((C, B)\) to the functor
category \(B^C\), and the arrow function sends a pair of functors \(F : B → B'\)
and \(G : C → C\) to the functor

\[F^G : B^C → B'^C\]
defined on objects \(S ∈ B^C\) as \(F^G S = F(S) ∘ G\) and on arrows \(τ : S → T\) in
\(B^C\) as \(F^G τ = F(τ) ∘ G\). Note, for example, that \(F^C\) is just “compose with
\(F\) on the left” while \(B^F\) is “compose with \(G\) on the right”. This functor
is an exact analogue to the hom-functor \(\text{Set}^\text{op} × \text{Set} → \text{Set}\).

**Exercises**

1. For small categories \(A, B,\) and \(C\) establish a bijection

\[\text{Cat}(A × B, C) ≅ \text{Cat}(A, C^B).\]

and show it natural in \(A, B,\) and \(C\). Hence show that \(- × B : \text{Cat} → \text{Cat}\) has
a right adjoint.

2. For categories \(A, B,\) and \(C\) establish natural isomorphisms

\[(A × B)^C ≅ A^C × B^C, \quad C^{A × B} ≅ (C^B)^A.\]

Compare the second isomorphism with the bijection of Exercise 1.

3. Use Theorem 1 to show that horizontal composition is a functor

\[• : A^B × B^C → A^C.\]

4. Let \(G\) be a topological group with identity element \(e\), while \(σ, σ', τ, τ'\),
are continuous paths in \(G\) starting and ending at \(e\) (thus, if \(I\) is the unit interval, \(σ : I → G\)
is continuous with \(σ(0) = e = σ(1)\)). Define \(τ • σ\) to be the path \(σ\) followed by
the path \(τ\), as in (1.5.1). Define \(τ • σ\) to be the pointwise product of \(τ\) and \(σ\), so that
\((σ • τ) = (τ • σ) = σ") for \(0 ≤ t ≤ 1\). Prove that the interchange law (5) holds.

5. (Hilton-Eckmann). Let \(S\) be a set with two (everywhere defined) binary operations
\(• : S × S → S, • : S × S → S\) which both have the same (two-sided) unit element \(e\)
and which satisfy the interchange identity (5). Prove that • and • are equal, and
that each is commutative.

6. Combine Exercises 4 and 5 to prove that the fundamental group of a topological
space is abelian.

7. If \(T : A → D\) is a functor, show that its arrow functions \(T_{ab} : A(a, b) → D(Ta, Tb)\)
define a natural transformation between functors \(A^{op} × A → Set\).

8. For the identity functor \(A\) of any category, the natural transformations
\(\alpha : A → A, \beta : A → A\) form a commutative monoid. Find this monoid in the cases \(C = \text{Grp, Ab, and Set}.\)
6. Comma Categories

There is another general construction of a category whose objects are certain arrows, as in the following several special cases.

If \( b \) is an object of the category \( C \), the category of objects under \( b \) is the category \( (b \downarrow C) \) with objects all pairs \( \langle f, c \rangle \), where \( c \) is an object of \( C \) and \( f: b \rightarrow c \) an arrow of \( C \), and with arrows \( h: \langle f, c \rangle \rightarrow \langle f', c' \rangle \) those arrows \( h: c \rightarrow c' \) of \( C \) for which \( f = f' \). Thus an object of \( (b \downarrow C) \) is just an arrow in \( C \) from \( b \) and an arrow of \( (b \downarrow C) \) is a commutative triangle with top vertex \( b \). In displayed form:

\[
\begin{array}{ccc}
  \text{objects} & \langle f, c \rangle: & \begin{array}{c}
  f \\
  \downarrow \\
  c
  \end{array} \\
  \text{arrows} & \langle f, c \rangle \rightarrow \langle f', c' \rangle: & \begin{array}{c}
  f \\
  \downarrow \\
  c
  \end{array} \rightarrow \begin{array}{c}
  f' \\
  \downarrow \\
  c'
  \end{array}
\end{array}
\]

(1)

The composition of arrows in \( (b \downarrow C) \) is then given by the composition in \( C \) of the base arrows \( h \) of these triangles.

For example, if \( h \) denotes any one-point set, while \( X \) is any set, each function \( \rightarrow X \) is just a selection of a point in the set \( X \); hence \( (h \downarrow \text{Set}) \) is just the category of pointed sets. Similarly, \( (Z \downarrow \text{Ab}) \) is the category of abelian groups, each with a selected element.

If \( a \) is an object of \( C \), the category \( (C \downarrow a) \) of objects over \( a \) has

\[
\begin{array}{ccc}
  \text{objects} & \begin{array}{c}
  c \\
  \downarrow \\
  a
  \end{array} \\
  \text{arrows} & \begin{array}{c}
  f \\
  \downarrow \\
  a
  \end{array} \rightarrow \begin{array}{c}
  f' \\
  \downarrow \\
  a
  \end{array}
\end{array}
\]

(2)

the triangle commutative. For example, \( h \) is terminal in \( \text{Set} \) so there is always a unique \( X \rightarrow h \); therefore \( (\text{Set} \downarrow h) \) isomorphic to \( \text{Set} \). Or again, \( \mathbb{Z} \) is a ring, and the category \( (\text{Ring} \downarrow \mathbb{Z}) \) is the category whose objects are rings equipped with a morphism \( e: R \rightarrow \mathbb{Z} \) (called a ring \( R \) with an "augmentation" \( e \)) and whose morphisms are morphisms of rings preserving the augmentation.

If \( b \) is an object of \( C \) and \( S: D \rightarrow C \) a functor, the category \( (b \downarrow S) \) of objects \( S \)-under \( b \) has as objects all pairs \( \langle f, d \rangle \) with \( d \in \text{Obj} D \) and \( f: b \rightarrow S \cdot d \) and as arrows \( h: \langle f, d \rangle \rightarrow \langle f', d' \rangle \) all those arrows \( h: d \rightarrow d' \) in \( D \) for which \( f' = S \cdot h \cdot f \). In pictures,

\[
\begin{array}{ccc}
  \text{objects} & \begin{array}{c}
  f \\
  \downarrow \\
  S \cdot d
  \end{array} \\
  \text{arrows} & \begin{array}{c}
  f \\
  \downarrow \\
  S \cdot d
  \end{array} \rightarrow \begin{array}{c}
  f' \\
  \downarrow \\
  S \cdot d'
  \end{array}
\end{array}
\]

(3)

Again, composition is given by composition of the arrows \( h \) in \( D \). Note especially that equality of arrows in \( (b \downarrow S) \) means their equality as arrows of \( D \).
Constructions on Categories

here \( d_0, d_1 \) are the two functors \( 1 \to 2 \), the functor category \( C^2 \) is just the category of arrows \( f \) of \( C \), and so the functors \( C^0, C^1 \) (defined as at the end of the last section) are simply the functors which send each arrow \( f \) of \( C \) to its domain and its codomain, respectively. The functors \( P \) and \( Q \) (called the projections of the comma category) and the functor \( R \) are defined (on objects) as suggested in the diagram

\[
\begin{align*}
e & \mapsto T e \land (f: T e \to S d) \mapsto S d \land d.
\end{align*}
\]

(6)

Exercises

1. If \( K \) is a commutative ring, show that the comma category \((K\downarrow \text{CRng})\) is the (usual) category of all small commutative \(K\)-algebras.
2. If \( t \) is a terminal object in \( C \), prove that \((C \downarrow t)\) is isomorphic to \( C \).
3. Complete (6) by defining \( P, Q \), and \( R \) on arrows.
4. (S. A. Hug.) Given functors \( T, S : D \to C \), show that a natural transformation \( \tau : T \to S \) is the same thing as a functor \( \tau : D \to (T \downarrow S) \) such that \( P \tau = Q \tau = \text{id}_D \), with \( P \) and \( Q \) the projections of (5).
5. Given any commutative diagram of categories and functors

\[
\begin{array}{c}
X \\
\downarrow & \downarrow \quad \downarrow \\
E \to C \cup C^2 \to C \to D
\end{array}
\]

(bottom row as in (5)), prove that there is a unique functor \( L : X \to (T \downarrow S) \) for which \( P = P L, Q = Q L \), and \( R = R L \). (This describes \((T \downarrow S)\) as a “pull-back”, c.f. §III.4).

6. (a) For fixed small \( C, D, \) and \( E \), show that \((T, S) \to (T \downarrow S)\) is the object function of a functor \((C_{\text{Graph}} \times (C^2)) \to \text{Cat}\).
(b) Describe a similar functor for variable \( C, D, \) and \( E \).

7. Graphs and Free Categories

Recall that a (directed) graph \( G \) (§I.2) is a set \( O \) of objects (vertices), and a set \( A \) of arrows \( f \) (edges), and a pair of functions \( A \rightharpoonup O \):

\[
A \rightharpoonup O, \quad \partial_0 f = \text{domain } f, \quad \partial_1 f = \text{codomain } f.
\]

A morphism \( D : G \to G' \) of graphs is a pair of functions \( D_0 : O \to O' \) and \( D_1 : A \to A' \) such that

\[
D_0 \partial_0 f = \partial_0 D_1 f \quad \text{and} \quad D_0 \partial_1 f = \partial_1 D_1 f.
\]

for every arrow \( f \in A \). These morphisms, with the evident composition, are the arrows of the category \( \text{Grph} \) of all small graphs (a graph is small if both \( O \) and \( A \) are small sets). Each graph may be pictured by a diagram of vertices (objects) and arrows, just like the diagram for a category except that neither composite arrows nor identity arrows are provided. Hence a graph is often called a diagram scheme or a precategory.

Every category \( C \) determines a graph \( UC \) with the same objects and arrows, forgetting which arrows are composites and which are identities.

Every functor \( F : C \to C \) is also a morphism \( UF : UC \to UC \) between the corresponding graphs. These observations define the forgetful functor \( U : \text{Cat} \to \text{Grph} \) from small categories to small graphs.

Let \( O \) be a fixed set. An \( O \)-graph will be one with \( O \) as its set of objects; a morphism \( D \) of \( O \)-graphs will be one with \( D_0 : O \to O \) the identity. The simplest \( O \)-graph on \( O \) is \( O \rightharpoonup O \), with both functions domain and range the identity. If \( A \) and \( B \) are (the sets of arrows of) two \( O \)-graphs, the product over \( O \) is

\[
A \times O B = \{ \langle g, f \rangle \mid \partial_0 g = \partial_1 f, g \in A, f \in B \};
\]

it is the set of “composable pairs” of arrows \( \cdot \to \cdot \to \cdot \). The definitions

\[
\partial_0 \langle g, f \rangle = \partial_0 f \quad \partial_1 \langle g, f \rangle = \partial_1 g
\]

make this set an \( O \)-graph. This product operation on \( O \)-graphs is associative, since for any three \( O \)-graphs \( A, B, C \) there is an evident isomorphism

\[
A \times_O (B \times_O C) \cong (A \times_O B) \times_O C.
\]

For the special \( O \)-graph \( O \) there is also an isomorphism \( A \cong A \times_O O \) given by \( f \mapsto \langle f, \partial_0 f \rangle \).

A category with objects \( O \) may be described as an \( O \)-graph equipped with two morphisms \( c : A \times_O A \rightharpoonup A \) and \( i : O \rightharpoonup A \) of \( O \)-graphs (composition and identity) such that the diagrams

\[
\begin{array}{c}
(A \times_O A) \times_O A \cong A \times_O (A \times_O A) \quad A \times_O A \cong O \times_O A \\
\downarrow c \quad \downarrow c
\end{array}
\]

are commutative, where \( 1 \times c \) is short for \( 1 \times_O c \), etc. Indeed, composable arrows \( \langle g, f \rangle \) have a composite given by \( c c \) as \( c(g, f) \), each object \( b \in O \) has an identity arrow given by \( i(b) \in A \), while the first diagram states that composition is associative and the second that each \( i(d) \) acts as a left and right identity for composition. In this sense, a category is like a monoid, as described in the introduction: A \( O \)-graph, and product of sets by \( \times_O \).

Any \( O \)-graph \( G \) may be used to “generate” a category \( C \) on the same set \( O \) of objects; the arrows of this category will be the “strings” of composable arrows of \( G \), so that an arrow of \( C \) from \( b \) to \( a \) may be pictured
as a path from $b$ to $a$, consisting of successive edges of $G$. This category $C$ will be written $C = C(G)$ and called the free category generated by the graph $G$. Its basic properties may be stated as follows.

Theorem 1. Let $G = (A \xrightarrow{f} O)$ be a small graph. There is a small category $C = C_{G}$ with $O$ as set of objects and a morphism $P : G \to UC$ of graphs from $G$ to the underlying graph $UC$ of $C$ with the following property. Given any category $B$ and any morphism $D : G \to UB$ of graphs, there is a unique functor $D' : C \to B$ with $(UD) \circ P = D$, as in the commutative diagram

\[
\begin{array}{ccc}
C & \xrightarrow{P} & UC \\
\downarrow & & \downarrow \text{id}_{UC} \\
B & \xrightarrow{D} & UB
\end{array}
\]

In particular, if $B$ has $O$ as set of objects and $D$ is a morphism of $O$-graphs, then $D'$ is the identity on objects.

The property of $P$ stated in (4) is equivalent to stating that the arrow $P : G \to UC$ is an initial object in the comma category $(G \downarrow U)$. Hence $P$ is unique up to an isomorphism (of $C$). Similar properties appear often; we shall say that $P$ is “universal” among morphisms from $G$ to the underlying-graph functor $U$.

Proof. Take the objects of $C$ to be those of $G$ and the arrows of $C$ to be the finite strings (or “paths”)

\[a_1 \xrightarrow{f_1} a_2 \xrightarrow{f_2} \cdots \xrightarrow{f_n} a_n\]

composed of $n$ objects $a_1, \ldots, a_n$ of $G$ connected by $n - 1$ arrows $f_1 : a_1 \to a_2$ of $G$. Regard each such string as an arrow $\langle a_1, f_1, \ldots, f_{n-1}, a_n \rangle : a_1 \to a_n$ in $C$, and define the composite of two strings by juxtaposition (i.e., by concatenation), identifying the common end. This composition is manifestly associative, and strings $\langle a_1 \rangle$ of length $n = 1$ are its identities. Every string of length $n > 1$ is a composite of strings of length 2:

\[\langle a_1, f_1, a_2, \ldots, a_{n-1}, f_{n-1}, a_n \rangle = \langle a_1, f_1, a_2, \ldots, f_{n-2}, a_{n-1}, f_{n-1}, a_n \rangle = \cdots = \langle a_1, f_1, a_2 \rangle .\]

The desired morphism $P : G \to UC$ of graphs sends each arrow $f : a_1 \to a_2$ of the given graph $G$ to the string $\langle a_1, f, a_2 \rangle$ of length 2.

Now consider any other morphism $D : G \to UB$ of the given graph $G$ to the underlying graph of some category $B$. If there is a functor $D' : C \to B$ with $UD' \circ P = D$, as in the commutative diagram (4), then $D'$ must be $D' \langle a \rangle = Da$ on objects and $D' \langle a_1, f_1, a_2 \rangle = Df_1$. Since any string of length $n > 1$ is a composite (5) in $C$, $D'$ must be given by

\[D' \langle a_1, f_1, a_2, \ldots, a_{n-1}, f_{n-1}, a_n \rangle = Df_{n-1} \circ \cdots \circ Df_1 .\]

Conversely, this formula does define a functor $D' : B \to C$ for which the indicated diagram commutes, q.e.d.

Quotient Categories

Here are some easy examples. For the graph consisting of a single arrow $f$ with $\partial_0 f = \partial_1 f$, the free category consists of all arrows $1, f, f^2, \ldots$. For the graph consisting of a single arrow $g$ with different ends, the free category consists of this arrow plus two identity arrows (one at each end). For the graph $\xrightarrow{g} \xrightarrow{h}$ with three different vertices in the free graph is a commutative triangle (add one composite arrow and three identity arrows).

When $O$ consists of one point, the graph $G$ reduces simply to a set $X$ (the set $X = A$ of arrows) and the theorem provides the familiar construction of a free monoid $M$ generated by $X$, as follows:

Corollary 2. To any set $X$ there is a monoid $M$ and a function $p : X \to UM$, where $UM$ is the underlying set of $M$, with the following universal property: For any monoid $L$ and any function $h : X \to UL$ there is a unique morphism $h' : M \to L$ of monoids with $h = U h' \circ p$.

The elements of $M$ are the identity and strings $\langle x_1, \ldots, x_{n-1} \rangle$, for $x_i \in X$.

Graphs may be used to describe diagrams. If $G$ is any graph, a diagram of the shape $G$ in the category $B$ may be defined to be a morphism $D : G \to UB$ of graphs. By the Theorem, these morphisms $D$ correspond exactly to functors $D' : C_{G} \to B$, via the bijection $D' \mapsto D = UD' \circ P$. This bijection

\[\text{Cat}(C_{G}, B) \cong \text{Grph}(G, UB)\]

is natural in $G$ and $B$, so asserts that $C : \text{Grph} \to \text{Cat}$ is left adjoint to the forgetful functor $U : \text{Cat} \to \text{Grph}$.

Exercises

1. Define “opposite graph” and “product of two graphs” to agree with the corresponding definitions for categories (i.e., so that the functor $U$ will preserve opposites and products).
2. Show that every finite ordinal number is a free category.
3. Show that each graph $G$ generates a free groupoid $F$ (i.e., one which satisfies Theorem 1 with “category $C$” replaced by “groupoid $F$” and “category $B$” by “groupoid $E$”). Deduce as a corollary that every set $X$ generates a free group.

8. Quotient Categories

Certain categories may be described by generators and relations, as follows:

Proposition 1. For a given category $C$, let $R$ be a function which assigns to each pair of objects $a, b$ of a binary relation $R_{a,b}$ on the hom-set $C(a, b)$. Then there exists a category $C/R$ and a functor $Q : Q : C \to C/R$ such that (i) If $f \in R_{a,b}$, then $Qf = Qf'$ in $C/R$; (ii) If $H : C \to D$ is an
functor from $C$ for which $fR_{a,b}f'$ implies $Hf = Hf'$ for all $f$ and $f'$, then there is a unique functor $H^*: C/R 	o D$ with $H^*Q_R = H$. Moreover, the functor $Q_R$ is a bijection on objects.

Put briefly: $Q$ is the universal functor on $C$ with $Qf = Qf'$ whenever $fRf'$.

For example, if $C=\text{Top}$ and $fRf'$ means that $f$ is homotopic to $f'$, then the desired quotient category $C/R$ is just the category $\text{Top}_{\mathbb{R}}$ of §1.7, with objects topological spaces and arrows homotopy classes of continuous maps. This direct construction is possible for $\text{Top}_{\mathbb{R}}$ because the relation of homotopy between maps is an equivalence relation preserved by composition. The general case requires a preliminary construction on the relation $R$ to achieve these properties.

Sketch of proof. Call $R$ a congruence on $C$ if (i) For each pair $a, b$ of objects, $R_{a,b}$ is a reflexive, symmetric, and transitive relation on $C(a,b)$; (ii) If $f, f': a \to b$ have $fR_{a,b}f'$, then for all $g: a \to a$ and all $h: b \to b'$ one has $(hfg)R_{a,b}(h'f'g)$. Given any $R$, there is a least congruence $R'$ on $C$ with $R \subseteq R'$ (proof as exercise). Now take the objects of $C/R$ to be the objects of $C$, and take each hom-set $(C/R)(a,b)$ to be the quotient $C(a,b)/R_{a,b}$ of $C(a,b)$ by the equivalence relation $R$ there. Since the relation is preserved by composition, the composite in $C$ carries over to $C/R$ by the evident projection $Q: C \to C/R$. Now for any functor $H: C \to D$ the sets $S_{a,b} = \{f, f': a \to b \mid Hf = Hf'\}$ evidently form a congruence on $C$. Thus, if $S \supseteq R$, one also has $S \supseteq R'$, and $H$ factors as $H = H^*Q_R$, as required.

In case $C$ is the free category generated by a graph $G$ we call $C/R$ the category with generators $G$ and relations $R$. For example, $3$ may be described as the category generated by three objects $0, 1, 2$, three arrows $f: 0 \to 1$, $g: 1 \to 2$, and $h: 0 \to 2$, and one relation $h = g f$. As a special case, (one object) we may speak of a monoid given by generators and relations.

Exercises

1. Show that the category generated by the graph

![Graph](attachment:image.png)

with the one relation $g'f = f'g$ has four identity arrows and exactly five non-identity arrows $f, g, f', g'$ and $g'f = f'g$.

2. If $C$ is a group $G$ (regarded as a category with one object) show that to each congruence $R$ on $C$ there is a normal subgroup $N$ of $G$ with $fRg$ if and only if $g^{-1}f \in N$.

Notes.

The leading idea of this chapter is to make the simple notion of a functor apply to complex cases by defining suitable complex categories – the opposite category for contravariant functors, the product category for bifunctors, the functor category really as an adjoint to the product, and the comma category to reduce universal arrows to initial objects. The importance of the use of functor categories (sometimes called “categories of diagrams”) was emphasized by Grothendieck [1957] and Freyd [1964]. The notion of a comma category, often used in special cases, was introduced in full generality in Lawvere’s (unpublished) thesis [1963], in order to give a set-free description of adjoint functors. For a time it was a sort of secret tool in the arsenal of knowledgeable experts.

Duality has a long history. The duality between point and line in geometry, especially projective geometry, led to a sharp description of axiomatic duality in the monumental treatise by Veblen-Young on projective geometry. The explicit description of duality by opposite categories is often preferable, as in the Pontrjagin duality which appears (§IV.3) as an equivalence between categories, or as an equivalence between a category and an opposite category (see Negrepontis [1971]).
III. Universals and Limits

Universal constructions appear throughout mathematics in various guises—as universal arrows to a given functor, as universal arrows from a given functor, or as universal elements of a set-valued functor. Each universal determines a representation of a corresponding set-valued functor as a hom-functor. Such representations, in turn, are analyzed by the Yoneda Lemma. Limits are an important example of universals—both the inverse limits (= projective limits = limits = left roots) and their duals, the direct limits (= inductive limits = colimits = right roots). In this chapter we define universals and limits and examine a few basic types of limits (products, pullbacks, and equalizers ...). Deeper properties will appear in Chapter IX on special limits, while the relation to adjoints will be treated in Chapter V.

1. Universal Arrows

Given the forgetful functor \( U : \text{Cat} \rightarrow \text{Grph} \) and a graph \( G \), we have constructed (§ II.7) the free category \( C \) on \( G \) and the morphism \( P : G \rightarrow UC \) of graphs which embeds \( G \) in \( C \), and we have shown that this arrow \( P \) is “universal” from \( G \) to \( U \). A similar universality property holds for the morphisms embedding generators into free algebraic systems of other types. Here is the general concept.

**Definition.** If \( S : D \rightarrow C \) is a functor and \( c \) an object of \( C \), a universal arrow from \( c \) to \( S \) is a pair \( < r, u > \) consisting of an object \( r \) of \( D \) and an arrow \( u : c \rightarrow Sr \) of \( C \), such that to every pair \( < d, f > \) with \( d \) an object of \( D \) and \( f : c \rightarrow Sd \) an arrow of \( C \), there is a unique arrow \( f^\prime : d \rightarrow D \) with \( Sf^\prime u = f \). In other words, every arrow \( f \) to \( S \) factors uniquely through the universal arrow \( u \), as in the commutative diagram

\[
\begin{array}{ccc}
  c & \rightarrow & Sr \\
  \downarrow^u & & \downarrow^f \\
  c & \rightarrow & Sd
\end{array}
\]

Equivalently, \( u : c \rightarrow Sr \) is universal from \( c \) to \( S \) when the pair \( < r, u > \) is an initial object in the comma category \( (c \downarrow S) \), whose objects are the arrows \( c \rightarrow Sd \). As with any initial object, it follows that \( < r, u > \) is unique up to isomorphism in \( (c \downarrow S) \); in particular, the object \( r \) of \( D \) is unique up to isomorphism in \( D \). This remark is typical of the use of comma categories.

This notion of a universal arrow has a great variety of examples: we list a few:

**Bases of Vector Spaces.** Let \( \text{Vect} \) denote the category of all vector spaces over a fixed field \( K \), with arrows linear transformations, while \( U : \text{Vect} \rightarrow \text{Set} \) is the forgetful functor, sending each vector space \( V \) to the set of its elements. For any set \( X \) there is a familiar vector space \( V_X \) with \( X \) as a set of basis vectors; it consists of all formal \( K \)-linear combinations of the elements of \( X \). The function which sends each \( x \in X \) into the same \( x \) regarded as a vector of \( V_X \) is an arrow \( j : X \rightarrow U(V_X) \). For any other vector space \( W \), it is a fact that each function \( f : X \rightarrow U(W) \) can be extended to a unique linear transformation \( f^\prime : V_X \rightarrow W \) with \( Uf^\prime = f \). This familiar fact states exactly that \( j \) is a universal arrow from \( X \) to \( U \).

**Free Categories from Graphs.** Theorem II.7.1 states exactly that the functor \( P : G \rightarrow UC \) is universal. The same observation applies to the free monoid on a given set of generators, the free group on a given set of generators, the free \( R \)-module (over a given ring \( R \)) on a given set of generators, the polynomial algebra over a given commutative ring in a given set of generators, and so on in many cases of free algebraic systems.

**Fields of Quotients.** To any integral domain \( D \) a familiar construction gives a field \( Q(D) \) of quotients of \( D \) together with a monomorphism \( j : D \rightarrow Q(D) \) (which is often formulated by making \( D \) a subdomain of \( Q(D) \)). This field of quotients is usually described as the smallest field containing \( D \), in the sense that for each \( D \subset K \) with \( K \) a field there is a monomorphism \( f : Q(D) \rightarrow K \) of fields which is the identity on the common subdomain \( D \). However, this inclusion \( D \subset K \) may readily be replaced by any monomorphism \( D \subset K \) of domains. Hence our statement means that the pair \( < Q(D),j > \) is universal for the forgetful functor \( \text{Fld} \rightarrow \text{Dom} \) from the category of fields to that of domains—provided we take arrows of \( \text{Dom} \) to be the monomorphisms of integral domains (note that a homomorphism of fields is necessarily a monomorphism). However, for the larger category \( \text{Dom} \) with arrows all homomorphisms of integral domains there does not exist a universal arrow from each domain. For instance, for the domain \( Z \) of integers there is for each prime \( p \) a homomorphism \( Z \rightarrow Z_p \); the reader should observe that this makes impossible the construction of a universal arrow from \( Z \) to the functor \( \text{Fld} \rightarrow \text{Dom} \).

**Complete Metric Spaces.** Let \( \text{Met} \) be the category of all metric spaces \( X, Y, ... \) with arrows \( X \rightarrow Y \) those functions which preserve the metric
Universal Arrows

(And which therefore are necessarily injections). The complete metric spaces form (the objects of) a full subcategory. The familiar completion \( \hat{X} \) of a metric space \( X \) provides an arrow \( X \to \hat{X} \) which is universal for the evident forgetful functor (from complete metric spaces to metric spaces).

In many other cases, the function embedding a mathematical object in a suitably completed object can be interpreted as a universal arrow. The general fact of the uniqueness of the universal arrow implies the uniqueness of the completed object, up to a unique isomorphism (who wants more?)

The idea of universality is sometimes expressed in terms of "universal elements". If \( D \) is a category and \( H : D \to \text{Set} \) a functor, a universal element of the functor \( H \) is a pair \( \langle r, e \rangle \) consisting of an object \( r \in D \) and an element \( e \in Hr \) such that for every pair \( \langle d, x \rangle \) with \( x \in Hd \) there is a unique arrow \( f : r \to d \) of \( D \) with \( (Hf)e = x \).

Many familiar constructions are naturally examples of universal elements. For instance, consider an equivalence relation \( E \) on a set \( S \), the corresponding quotient set \( S/E \) consisting of the equivalence classes of elements of \( S \) under \( E \), and the projection \( p : S \to S/E \) which sends each \( s \in S \) to its \( E \)-equivalence class. Now \( S/E \) has the familiar property that any function \( f \) on \( S \) which respects the equivalence relation can be regarded as a function on \( S/E \). More formally, this means that if \( f : S \to X \) has \( fs = fs' \) whenever \( sE = s' \), then \( f \) can be written as a composite \( f = fp \) for a unique function \( f' : S/E \to X \):

\[
\begin{array}{ccc}
S & \xrightarrow{p} & S/E \\
\| & & \downarrow f \\
S & \xrightarrow{f} & X.
\end{array}
\]

This states exactly that \( \langle S/E, p \rangle \) is a universal element for that functor \( H : \text{Set} \to \text{Set} \) which assigns to each set \( X \) the set \( HX \) of all those functions \( f : S \to X \) for which \( sE = s' \) implies \( fs = fs' \).

Again, let \( N \) be a normal subgroup of a group \( G \). The usual projection \( p : G \to G/N \) which sends each \( g \in G \) to its coset \( pg = gN \) in the quotient group \( G/N \) is a universal element for that functor \( H : \text{Grp} \to \text{Set} \) which assigns to each group \( G \) the set \( HG \) of all those homomorphisms \( f : G \to G' \) which kill \( N \) (have \( fN = 1 \)). Indeed, every such homomorphism factors as \( f = f'p \), for a unique \( f' : G/N \to G' \). Now the quotient group is usually described as a group whose elements are cosets. However, once the cosets are used to prove this one "universal" property of \( p : G \to G/N \), all other properties of quotient groups – for example, the isomorphism theorems – can be proved with no further mention of cosets (see Mac Lane-Birkhoff [1967]). All that is needed is the existence of a universal element

\( p \) of the functor \( H \). For that matter, even this existence could be proved without using cosets (see §V.6).

Tensor products provide another example of universal elements. Given two vector spaces \( V \) and \( V' \) over the field \( K \), the function \( H \) which assigns to each vector space \( W \) the set \( HW = \text{Bilin}(V, V'; W) \) of all bilinear functions \( V \times V' \to W \) is the object function of a functor \( H : \text{Vect} \to \text{Set} \), and the usual construction of the tensor product provides both a vector space \( V \otimes V' \) and a bilinear function \( \otimes : V \times V' \to V \otimes V' \), usually written \( \langle v, v' \rangle \mapsto v \otimes v' \), so that the pair \( \langle V \otimes V', \otimes \rangle \) is a universal element for the functor \( H = \text{Bilin}(V, V'; -) \). This applies equally well when the field \( K \) is replaced by a commutative ring (and vector spaces by \( K \)-modules).

The notion "universal element" is a special case of the notion "universal arrow". Indeed, if \( * \) is the set with one point, then any element \( e \in Hr \) can be regarded as an arrow \( e : * \to Hr \) in \( \text{Ens} \). Thus a universal element \( \langle r, e \rangle \) for \( H \) is exactly a universal arrow from \( * \) to \( H \). Conversely, if \( C \) has small hom-sets, the notion "universal arrow" is a special case of the notion "universal element". Indeed, if \( S : D \to C \) is a functor and \( c \in C \) is an object, then \( \langle r, uc : c \to Sr \rangle \) is a universal arrow from \( r \) to \( c \) if and only if the pair \( \langle r, uc \in C(c, Sr) \rangle \) is a universal element of the functor \( H = C(c, -) \). This is the functor which acts on objects \( d \) and arrows \( h \) of \( D \) by

\[
d \mapsto C(c, Sd), \quad h \mapsto C(c, Sh).
\]

Hitherto we have treated universal arrows from an object \( c \in C \) to a functor \( S : D \to C \). The dual concept is also useful. A universal arrow from \( S \) to \( c \) is a pair \( \langle r, v \rangle \) consisting of an object \( r \in D \) and an arrow \( v : Sr \to c \) with codomain \( c \) such that to every pair \( \langle d, f \rangle \) with \( f : Sd \to c \) there is a unique \( f' : d \to r \) with \( f = v \circ f' \), as in the commutative diagram

\[
\begin{array}{ccc}
d & \xrightarrow{Sd} & c \\
\| & & \downarrow v \\
r & \xleftarrow{sr} & c.
\end{array}
\]

The projections \( p : a \times b \to a, g : a \times b \to b \) of a product in \( C \) (for \( C = \text{Grp}, \text{Set}, \text{Cat}, \ldots \)) are examples of such a universal. Indeed, given any other pair of arrows \( f : c \to a, g : c \to b \) to \( a \) and \( b \), there is a unique \( h : c \to a \times b \) with \( ph = f, qh = g \). Therefore \( \langle p, q \rangle \) is a "universal pair".

To make it a universal arrow, introduce the diagonal functor \( \Delta : C \to C \times C \), with \( \Delta c = \langle c, c \rangle \). Then the pair \( f, g \) above becomes an arrow \( \langle f, g \rangle : \Delta c \to \langle a, b \rangle \) in \( C \times C \), and \( \langle p, q \rangle \) is a universal arrow from \( \Delta \) to the object \( \langle a, b \rangle \).
Similarly, the kernel of a homomorphism (in Ab, Grp, Rng, R-Mod, ...) is a universal, more exactly, a universal for a suitable contravariant functor. Note that we say "universal arrow to S" and "universal arrow from S" rather than "universal" and "couniversal".

Exercises

1. Show how each of the following familiar constructions can be interpreted as a universal arrow:
   (a) The integral group ring of a group (better, of a monoid).
   (b) The tensor algebra of a vector space.
   (c) The exterior algebra of a vector space.
2. Find a universal element for the contravariant power set functor \( P : \text{Set}^\text{op} \to \text{Set} \).
3. Find (from any object) universal arrows to the following forgetful functors:
   \( \text{Ab} \to \text{Grp}, \text{Rng} \to \text{Ab} \) (forget the multiplication).
4. Use only universality (of projections) to prove the following isomorphisms of group theory:
   (a) For normal subgroups \( M, N \) of \( G \) with \( M \subseteq N \), \( (G/M)/(N/M) \cong G/N \).
   (b) For subgroups \( S \) and \( N \) of \( G \), \( G \) normal, with join \( SN \), \( SN/N \cong S/S \cap N \).
5. Show that the quotient \( K \)-module \( A/S \) (where \( S \) a submodule of \( A \)) has a description by universality. Derive isomorphism theorems.
6. Describe quotients of a ring by a two-sided ideal by universality.
7. Show that the construction of the polynomial ring \( K[x] \) in an indeterminate \( x \) over a commutative ring \( K \) is a universal construction.

2. The Yoneda Lemma

Next we consider some conceptual properties of universality. First, universality can be formulated with hom-sets, as follows:

**Proposition 1.** For a functor \( S : D \to C \) a pair \( \langle r, u : c \to Br \rangle \) is universal from \( c \) to \( S \) if and only if the function sending each \( f' : r \to d \) into \( Sf' \circ u : c \to Sd \) is a bijection of hom-sets

\[
D(r, d) \cong C(c, Sd). \tag{1}
\]

This bijection is natural in \( d \). Conversely, given \( r \) and \( c \), any natural isomorphism (1) is determined in this way by a unique arrow \( u : c \to Br \) such that \( \langle r, u \rangle \) is universal from \( c \) to \( S \).

**Proof.** The statement that \( \langle r, u \rangle \) is universal is exactly the statement that \( f' \mapsto Sf' \circ u = f \) is a bijection. This bijection is natural in \( d \), for if \( g' : d \to d' \), then \( S(g' \circ f') \circ u = Sg' \circ (Sf' \circ u) \).

Conversely, a natural isomorphism (1) gives for each object \( d \) of \( D \) a bijection \( \varphi_d : D(r, d) \to C(c, Sd) \). In particular, choose the object \( d \) to be \( r \);

\[
\text{the identity } 1 \in D(r, r) \text{ then goes by } \varphi_r \text{ to an arrow } u : c \to Br \text{ in } C. \]

For any \( f' : r \to d \) the diagram

\[
\begin{array}{ccc}
D(r, r) & \xrightarrow{\varphi_r} & C(c, Sd) \\
\downarrow{D(r, f')} & & \downarrow{C(c, Sf')} \\
D(r, d) & \xrightarrow{\varphi_d} & C(c, Sd)
\end{array}
\]

commutes because \( \varphi \) is natural. But in this diagram, \( 1 \in D(r, r) \) is mapped (top and right) to \( Sf' \circ u \) and (left and bottom) to \( \varphi_d(f') \). Since \( \varphi_d \) is a bijection, this states precisely that each \( f : c \to Sd \) has the form \( f = Sf' \circ u \) for a unique \( f' \). This is precisely the statement that \( \langle r, u \rangle \) is universal.

If \( C \) and \( D \) have small hom-sets, this result (1) states that the functor \( C(c, S) \to \text{Set} \) is naturally isomorphic to a covariant hom-functor \( D(r, -) \). Such isomorphisms are called representations:

**Definition.** Let \( D \) have small hom-sets. A representation of a functor \( K : D \to \text{Set} \) is a pair \( \langle r, \psi \rangle \), with \( r \) an object of \( D \) and

\[
\psi : D(r, -) \cong K \tag{3}
\]

a natural isomorphism. The object \( r \) is called the representing object. The functor \( K \) is said to be representable when such a representation exists.

Up to isomorphism, a representable functor is just thus a covariant hom-functor \( D(r, -) \). This notion can be related to universal arrows as follows.

**Proposition 2.** Let \( * \) denote any one-point set and let \( D \) have small hom-sets. If \( \langle r, u : * \to Kr \rangle \) is a universal arrow from \( * \) to \( K : D \to \text{Set} \), then the function \( \psi \) which for each object \( d \) of \( D \) sends the arrow \( f' : r \to d \) to \( K(f')(u) \in Kd \) is a representation of \( K \). Every representation of \( K \) is obtained in this way from exactly one such universal arrow.

**Proof.** For any set \( X \), a function \( f : * \to X \) from the one-point set \( * \) to \( X \) is determined by the element \( f(*) \in X \). This correspondence \( f \mapsto f(*) \) is a bijection \( \text{Set}(*, X) \to X \), natural in \( X \in \text{Set} \). Composing with \( K \) yields a natural isomorphism \( \text{Set}(*, K-) \to K \). This plus the representation \( \psi \) of (3) gives

\[
\text{Set}(*, K-) \cong K \cong D(r, -). \tag{4}
\]

Therefore a representation of \( K \) amounts to a natural isomorphism \( \text{Set}(*, K-) \cong D(r, -) \). The proposition thus follows from the previous one.

A direct proof is equally easy: Given the universal arrow \( u \), the correspondence \( f' \mapsto K(f')(u \circ *) \) is a representation; given a representation \( \psi \) as in (3), \( \psi \), maps \( 1 : r \to r \) to an element of \( K r \), which is a universal element, hence also a universal arrow \( * \to Kr \).
The Yoneda Lemma

Observe that each of the notions “universal arrow”, “universal element”, and “representable functor” subsumes the other two. Thus, a universal arrow from \( c \) to \( S : D \rightarrow C \) amounts (Proposition 1) to a natural isomorphism \( D(r, c) \rightarrow C(c, Sd) \) and hence to a representation of the functor \( C(c, S -) : D \rightarrow \text{Set} \) or equally well to a universal element for the same functor.

The argument for Proposition 1 rested on the observation that each natural transformation \( \phi : D(r, -) \rightarrow K \) is completely determined by the image under \( \phi \) of the identity \( 1 : r \rightarrow r \). This fact may be stated as follows:

**Lemma.** (Yoneda). If \( K : D \rightarrow \text{Set} \) is a functor from \( D \) and \( r \) an object in \( D \) (for \( D \) a category with small hom-sets), there is a bijection

\[
y : \text{Nat}(D(r, -), K) \cong K_r
\]

which sends each natural transformation \( \alpha : D(r, -) \rightarrow K \) to \( \alpha_1 \), the image of the identity \( r \rightarrow r \).

The proof is indicated by the following commutative diagram:

\[
\begin{array}{ccc}
D(r, r) & \xrightarrow{\alpha} & K(r) \\
f_* = D(r, f) & \downarrow & k(f) \\
D(r, d) & \xrightarrow{\alpha_d} & K(d)
\end{array}
\]

Corollary. For objects \( r, s \in D \), each natural transformation \( D(r, -) \rightarrow D(s, -) \) has the form \( D(h, -) \) for a unique arrow \( h : r \rightarrow s \).

The Yoneda map \( y \) of (4) is natural in \( K \) and \( r \). To state this fact formally, we must consider \( K \) as an object in the functor category \( \text{Set}^D \), regard both domain and codomain of the map \( y \) as objects of the pair \( \langle K, r \rangle \), and consider this pair as an object in the category \( \text{Set}^D \times D \). The codomain for \( y \) is then the evaluation functor \( E \), which maps each pair \( \langle K, r \rangle \) to the value \( Kr \) of the functor \( K \) at the object \( r \); the domain is the functor \( N \) which maps the object \( \langle K, r \rangle \) to the set \( \text{Nat}(D(r, -), K) \) of all natural transformations and which maps a pair of arrows \( F : K \rightarrow K', f : r \rightarrow r' \rightarrow \text{Nat}(D(f, -), F) \). With these observations we may at once prove an addendum to the Yoneda Lemma:

**Lemma.** The bijection of (4) is a natural isomorphism \( y : N \rightarrow E \) between the functors \( E, N : \text{Set}^D \times D \rightarrow \text{Set} \).

The object function \( r \mapsto D(r, -) \) and the arrow function

\[
(f : s \rightarrow r) \mapsto D(f, -) : D(r, -) \rightarrow D(s, -)
\]

for \( f \) an arrow of \( D \) together define a faithful functor

\[
Y : D^\text{op} \rightarrow \text{Set}^D
\]

called the Yoneda functor. Its dual is another such functor

\[
Y' : D \rightarrow \text{Set}^D
\]

(also faithful) which sends \( f : s \rightarrow r \) to the natural transformation

\[
D(-, f) : D(-, s) \rightarrow D(-, r) : D^\text{op} \rightarrow \text{Set}.
\]

\( D \) must have small hom-sets if these functors are to be defined (because \( \text{Set} \) is the category of all small sets). For larger \( D \), the Yoneda lemmas remain valid if \( \text{Set} \) is replaced by any category \( \text{Ens} \) whose objects are sets \( X, Y, \ldots \), and for which \( \text{Ens}(X, Y) \) is the set of all functions from \( X \) to \( Y \), provided of course that \( D \) has hom-sets which are objects in \( \text{Ens} \). (The meaning of naturality is not altered by further enlargement of \( \text{Ens} \); see Exercise 4.)

**Exercises**

1. Let functors \( K, K' : D \rightarrow \text{Set} \) have representations \( \langle r, \psi \rangle \) and \( \langle r', \psi' \rangle \), respectively. Prove that to each natural transformation \( \tau : K \rightarrow K' \), there is a unique morphism \( h : r' \rightarrow r \) of \( D \) such that

\[
\tau \circ \psi = \psi' : D(h, -) : D(r, -) \rightarrow K'.
\]

2. State the dual of the Yoneda Lemma (\( D \) replaced by \( D^\text{op} \)).

3. (Kan; the coyoneda lemma) For \( K : D \rightarrow \text{Set}, \ast \in K \) is the category of elements \( x \in Kd, \ast : \ast \mapsto D \) is the projection \( x \in Kd \rightarrow d \) and for each \( a \in D, a : \ast \mapsto D \) is the diagonal functor sending everything to the constant value \( a \). Establish a natural isomorphism

\[
\text{Nat}(K, D(a, -)) \cong \text{Nat}(a, D).
\]

4. (Naturality is not changed by enlarging the codomain category.) Let \( E \) be a full subcategory of \( E' \). For functors \( K, L : D \rightarrow E, \text{with } J : E \rightarrow E' \) the inclusion, prove that \( \text{Nat}(K, L) \cong \text{Nat}(J K, J L) \).

3. Coproducts and Colimits

We introduce colimits by a variety of special cases, each of which is a universal.

**Coproducts.** For any category \( C \), the diagonal functor \( \Delta : C \rightarrow C \times C \) is defined on objects by \( \Delta(c) = \langle c, c \rangle \), on arrows by \( \Delta(f) = \langle f, f \rangle \). A universal arrow from an object \( \langle a, b \rangle \) of \( C \times C \) to the functor \( \Delta \) is called a coproduct diagram. It consists of an object \( c \) of \( C \) and an arrow \( \langle a, b \rangle \rightarrow \langle c, c \rangle \) of \( C \times C \); that is, a pair of arrows \( a : c \rightarrow a, b : c \rightarrow b \) to a common codomain \( c \). This pair has the familiar universal property: For any pair of arrows \( f : a \rightarrow d, g : b \rightarrow d \) there is a unique \( f' : c \rightarrow d \) with \( f = f' \circ i, g = f' \circ j \). When such a coproduct diagram exists,
the object \( c \) is necessarily unique (up to isomorphism in \( C \)); it is written \( c = a \sqcup b \) or \( c = a + b \) and is called a coproduct object. The coproduct diagram then is

\[
\begin{array}{c}
a \\
\downarrow a \sqcup b \\
\downarrow b
\end{array}
\]

the arrows \( i \) and \( j \) are called the injections of the coproduct \( a \sqcup b \) (though they are not required to be injective as functions). The universality of this diagram states that any diagram of the following form can be filled in uniquely (at \( h \)) so as to be commutative:

\[
\begin{array}{c}
a \\
\downarrow a \sqcup b \\
\downarrow b
\end{array}
\]

Hence the assignment \( \langle f, g \rangle \mapsto h \) is a bijection

\[
C(a, d) \times C(b, d) \cong C(a \sqcup b, d)
\]

natural in \( d \), with inverse \( h \mapsto \langle h_1, h_2 \rangle \). If every pair of objects \( a, b \) in \( C \) has a coproduct then, choosing a coproduct diagram for each pair, the coproduct \( a \sqcup b \) maps \( a \otimes k \) for arrows \( a \rightarrow a' \), \( b \rightarrow b' \) as the unique arrow \( a \otimes k : a \sqcup b \rightarrow a' \sqcup b' \) with \( (a \otimes k) i = i', (a \otimes k) j = j' \) (draw the diagram!).

The diagram (1) is more familiar in other guises. For example, in Set take \( a \sqcup b \) to be a disjoint union of the sets \( a \) and \( b \) (i.e., a union of disjoint copies of \( a \) and \( b \)), while \( i \) and \( j \) are the inclusion maps \( a \subseteq a \sqcup b \), \( b \subseteq a \sqcup b \). Now a function \( h \) on a disjoint union is uniquely determined by independently giving its values on \( a \) and on \( b \); i.e., by giving the composites \( h i \) and \( h j \). This says exactly that diagram (1) can be filled in uniquely at \( h \). To be sure, a disjoint union is not unique, but it is unique up to a bijection, as befits a universal.

The coproduct of any two objects exists in many of the familiar categories, where it has a variety of names as indicated in the following list:

- **Set**: disjoint union of sets,
- **Top**: disjoint union of spaces,
- **Top**: wedge product (join two spaces at base point),
- **Ab, R-Mod**: direct sum \( A \oplus B \),
- **Grp**: free product,
- **CRing**: tensor product \( R \otimes S \).

In a preorder \( P \), a least upper bound \( a \lor b \) of two elements \( a \) and \( b \), if it exists, is an element \( a \lor b \) with the properties (i) \( a \leq a \lor b, b \leq a \lor b \); and (ii) if \( a \leq c \) and \( b \leq c \), then \( a \lor b \leq c \). These properties state exactly that \( a \lor b \) is a coproduct of \( a \) and \( b \) in \( P \), regarded as a category.

### Infinite Coproducts

In the description of the coproduct, replace \( C \times C = C^2 \) by \( C^X \) for any set \( X \). Here the set \( X \) is regarded as a discrete category, so the functor category \( C^X \) has as its objects the \( X \)-indexed families \( a = \{a_x \mid x \in X \} \) of objects of \( C \). The corresponding diagonal functor \( d : C \rightarrow C^X \) sends each \( c \) to the constant family (all \( c_x = c \)). A universal arrow from \( a \) to \( d \) is an \( X \)-fold coproduct diagram; it consists of a coproduct object \( \sqcup_x a_x \) in \( C \) and arrows (coproduct injections) \( i_x : a_x \rightarrow \sqcup_x a_x \) of \( C \) with the requisite universal property. This universal property states that the assignment \( f \mapsto \{f_x \mid x \in X \} \) is a bijection

\[
\text{C}(\sqcup_x a_x, c) \cong \prod_{x \in X} \text{C}(a_x, c),
\]

natural in \( c \). In Set, a coproduct is an \( X \)-fold disjoint union.

**Coproducts.** If the factors in a coproduct are all equal \( a_x = b \) for all \( x \), the coproduct \( \sqcup_x a_x \) is called a copower and is written \( X \cdot b \), so that

\[
C(X \cdot b, c) \cong C(b, c)^X,
\]

natural in \( c \). For example, in Set, with \( b = Y \) a set, the copower \( X \cdot Y = X \times Y \) is the cartesian product of the sets \( X \) and \( Y \).

**Cokernels.** Suppose that \( C \) has a zero object \( 0 \), so that for any two objects \( b, c \) in \( C \) there is a zero arrow \( 0 : b \rightarrow c \). The cokernel of \( f : a \rightarrow b \) is then an arrow \( u : b \rightarrow c \) such that (i) \( uf = 0 : a \rightarrow c \); (ii) if \( h : b \rightarrow c \) has \( hf = 0 \), then \( h = hu \) for a unique arrow \( h' : e \rightarrow c \). The picture is

\[
\begin{array}{c}
a \\
\downarrow f \\
b \\
\downarrow u \\
\downarrow 0
\end{array}
\]

In Ab, the cokernel of \( f : A \rightarrow B \) is the projection \( B \rightarrow B/fA \) to a quotient group of \( B \), and in many other such categories a cokernel is essentially a suitable quotient object. However, in categories without a zero object cokernels are not available. Hence we consider more generally certain "coequalizers".

**Coequalizers.** Given in \( C \) a pair \( f, g : a \rightarrow b \) of arrows with the same domain \( a \) and the same codomain \( b \), a coequalizer of \( \langle f, g \rangle \) is an arrow \( u : b \rightarrow e \) (or, a pair \( \langle u, w \rangle \)) such that (i) \( uf = ug \); (ii) if \( h : b \rightarrow c \) has \( hf = hg \), then \( h = hu \) for a unique arrow \( h' : e \rightarrow c \). The picture is

\[
\begin{array}{c}
a \\
\downarrow f \\
b \\
\downarrow g \\
\downarrow u \\
\downarrow 0
\end{array}
\]

A coequalizer \( u \) can be interpreted as a universal arrow as follows. Let \( \downarrow \downarrow \) denote the category which has precisely two objects and two
non-identity arrows from the first object to the second; thus the category is 
\[ \cdot \rightrightarrows \cdot \]. Form the functor category \( C^{\rightarrow\rightarrow} \). An object in \( C^{\rightarrow\rightarrow} \) is then a functor from \( \cdot \rightrightarrows \cdot \) to \( C \); that is, a pair \( \langle f, g \rangle : a \rightrightarrows b \) of parallel arrows \( a \rightrightarrows b \) in \( C \). An arrow in \( C^{\rightarrow\rightarrow} \) from one such pair \( \langle f, g \rangle \) to another \( \langle f', g' \rangle \) is a natural transformation between the corresponding functors; this means that it is a pair \( \langle h, k \rangle \) of arrows \( h : a \rightrightarrows a' \) and \( k : b \rightrightarrows b' \) in \( C \)

\[
\begin{array}{ccc}
a & \xrightarrow{f} & b \\
\downarrow & & \downarrow \\
h & \xrightarrow{g} & k \\
a' & \xleftarrow{f'} & b' \\
\end{array}
\]

\[ kg = g'h, \]

\[ k = f' h, \]

which make the \( f \)-square and the \( g \)-square commute. There is also a 
diagonal functor \( \Delta : C \rightarrow C^{\rightarrow\rightarrow} \), defined on objects \( c \) and arrows \( r \) of \( C \) as

\[
\begin{array}{ccc}
c & \xrightarrow{r} & c' \\
\downarrow & & \downarrow \\
r & \xrightarrow{r'} & c' \\
\end{array}
\]

in symbols, \( \Delta c = \langle 1_c, 1_c \rangle \) and \( \Delta r = \langle r, r' \rangle \). Now given the pair \( \langle f, g \rangle : a \rightrightarrows b \), an arrow \( h : b \rightrightarrows c \) with \( hf = hg \) is the same thing as an arrow \( \langle hf = hg, h \rangle : \langle f, g \rangle \rightarrow \langle 1_b, 1_c \rangle \) in the functor category \( C^{\rightarrow\rightarrow} \):

\[
\begin{array}{ccc}
a & \xrightarrow{f} & b \\
\downarrow & & \downarrow \\
h & \xrightarrow{g} & h \\
\end{array}
\]

\[ hf = hg. \]

In other words, the arrows \( h \) which "coequalize" \( f \) and \( g \) are the arrows from \( \langle f, g \rangle \) to \( \Delta \). Therefore a coequalizer \( \langle e, u \rangle \) of the pair \( \langle f, g \rangle \) is just a universal arrow from \( \langle f, g \rangle \) to the functor \( \Delta \).

Coequalizers of any set of maps from \( a \) to \( b \) are defined in the same way.

In \( \text{Ab} \), the coequalizer of two homomorphisms \( f, g : A \rightarrow B \) is the projection \( B \rightarrow B/(f - g)A \) on a quotient group of \( B \) (by the image of the difference homomorphism). In \( \text{Set} \), the coequalizer of two functions \( f, g : X \rightarrow Y \) is the projection \( p : Y \rightarrow Y/E \) on the quotient set of \( Y \) by the least equivalence relation \( E \subseteq Y \times Y \) which contains all \( \langle f(x), g(x) \rangle \) for \( x \in X \). The same construction, using the quotient topology, gives coequalizers in \( \text{Top} \).

Pushouts. Given in \( C \) a pair \( f : a \rightarrow b, g : a \rightarrow c \) of arrows with a common domain \( a \), a pushout of \( \langle f, g \rangle \) is a commutative square, such as that on

\[
\begin{array}{ccc}
a & \xrightarrow{f} & b \\
\downarrow & & \downarrow \\
h & \xrightarrow{g} & k \\
\end{array}
\]

such that to every other commutative square (right above) built on \( f, g \) there is a unique \( t : r \rightrightarrows s \) with \( tu = h \) and \( tv = k \). In other words, the pushout is the universal way of filling out a commutative square on the sides \( f, g \). It may be interpreted as a universal arrow. Let \( \cdot \rightrightarrows \cdot \) denote the category which looks just like that. An object in the functor category \( C^{\rightarrow\rightarrow} \) is then a pair of arrows \( \langle f, g \rangle \) in \( C \) with a common domain, while \( \Delta(c) = \langle 1_c, 1_c \rangle \) is the object function of an evident "diagonal" functor \( \Delta : C \rightarrow C^{\rightarrow\rightarrow} \). A commutative square \( hf = kg \) as on the right above can then be read as an arrow

\[
\begin{array}{ccc}
(f, g) & \xrightarrow{b} & (c, k) \\
\downarrow & & \downarrow \\
\Delta(s) & \rightarrow & \Delta(s)
\end{array}
\]

in \( C^{\rightarrow\rightarrow} \) from \( \langle f, g \rangle \) to \( \Delta s \). The pushout is a universal such arrow. Its vertex \( r \), which is uniquely determined up to (a unique) isomorphism, is often written as a coproduct "over \( a \)"

\[ r = b \coprod_A c = b \coprod_{\langle f, g \rangle} c, \]

and called a "fibered sum" or (the vertex of) a "cocartesian square".

In \( \text{Set} \), the pushout of \( \langle f, g \rangle \) always exists; it is the disjoint union \( b \coprod c \) with the elements \( f(x) \) and \( g(x) \) identified for each \( x \in a \). A similar construction gives pushouts in \( \text{Top} \) – they include such useful constructions as adjunction spaces. Pushouts exist in \( \text{Grp} \); in particular, if \( f \) and \( g \) above are monic in \( \text{Grp} \), the arrows \( u \) and \( v \) of the pushout square are also monic, and the vertex \( r \) is called the "amalgamated product" of \( b \) with \( c \).

Cokernel Pair. Given an arrow \( f : a \rightarrow b \) in \( C \), the pushout of \( f \) with \( f \) is called the cokernel pair of \( f \). Thus the cokernel pair of \( f \) consists of an object \( r \) and a parallel pair of arrows \( u, v : b \rightrightarrows r \), with domain \( b \), such that \( uf = vf \) and such that to any parallel pair \( h, k : b \rightrightarrows s \) with \( hf = kf \) there is a unique \( t : r \rightrightarrows s \) with \( tu = h \) and \( tv = k \):

\[
\begin{array}{ccc}
a & \xrightarrow{f} & b \\
\downarrow & & \downarrow \\
\uparrow t & & \uparrow k \\
\end{array}
\]

\[ uf = vf, \]

\[ hf = kf. \]
Colimits. The preceding cases all deal with particular functor categories and have the following pattern. Let $C$ and $J$ be categories ($J$ for index category, usually small and often finite). The diagonal functor

$$
\Delta : C \to C^J
$$

sends each object $c$ to the constant functor $\Delta c$ – the functor which has the value $c$ at each object $i \in J$ and the value $1$, at each arrow of $J$. If $f : c \to c'$ is an arrow of $C$, $\Delta f$ is the natural transformation $\Delta f : \Delta c \to \Delta c'$ which has the same value $f$ at each object $i$ of $J$. Each functor $F : J \to C$ is an object of $C^J$. A universal arrow $(r, u)$ from $F$ to $\Delta$ is called a colimit (a “direct limit” or “inductive limit”) diagram for the functor $F$. It consists of an object $r$ of $C$, usually written $r = \text{Colim } F$ or $r = \text{Colim } F$, together with a natural transformation $\tau : F \to \Delta r$ which is universal among natural transformations $\tau : F \to \Delta c$. Since $\Delta c$ is the constant functor, the natural transformation $\tau$ consists of arrows $\tau_i : F_i \to c$ of $C$, one for each object $i$ of $J$, with $\tau_i \cdot F_u = \tau_i$ for each arrow $u : i \to j$ of $J$. Pictorially, all the squares in the following schematic diagram (for a special choice $\Delta c$)

![Diagram](image)

must commute. It is convenient to visualize these diagrams with all the “bottom” objects identified. For this reason, a natural transformation $\tau : F \to \Delta c$, often written as $\tau : F \to c$, omitting $\Delta$, is called a cone from the base $F$ to the vertex $c$, as in the figure

![Diagram](image)

In this language, a colimit of $F : J \to C$ consists of an object $\text{Lim } F \in C$ and a cone $\mu : F \to \Delta (\text{Lim } F)$ from the base $F$ to the vertex $\text{Lim } F$ which is universal: For any cone $\tau : F \to \Delta c$ from the base $F$ there is a unique arrow $\tau : \text{Lim } F \to c$ with $\tau_i = \tau \mu_i$ for every index $i \in J$. We call $\mu$ the limiting cone or the universal cone (from $F$).

For example, let $J = \omega = \{0 \to 1 \to 2 \to 3 \to \ldots\}$ and consider a functor $F : \omega \to \text{Set}$ which maps every arrow of $\omega$ to an inclusion (subset in set). Such a functor $F$ is simply a nested sequence of sets $F_0 \subseteq F_1 \subseteq F_2 \subseteq \ldots$. The union $U$ of all sets $F_i$, with the cone given by the inclusion maps

![Diagram](image)

Exercise

1. In the category of commutative rings, show that $R \to R \otimes S \to S$, with maps $r \mapsto r \otimes 1, 1 \otimes s \mapsto s$, is a coproduct diagram.

2. If a category has (binary) coproducts and coequalizers, prove that it also has pushouts. Apply to $\text{Set, Grp}$, and $\text{Top}$.

3. In the category $\text{Matr}_k$ of $k \times k$ matrices, describe the coequalizer of two $m \times n$ matrices $A, B$ (i.e., of two arrows $n \to m$ in $\text{Matr}_k$).

4. Describe coproducts (and show that they exist in $\text{Cat}$, in $\text{Mon}$, and in $\text{Grp}$).

5. If $E$ is an equivalence relation on a set $X$, show that the usual set $X/E$ of equivalence classes can be described by a coequalizer in $\text{Set}$.

6. Show that $a$ and $b$ have a coproduct in $C$ if and only if the following functor is representable: $C(a, -) \times C(b, -) : C \to \text{Set}$, by $c \mapsto C(a, c) \times C(b, c)$.

7. (Every abelian group is a colimit of its finitely generated subgroups.) If $A$ is an abelian group, and $J_a$ the preorder with all finitely generated subgroups $S \subseteq A$ ordered by inclusion, show that $A$ is the colimit of the evident functor $J_a \to \text{Ab}$. Generalize.

4. Products and Limits

The limit notion is dual to that of a colimit. Given categories $C, J$, and the diagonal functor $\Delta : C \to C^J$, a limit for a functor $F : J \to C$ is a universal arrow $(r, \mu)$ from $\Delta F$ to $C$. It consists of an object $r$ of $C$, usually written $r = \text{Lim } F$ or $r = \text{Lim } F$ and called the limit object (the “inverse limit” or “projective limit”) of the functor $F$, together with a natural transformation $\nu : \Delta r \to F$ which is universal among natural transformations $\tau : \Delta c \to F$, for objects $c$ of $C$. Since $\Delta c : J \to C$ is the constant functor, this natural transformation $\tau$ consists of cone $\tau_i : c \to F_i$ of $C$ for each object $i$ of $J$ such that for every arrow $u : i \to j$ of $J$ one has $\tau_j = F_u \cdot \tau_i$. We may call $\tau_i : c \to F_i$ a cone to the base $F_i$ from the vertex $c$. (We say "cone to the base $F_i"$ rather than "cocone"). The universal property of $\nu$ is this: It is a cone to the base $F$ from the vertex $\text{Lim } F$; for any cone $\tau$ to $F$ from an object $c$, there is a unique arrow $\tau : c \to \text{Lim } F$ such that $\tau_i = \nu \cdot \tau_i$ for all $i$. The situation may be pictured as

![Diagram](image)
family of objects \(a_j \in C\), while a cone with vertex \(c\) and base \(a_j\) is just a \(J\)-indexed family of arrows \(f_j : c \to a_j\). A universal cone \(p_j : \Pi_j a_j \to a_j\) thus consists of an object \(\Pi_j a_j\), called the product of the factors \(a_j\), and of arrows \(p_j\), called the projections of the product, with the following universal property: To each \(J\)-indexed family (= cone) \(f_j : c \to a_j\) there is a unique \(f\)

\[
f : c \to \Pi_j a_j, \quad \text{with} \quad p_j f = f_j, \quad j \in J.
\]

The arrow \(f\) uniquely determined by this property is called the map (to the product) with components \(f_j, j \in J\). Also \(\{f_j | j \in J\} \to f\) is a bijection

\[
\Pi_j C(c, a_j) \cong C(c, \Pi_j a_j),
\]

natural in \(c\). Here the right hand product is that in \(C\), while the left-hand product is taken in \(\textbf{Set}\) (where we assume that \(C\) has small hom-sets).

Observe that the hom-functor \(C(c, -)\) carries products in \(C\) to products in \(\textbf{Set}\) (see §V.4). Products over any small set \(J\) exist in \(\textbf{Set}\), in \(\textbf{Top}\), and in \(\textbf{Grp}\); in each case they are just the familiar cartesian products.

Powers. If the factors in a product are all equal \((a_i = b \in C\) for all \(i)\), the product \(\prod_i a_i = \prod_i b_i\) is called a power and is written \(\prod_i b = b^J\), so that

\[
C(c, b^J) \cong C(c, b)_J,
\]

natural in \(c\). The power on the left is that in \(\textbf{Set}\), where every small power \(X^J\) exists (and is the set of all functions \(J \to X\)).

Equalizers. If \(J = \{1\}\), a functor \(F : \{1\} \to C\) is a pair \(f, g : b \to a\) of parallel arrows of \(C\). A limit object \(d\) of \(F\), when it exists, is called an equalizer (or, a "difference kernel") of \(f\) and \(g\). The limit diagram is

\[
d \to \xymatrix{ b \ar[r]^f & a, \quad f e = g e; \quad e \in C(a,b)}
\]

(the limit arrow \(e\) amounts to a cone \(a \to d \to b\) from the vertex \(d\)). The limit arrow is often called the equalizer of \(f\) and \(g\); its universal property reads: To any \(h : c \to d\) with \(fh = gh\) there is a unique \(c \to d\) with \(eh = h\).

In \(\textbf{Set}\), the equalizer always exists; \(d\) is the set \(\{x \in b | f x = g x\}\) and \(e : d \to b\) is the injection of this subset of \(b\) into \(b\). In \(\textbf{Top}\), the equalizer has the same description (\(d\) has the subspace topology). In \(\textbf{Ab}\) the equalizer \(d\) of \(f\) and \(g\) is the usual kernel of the difference homomorphism \(f - g : b \to a\).

Equalizers for any set of arrows from \(b\) to \(a\) are described similarly. Every equalizer \(e\) is necessarily a monic.
Pullbacks. If \( J = (\to \to) \), a functor \( F: (\to \to) \to C \) is a pair of arrows \( b \mapsto a \), \( d \) of \( C \) with a common codomain \( a \). A cone over such a functor is a pair of arrows from a vertex \( c \) such that the square (on the left)

\[
\begin{array}{ccc}
& c & \\
& k & \downarrow & \downarrow \ \\
& b & \rightarrow & a, \\
& b & \rightarrow & a \\
\end{array}
\]

(8)

commutes. A universal cone is then a commutative square of this form, with new vertex written \( b \times_a d \) and arrows \( p, q \) as shown on the right, such that for any square with vertex \( c \) there is a unique \( r: c \rightarrow b \times_a d \) with \( k = qr, h = pr \). The square formed by this universal cone is called a pullback square or a "cartesian square" and the vertex \( b \times_a d \) of the universal cone is called a pullback, a "fibered product", or a product over (the object) \( a \). This construction, possible in many categories, first became prominent in the category \( \text{Top} \). If \( g: d \rightarrow a \) is a "fiber map" (of some type) with "base" \( a \) and \( f \) is a continuous map into the base, then the projection \( p \) of the pullback is the "induced fiber map" (of the type considered).

The pullback of a pair of equal arrows \( f: b \rightarrow a \rightarrow b, f \), when \( f \) is an epimorphism, is called the kernel pair of \( f \). It is an object \( d \) and a pair of arrows \( p, q: d \rightarrow b \) such that \( fp =fq: d \rightarrow a \) and such that any pair \( h, k: c \rightarrow a \) with \( fh = fk \) can be written as \( h = pr, k = qr \) for a unique \( r: c \rightarrow d \).

If \( J = 0 \) is the empty category, there is exactly one functor \( 0 \rightarrow C \); namely, the empty functor; a cone over this functor is just an object \( c \in C \) (i.e., just a vertex). Hence a universal cone on \( 0 \) is an object \( t \) of \( C \) such that each object \( c \in C \) has a unique arrow \( c \rightarrow t \). In other words, there is a limit of the empty functor to \( C \) is a terminal object of \( C \).

Limits are sometimes defined for diagrams rather than for functors. In detail, let \( C \) be a category, \( UC \) the underlying graph of \( C \), and \( G \) any graph. Then a diagram in \( C \) of shape \( G \) is a morphism \( D: G \rightarrow UC \) of graphs. Now define a cone \( \mu: c \rightarrow D \) to be a function assigning to each object \( i \in G \) an arrow \( \mu_i: c \rightarrow D_i \) of \( C \) such that \( Dh \circ \mu_i = \mu_j \) for every arrow \( h: i \rightarrow j \) of the graph \( G \). This is just the previous definition of a cone (a natural transformation) \( \mu: \Delta c \rightarrow D_c \), coupled with the observation that this definition uses the composition of arrows in \( C \) but not in the domain \( G \) of \( D \). A limit for the diagram \( D \) is now a universal cone \( \lambda: c \rightarrow D \).

This variation on the definition of a limit yields no essentially new information. For, let \( FG \) be a free category generated by the graph \( G \), and \( P: G \rightarrow UF(G) \) the corresponding universal diagram. Then each diagram \( D: G \rightarrow UC \) can be written uniquely as \( D = UD'P \) for a (unique) functor \( D': FG \rightarrow C \), and one readily observes that limits (and limiting cones) for \( D' \) correspond exactly to those for \( D \).

**Exercises**

1. In \( \text{Set} \), show that the pullback of \( f: X \rightarrow Z \) and \( g: Y \rightarrow Z \) is given by the set of pairs \( \{(x, y) \mid x \in X, y \in Y, f(x) = g(y)\} \). Describe pullbacks in \( \text{Top} \).

2. Show that the usual cartesian product over an index set \( J \), with its projections, is a (categorical) product in \( \text{Set} \) and in \( \text{Top} \).

3. If the category \( J \) has an initial object \( 0 \), prove that every functor \( F: J \rightarrow C \) to any category \( C \) has a limit, namely \( F(0) \). Dualize.

4. In any category, prove that \( f: a \rightarrow b \) is epimorphism if and only if the following square is a pullback:

\[
\begin{array}{ccc}
& a & \\
& f & \downarrow & \downarrow \ \\
& b & \rightarrow & b \\
\end{array}
\]

5. In a pullback square (8), show that \( f \) is right cancellable implies \( g \) monic.

6. In \( \text{Set} \), show that the kernel pair of \( f: X \rightarrow Y \) is given by the equivalence relation \( E = \{(x, y) \mid x, y \in X \text{ and } f(x) = f(y)\} \), with the suitable maps \( E \rightarrow \text{Set} \).

7. (Kernel pairs via products and equalizers.) If \( C \) is the category of sets and \( \text{Set} \), show that the kernel pair of \( f: a \rightarrow b \) may be expressed in terms of the projections \( p_1, p_2: a \times b \rightarrow a \), \( p_1, p_2, e \), where \( e \) is the equalizer of \( f, p_2 \). (cf. Exercise 6). Dualize.

8. Consider the following commutative diagram

\[
\begin{array}{ccc}
& & \\
& & \downarrow \ \\
& & \\
\end{array}
\]

(a) If both squares are pullbacks, prove that the outside rectangle (with top and bottom edges the evident composites) is a pullback.

(b) If the outside rectangle and the right-hand square are pullbacks, so is the left-hand square.

9. (Equalizers via products and pullbacks.) Show that the equalizer of \( f, g: b \rightarrow a \) may be constructed as the pullback of \( (1_b, f): b \times b \rightarrow b \times a \rightarrow b \) (cf. Exercise 6). Dualize.

10. If \( C \) has pullbacks and a terminal object, prove that \( C \) has all products and equalizers.

5. Categories with Finite Products

A category \( C \) is said to have finite products if to any finite number of objects \( c_1, \ldots, c_n \) of \( C \) there exists a product diagram, consisting of a product object \( c_1 \times \cdots \times c_n \) and \( n \) projections \( p_i: c_1 \times \cdots \times c_n \rightarrow c_i, \) for \( i = 1, \ldots, n, \) with the usual universal property. In particular, \( C \) then has a product of \( n \) objects, which is simply a terminal object \( t \) in \( C \), as well
as a product for any two objects. The diagonal map \( \delta_c : c \rightarrow c \times c \) is defined for each \( c \) by \( p_1 \delta_c = 1_c = p_2 \delta_c \); it is a natural transformation.

**Proposition 1.** If a category \( C \) has a terminal object \( t \) and a product diagram \( a \leftarrow a \times b \rightarrow b \) for any two of its objects, then \( C \) has all finite products. The product objects provide, by \( \langle a, b \rangle \mapsto a \times b \), a bifunctor \( C \times C \rightarrow C \).

For any three objects \( a, b, \) and \( c \) there is an isomorphism
\[
\alpha = \alpha_{a,b,c} : a \times (b \times c) \cong (a \times b) \times c
\]
which are natural in \( a, b, \) and \( c \). For any object \( a \) there are isomorphisms
\[
\lambda = \lambda_a : t \times a \cong a = \eta = \eta_a : a \times t \cong a
\]
which are natural in \( a \), where \( t \) is the terminal object of \( C \).

**Proof.** A product of one object \( c \) is just the diagram \( c \rightarrow c \) formed with the identity map of \( c \), so is present in any category. Now suppose that any two objects \( a_1, a_2 \) of \( C \) have a product. If we choose one such product diagram \( a_1 \rightarrow a_1 \times a_2 \rightarrow a_2 \) for each pair of objects, then \( x \times x \) becomes a morphism when \( f_1 \times f_2 \) is defined on arrows \( f_1, f_2 \). One may then form a product of three objects \( a, b, \) and \( c \) by forming the iterated product object \( a \times (b \times c) \) with projections as in the diagram

\[
\begin{array}{ccc}
  a \times (b \times c) & \rightarrow & b \times c \\
  \downarrow \alpha & & \downarrow \\
  c & \leftrightarrow & a \times (b \times c) \\
  \uparrow \alpha & & \uparrow \\
  a & \leftrightarrow & a \times (b \times c)
\end{array}
\]

The projections to \( a \) and the two indicated composites give three arrows from \( a \times (b \times c) \) to \( a \), \( b \), and \( c \) respectively. By the universality of the given projections (from two factors) it follows readily that these three arrows form a product diagram for \( a, b \), and \( c \). Product diagrams for more factors can be found by iteration in much the same way. For three factors, one could also form a product diagram by the iteration \( (a \times b) \times c \); the uniqueness of the product objects then yields a unique isomorphism \( a \times b \times c \rightarrow (a \times b) \times c \) commuting with the given projections to \( a, b, \) and \( c \). This is the isomorphism \( \alpha \) of the proposition, and it is natural. Finally, since every object has a unique arrow to the terminal object \( t \), the diagram \( t \rightarrow a \rightarrow a \) is a product diagram for \( t \) and \( a \). The uniqueness of the product object \( a \times a \) then yields an isomorphism \( \lambda_a : t \times a \rightarrow a \), and similarly \( \eta_a : a \times t \rightarrow a \). Naturality of \( \lambda \) and \( \eta \) follows.

The dual result holds for finite coproducts; in particular a coproduct of no factors is an initial object. For \( m \) objects \( a_j \), a coproduct diagram consists of \( m \) injections \( i_j : a_j \rightarrow a_1 \bigcup \cdots \bigcup a_m \) and any map \( f : a_1 \bigcup \cdots \bigcup a_m \rightarrow c \) is uniquely determined by its \( m \) components
\[
\begin{align*}
f_1 & : a_1 \rightarrow c, \\
f_2 & : a_2 \rightarrow c, \\
\vdots & \\
f_m & : a_m \rightarrow c.
\end{align*}
\]

In particular, if \( C \) has both finite products and finite coproducts, the arrows
\[
a_1 \bigcup \cdots \bigcup a_m \rightarrow b_1 \times \cdots \times b_n
\]
from a coproduct to a product are determined uniquely by an \( m \times n \) matrix of arrows \( \sum_{k=1}^{n} f_{j,k} = p_k f_j : a_j \rightarrow b_k \), where \( j = 1, \ldots, m \) and \( k = 1, \ldots, n \). Since \( C \) has finite products and coproducts, there is then a "canonical" arrow
\[
a_1 \bigcup \cdots \bigcup a_m \rightarrow a_1 \times \cdots \times a_n
\]

of the coproduct to the product—namely, that arrow which has the identity \( n \times n \) matrix (identities on the diagonal and zeroes elsewhere).

This canonical arrow may be an isomorphism (in \( Ab \) or \( R-Mod \)), a proper monic (in \( Top \) or \( Set \)) or a proper epi (in \( Grp \) or \( Rng \)).

**Exercises**

1. Prove that the diagonal \( \delta_c : c \rightarrow c \times c \) is natural in \( c \).
2. In any category with finite products, prove that the following diagrams involving the canonical maps \( \alpha, \eta, \lambda \) of (1) and (2) always commute:

\[
\begin{array}{ccc}
a \times (b \times (c \times d)) & \xrightarrow{\alpha} & (a \times b) \times (c \times d) \\
\downarrow{\alpha} & & \downarrow{(a \times b) \times (a \times c) \times d} \\
a \times ((b \times c) \times d) & \xrightarrow{\lambda} & (a \times b) \times c \\
\downarrow{\eta} & & \downarrow{a \times c} \\
t \times (b \times c) & \xrightarrow{\lambda} & (t \times b) \times c \\
\downarrow{\lambda} & & \downarrow{a \times c} \\
(\lambda \times 1) & \xrightarrow{\lambda \times 1} & \lambda & \xrightarrow{\lambda \times 1} & \lambda \\
\end{array}
\]

3. (a) Prove that \( \text{Cat} \) has pullbacks (cf. Exercise II.6.5).
(b) Show that the comma categories \( (b \downarrow C) \) and \( (C \downarrow a) \) are pullbacks in \( \text{Cat} \).
4. Prove that \( \text{Cat} \) has all small coproducts.
5. If \( B \) has (finite) products show that any functor category \( B^C \) also has (finite) products (calculated "pointwise").
6. Groups in Categories

We return to the idea of the introduction. Let \( C \) be a category with finite products and a terminal object \( t \). Then a monoid in \( C \) is a triple \( \langle c, \mu, \eta \rangle \) together with an arrow \( \xi : c \rightarrow c \) which makes the diagram (with \( \delta \) the diagonal)

\[
\begin{array}{ccc}
    c \times c & \rightarrow & c \\
    \downarrow \mu & & \downarrow \delta \\
    c & = & c
\end{array}
\]

(3)

commute (this suggests that \( \xi \) sends each \( c \in C \) to its right inverse).

By similar diagrams, one may define rings in \( C \), lattices in \( C \), etc.; the process applies to any type of algebraic system defined by operations and identities between them.

It is a familiar fact that if \( G \) is an ordinary group, so is the function set \( G^X \) for any \( X \); indeed the product of two functions \( f, f' \) in \( G^X \) is defined pointwise, as \( (f \cdot f')(x) = f(x) \cdot f'(x) \). In the present context this construction takes the following form.

**Proposition 1.** If \( C \) is a category with finite products, then an object \( c \) is a group (or, a monoid) in \( C \) if and only if the hom functor \( C(-, c) \) is a group (respectively, a monoid) in the functor category \( \text{Set}^{C^{op}} \).

**Proof.** Each multiplication \( \mu \) for \( c \) determines a corresponding multiplication \( \overline{\mu} \) for the hom-set \( C(-, c) : \text{Set}^{op} \rightarrow \text{Set} \), as the composite

\[
C(-, c) \times C(-, c) \xrightarrow{\overline{\mu}} C(-, c) \xrightarrow{\mu} C(-, c)
\]

where \( v = \mu_{\xi} = C(-, \mu) \), while the first natural isomorphism is that given (cf. (4.4)) by the definition of the product object \( c \times c \). Conversely, given any natural \( v \) as above, the Yoneda lemma proves that there is a unique \( \mu : c \times c \rightarrow c \) with \( v = \mu_{\xi} \). A “diagram chase” shows that \( \mu \) is associative if and only if \( \overline{\mu} \) is; the chase uses the definition of the associativity iso-

morphism \( \times \) by its commutation with the projections of the three-fold product. The rest of the proof is left as an exercise.

Since the functor category \( \text{Set}^{C^{op}} \) always has finite products (Exercise 5.5) we can consider objects \( c \) in \( C \) such that \( C(-, c) \) is a group in this functor category even if the category \( C \) does not have finite products; however, I know no real use of this added generality.

**Exercises**

(Throughout, \( C \) is a category with finite products and a terminal object \( t \).)

1. Describe the category of monoids in \( C \), and show that it has finite products.
2. Show that the category of groups in \( C \) has finite products.
3. Show that a functor \( T : B \rightarrow \text{Set} \) is a group in \( \text{Set}^{B} \) if and only if each \( TB \), for \( b \) an object of \( B \), is a monoid. Note that \( T \) is an ordinary group if each \( TF \), for \( F \) in \( B \), is a morphism of groups.
4. (a) If \( A \) is an abelian group (in \( \text{Set} \)) show that its multiplication \( A \times A \rightarrow A \), its unit \( 1 \rightarrow A \), and its inverse \( A \rightarrow A \) are all morphisms of groups (where \( A \times A \) is regarded as the direct product group). Deduce that \( A \) with these structure maps is a group in \( \text{Grp} \). (b) Prove that every group in \( \text{Grp} \) has this form.

**Notes**

The Yoneda Lemma made an early appearance in the work of the Japanese pioneer N. Yoneda [1964] with time its importance has grown. Representable functors probably first appeared in topology in the form of “universal examples”, such as the universal examples of cohomology operations (for instance, in J. P. Serre’s 1953 calculations of the cohomology, modulo 2, of Eilenberg-Mac Lane spaces).

Universal arrows are unique only up to isomorphism; perhaps this lack of absolute uniqueness is why the notion was slow to develop. Examples had long been present; the bold step of really formulating the general notion of a universal arrow was taken by Samuel in 1948; the general notion was then lavishly popularized by Bourbaki. The idea that the ordinary cartesian products could be described by universal properties of their projections was formulated about the same time (Mac Lane [1948, 1950]). On the other hand the notions of limit and colimit have a long history in various concrete examples. Thus colimits were used in the proofs of theorems in which infinite abelian groups are represented as unions of their finitely generated subgroups. Limits (over ordered sets) appear in the \( p \)-adic numbers of Hensel and in the construction of Czech homology and cohomology by limit processes as formalized by Pontrjagin. An adequate treatment of the natural isomorphisms occurring for such limits was a major motivation of the first Eilenberg-Mac Lane paper on category theory [1945]. E. H. Moore’s general analysis (about 1913) used limits over certain directed sets. In all these classical cases, limits appeared only for functors \( F : J \rightarrow C \) with \( J \) a linearly or partially ordered set. Then Kan [1960] took the step of considering limits for all functors, while Freyd [1964] for the general case used the word “root” in place of “limit”. His followers have chosen to extend the original word “limit” to this general meaning. Properties special to limits over directed sets will be studied in Chapter IX.
IV. Adjoints

1. Adjunctions

We now present a basic concept due to Kan, which provides a different formulation for the properties of free objects and other universal constructions. As motivation, we first reexamine the construction (§III.1) of a vector space \( V_X \) with basis \( X \). For a fixed field \( K \) consider the functors

\[
\text{Set} \xrightarrow{\mathcal{V}} \text{Vct},
\]

where, for each vector space \( W \),

\[
U(W) = \{ \text{all vectors in } W \},
\]

so that \( U \) is the forgetful functor, while, for any set \( X \),

\[
V(X) = \{ \text{vector space with basis } X \}.
\]

The vectors of \( V(X) \) are thus the formal finite linear combinations \( \sum r_i x_i \), with scalar coefficients \( r_i \in K \) and with each \( x_i \in X \), with the evident vector operations. Each function \( g : X \rightarrow U(W) \) extends to a unique linear transformation \( f : V(X) \rightarrow W \), given explicitly by

\[
f(\sum r_i x_i) = \sum r_i g(x_i)
\]

(i.e., formal linear combinations in \( V(X) \) to actual linear combinations in \( W \)).

This correspondence \( g \mapsto f \) has an inverse \( f \mapsto g \), the restriction of \( / \) to \( X \), hence is a bijection

\[
\psi : \text{Set}(U(W), X) \cong \text{Set}(X, U(W)).
\]

This bijection \( \psi = \psi_{X,W} \) is defined "in the same way" for all sets \( X \) and all vector spaces \( W \). This means that the \( \psi_{X,W} \) are the components of a natural transformation \( \psi \) when both sides above are regarded as functors of \( X \) and \( W \). It suffices to verify naturality in \( X \) and in \( W \) separately. Naturality in \( X \) means that for each arrow \( h : X' \rightarrow X \) the diagram

\[
\begin{array}{ccc}
\text{Vct}(V(X), W) & \xrightarrow{\psi} & \text{Set}(X, U(W)) \\
(\psi h)^* \downarrow & & \downarrow \psi \\
\text{Vct}(V(X'), W) & \xrightarrow{\psi} & \text{Set}(X', U(W))
\end{array}
\]

where \( h^* g = g = h \), will commute. This commutativity follows from the definition of \( \psi \) by a routine calculation, as does also the naturality in \( W \).

Note next several similar examples.

The free category \( C = \mathcal{F}G \) on a given (small) graph \( G \) is a functor

\[
\text{Grph} \rightarrow \text{Cat}:
\]

it is related to the forgetful functor \( U : \text{Cat} \rightarrow \text{Grph} \) by the fact (§11.7) that each morphism \( D : \text{Cat} \rightarrow \text{Grph} \) of graphs extends to a unique map \( F : \mathcal{F}G \rightarrow B \) of categories; moreover, \( D : \mathcal{F}G \rightarrow B \) is a natural isomorphism

\[
\text{Cat}(\mathcal{F}G, B) \cong \text{Grph} (G, U B).
\]

In the category of small sets, each function \( g : S \times T \rightarrow R \) of two variables can be treated as a function \( \psi : S \rightarrow \text{hom}(T, R) \) of one variable (in \( S \)) whose values are functions of a second variable (in \( T \)); explicitly,

\[
[(\psi g)_s]_t = g(s, t) \quad \text{for } s \in S, \quad t \in T.
\]

This describes \( \psi \) as a bijection

\[
\psi : \text{hom}(S \times T, R) \cong \text{hom}(S, \text{hom}(T, R)).
\]

It is natural in \( S, T, \) and \( R \). If we hold the set \( T \) fixed and define functors

\[
F, G : \text{Set} \rightarrow \text{Set}
\]

by \( F(S) = S \times T, \ G(R) = \text{hom}(T, R) \), the bijection takes the form

\[
\text{hom}(F(S), R) \cong \text{hom}(S, \text{hom}(T, R))
\]

natural in \( S \) and \( R \), and much like the previous examples.

For modules \( A, B, \) and \( C \) over a commutative ring \( K \) there is a similar isomorphism

\[
\text{hom}(A \otimes_K B, C) \cong \text{hom}(A, \text{hom}_K(B, C))
\]

natural in all three arguments.

**Definition.** Let \( A \) and \( X \) be categories. An adjunction from \( X \) to \( A \) is a triple \( (F, G, \eta) : X \rightarrow A \), where \( F \) and \( G \) are functors

\[
A(Fx, a) \cong X(x, Ga)
\]

natural in all three arguments.

The left hand side \( A(Fx, a) \) is the bifunctor

\[
X^{op} \times A \xrightarrow{A(Fx, a)} A^{op} \times A \cong \text{Set}
\]

which sends each pair of objects \( (x, a) \) to the hom-set \( A(Fx, a) \), and the right hand side is a similar bifunctor

\[
X^{op} \times A \rightarrow \text{Set}.
\]

Therefore the naturality of the bijection \( \eta \) means that for all \( k : a \rightarrow a' \) and \( h : x' \rightarrow x \) both the diagrams:

\[
\begin{array}{ccc}
A(Fx, a) & \xrightarrow{\eta} & X(x, Ga) \\
\downarrow k & & \downarrow (Gk)_* \\
A(Fx', a') & \xrightarrow{\eta} & X(x', Ga')
\end{array}
\]

\[
\begin{array}{ccc}
A(Fx, a) & \xrightarrow{\eta} & X(x, Ga) \\
\downarrow (Fh)^* & & \downarrow h^* \\
A(Fx', a') & \xrightarrow{\eta} & X(x', Ga')
\end{array}
\]

are commutative.
will commute. Here \( k \) is short for \( A(Fx, k) \), the operation of composition with \( k \), and \( h^* = X(h, Ga) \).

This discussion assumes that all the hom-sets of \( X \) and \( A \) are small.

If not, we just replace \( \text{Set} \) above by a suitable larger category \( \text{Ens} \) of sets.

An adjunction may also be described without hom-sets directly in terms of arrows. It is a bijection which assigns to each arrow \( f: Fx \to a \) an arrow \( \varphi f = \text{rad} f: x \to Ga \), the \textit{right adjoint} of \( f \), in such a way that the naturality conditions of (2),

\[
\varphi(f \circ h) = \varphi f \circ h, \quad \varphi(k \cdot f) = Gk \cdot \varphi f,
\]

hold for all \( f \) and all arrows \( h: x' \to x \) and \( k: a \to a' \). It is equivalent to require that \( \varphi^{-1} \) be natural; i.e., that for every \( h, k \) and \( g: x \to Ga \) one has

\[
\varphi^{-1}(gh) = \varphi^{-1} g \cdot Fh, \quad \varphi^{-1}(Gk \cdot g) = k \cdot \varphi^{-1} g.
\]

Given such an adjunction, the functor \( F \) is said to be a \textit{left-adjoint} for \( G \), while \( G \) is called a \textit{right adjoint} for \( F \). (Some authors write \( F \to G \); others say that \( F \) is the "adjoint" of \( G \) and \( G \) the "coadjoint" of \( F \), but other authors say the opposite; therefore we shall stick to "left" and "right" adjoints.)

Every adjunction yields a universal arrow. Specifically, set \( a = Fx \) in (1). The left hand hom-set of (1) then contains the identity \( 1: Fx \to Fx \); call its \( \varphi \)-image \( \eta_x \). By Yoneda's Proposition III.2.1, this \( \eta_x \) is a universal arrow

\[ \eta_x: x \to GFx, \quad \eta_x = \varphi(1_{Fx}) \]

from \( x \in X \) to \( G \). The adjunction gives such a universal arrow \( \eta_x \) for every object \( x \). Moreover, the function \( x \to \eta_x \) is a natural transformation \( I_x \to GF \) because every diagram

\[
\begin{array}{ccc}
x & \xrightarrow{\eta_x} & GFx \\
\downarrow h & \downarrow GFh & \\
x' & \xrightarrow{\varphi f} & GFx
\end{array}
\]

is commutative. This one proves by the calculation

\[ GFh \cdot \varphi(1_{Fx}) = \varphi(Fh \cdot 1_{Fx}) = \varphi(1_{Fh} \cdot Fh) = \varphi(1_{Fx}) \cdot h. \]

Based on the Eq. (3) describing the naturality of \( \varphi \). This calculation may also be visualized by the commutative diagram

\[
\begin{array}{ccc}
A(Fx', Fx') \xrightarrow{GFh} A(Fx, Fx) \xrightarrow{\varphi(1_{Fx})} A(Fx, Fx) \\
\downarrow & \downarrow \varphi & \\
X(x', GFx') \xrightarrow{\eta_{Fx'}} X(x', GFx) \xrightarrow{\eta_x} X(x, GFx),
\end{array}
\]

where \( h^* = X(h, 1) \) and \( h_* = X(1, h) \).

The bijection \( \varphi \) can be expressed in terms of the arrows \( \eta_x \) as

\[ \varphi(f) = G(f) \cdot \eta_x \quad \text{for} \quad f: Fx \to a; \]

indeed, by the naturality (3) of \( \varphi \) we may compute that

\[ \varphi(f) = \varphi(f \cdot 1_{Fx}) = Gf \cdot \varphi(1_{Fx}) = Gf \cdot \eta_x. \]

This computation may be visualized by chasing \( 1 \) around the commutative square

\[
\begin{array}{ccc}
A(Fx, Fx) & \xrightarrow{\varphi} & X(x, GFx) \\
\downarrow f & & \downarrow \eta_x \downarrow & \\
A(Fx, a) & \xrightarrow{\varphi f} & X(x, Ga)
\end{array}
\]

Dually, the adjunction gives a universal arrow from \( F \). Indeed, set \( x = Ga \) in the adjunction (1). The identity arrow \( 1: Ga \to Ga \) is now present in the right hand hom-set; its image under \( \varphi^{-1} \) is called \( \epsilon_a \).

\[ \epsilon_a: FGa \to a, \quad \epsilon_a = \varphi^{-1}(1_{Ga}), \quad a \in A, \]

and is a universal arrow from \( F \) to \( a \). As before, \( \epsilon \) is a natural transformation \( \epsilon: GF \to 1_A \),

\[ \varphi^{-1}(g) = \epsilon_a \cdot Fg \quad \text{for} \quad g: x \to Ga. \]

Finally, take \( x = Ga \). Then \( \epsilon_a = \varphi^{-1}(1_{Ga}) \) gives, by the formula (5) for \( \varphi \),

\[ 1_{Ga} = \varphi(\epsilon_a) = G(\epsilon_a) \cdot \eta_{Ga}. \]

This asserts that the composite natural transformation

\[ G \xrightarrow{\epsilon_a} GF \xrightarrow{\eta_{Ga}} G \]

is the identity transformation.

To summarize, we have proved

\textbf{Theorem 1.} An adjunction \((F, G, \varphi): X \to A\) determines

(i) A natural transformation \( \eta: I_x \to GF \) such that for each object \( x \) the arrow \( \eta_x \) is universal to \( G \) from \( x \), while the right adjoint of each \( f: Fx \to a \) is

\[ \varphi(f) = Gf \cdot \eta_x: x \to Ga; \]

(ii) A natural transformation \( \epsilon: GF \to 1_A \) such that each \( a \) is universal to \( F \) from \( A \), while each \( g: x \to Ga \) has left adjoint

\[ \varphi^{-1}(g) = \epsilon_a \cdot Fg: Fx \to a. \]

Moreover, both the following composites are the identities (of \( G \), resp. \( F \)).

\[ G \xrightarrow{\epsilon_a} GF \xrightarrow{\eta_{Ga}} G, \quad F \xrightarrow{\epsilon_a} FG \xrightarrow{\eta_{Fx}} F. \]
We call $\eta$ the unit and $\varepsilon$ the counit of the adjunction. (Formerly, we called $\eta$ a "front adjunction" and $\varepsilon$ a "back adjunction").

The given adjunction is actually already determined by various portions of all these data, in the following sense.

**Theorem 2.** Each adjunction $\langle F, G, \varphi \rangle : X \to A$ is completely determined by the items in any one of the following lists:

(i) Functors $F$, $G$, and a natural transformation $\eta : 1_x \to GFx$ such that each $\eta_x : x \to GFx$ is universal to $G$ from $x$. Then $\varphi$ is defined by (6).

(ii) The functor $G : A \to X$ and for each $x \in A$ an object $F_0x \in A$ and a universal arrow $\eta_x : x \to GF_0x$ from $x$ to $G$. Then the functor $F$ has object function $F_0$ and is defined on arrows $h : x \to x'$ by $GFh \cdot \eta_x = \eta_{x'} \cdot h$.

(iii) Functors $F$, $G$, and a natural transformation $\varepsilon : FG \to 1_A$ such that each $\varepsilon_a : FGa \to a$ is universal from $F$ to $a$. Here $\varphi^{-1}$ is defined by (7).

(iv) The functor $F : X \to A$ and for each $a \in A$ an object $G_0a \in X$ and an arrow $\eta_a : G_0a \to a$ universal from $F$ to $a$.

(v) Functors $F$, $G$ and natural transformations $\eta : I_k \to GF$ and $\varepsilon : FG \to 1_A$ such that both composites (8) are the identity transformations. Here $\varphi$ is defined by (6) and $\varphi^{-1}$ by (7).

Because of (v), we often denote the adjunction $\langle F, G, \varphi \rangle$ by $\langle F, G, \eta, \varepsilon \rangle : X \to A$.

**Proof.** Ad (i): The statement that $\eta_x$ is universal means that to each $f : x \to Ga$ there is exactly one $g$ as in the commutative diagram

$$\begin{array}{ccc}
Fx & \xrightarrow{\eta_x} & GFx \\
\downarrow f & & \downarrow g \\
X & \xrightarrow{\alpha} & Ga
\end{array}$$

This states precisely that $\psi(g) = Gg \cdot \eta_x$ defines a bijection

$$\psi : A(Fx, a) \to X(x, Ga)$$

This bijection $\psi$ is natural in $x$ because $\eta_x$ is natural, and natural in $a$ because $G$ is a functor, hence gives an adjunction $\langle F, G, \psi \rangle$. In case $\eta$ was the unit obtained from an adjunction $\langle F, G, \varphi \rangle$, then $\psi = \varphi$.

The data (ii) can be expanded to (i), and hence determine the adjunction. In (ii) we are given simply a universal arrow $\langle F_0x, \eta_x \rangle$ for every object $x \in X$; we shall show that there is exactly one way to make $F_0$ the object function of a functor $F$ for which $\eta : I_k \to GF$ will be natural. Specifically, for each $h : x \to x'$ the universality of $\eta_x$ states that there is exactly one arrow (dotted)

$$\begin{array}{ccc}
F_0x & \xrightarrow{\eta_x} & GF_0x \\
\downarrow h & & \downarrow \eta_x^h \\
F_0x' & \xrightarrow{\eta_x'} & GF_0x'
\end{array}$$

which can make the diagram commute. Choose this arrow as $Fh : F_0x \to F_0x'$; the commutativity states that $\eta$ is now natural, and it is easy to check that this choice of $Fh$ makes $F$ a functor.

The proofs of parts (iii) and (iv) are dual.

To prove part (v) we use $\eta$ and $\varepsilon$ to define functions

$$A(Fx, a) \xrightarrow{\alpha} X(x, Ga)$$

by $\varphi f = Gf \cdot \eta_x$ for each $f : Fx \to a$ and $\theta g = \varepsilon_a \cdot Fg$ for each $g : x \to Ga$. Then since $G$ is a functor and $\eta$ is natural

$$\varphi \theta g = G\varepsilon_a \cdot GFg \cdot \eta_x = G\varepsilon_a \cdot \eta_{Ga} \cdot g.$$

But our hypothesis (8) states that $G\varepsilon_a \cdot \eta_{Ga} = 1$. Hence $\varphi \theta = \text{id}$. Dually $\theta \varphi = \text{id}$. Therefore $\varphi$ is a bijection (with inverse $\theta$). It is clearly natural, hence is an adjunction (and, if we started with an adjunction, it is the one from which we started).

This theorem is very useful. For example, parts (ii) and (iv) construct an adjunction whenever we have a universal arrow from (or to) every object of a given category. For example, the category $C$ has finite products when for each pair $\langle a, b \rangle \in C \times C$ there is a universal arrow from $\Delta : C \to C \times C$ to $\langle a, b \rangle$. By the theorem above we conclude that the function $\langle a, b \rangle \to a \times b$ giving the product object is actually a functor $C \times C \to C$, and that this functor is right adjoint to the diagonal functor $\Delta$:

$$\varphi : (C \times C)(a, b) \cong C(a, a \times b).$$

Using the definition of the arrows in $C \times C$, this is

$$\varphi : C(a, c) \times C(c, b) \cong C(a, c \times b).$$

The counit of this adjunction (set $c = a \times b$ on the right) is an arrow $\langle a \times b, a \times b \rangle \to \langle a, b \rangle$; it is thus just a pair of arrows $a \cdot a \times b \to b$; namely, the projections $p : a \times b \to a$ and $q : a \times b \to b$ of the product. The adjunction $\varphi^{-1}$ sends each $c : a \times b$ to the pair $\langle pq, qf \rangle$; this is the way in which $\varphi$ is determined by the counit $\varepsilon$.

Similarly, if the category $C$ has coproducts $\langle a, b \rangle \mapsto a + b$, they define a functor $C \times C \to C$ which is a left adjoint to $\Delta$:

$$C(a + b, c) \cong (C \times C)(a, b + c).$$

All the other examples of limits (when they always exist) can be similarly read as examples of adjoints. In many further applications, it turns out that proving universality is an easy way of showing that adjoints are present.
On the other hand, part (v) of the theorem describes an adjunction by two simple identities

\[ \begin{array}{ccc} F & \xrightarrow{F \circ \eta_m} & GF \\ \downarrow{\varepsilon} & & \downarrow{\cong} \\ A & \xrightarrow{\gamma} & G \end{array} \]

(9)

on the unit and counit of the adjunction. These triangular identities make no explicit use of the objects of the categories \( A \) and \( X \), and so are easy to manipulate. As we shall soon see, this is convenient for discussing properties of adjunctions. (For some authors, these identities are said to make \( \eta \) a "quasi-inverse" to \( \varepsilon \).)

**Corollary 1.** Any two left-adjoints \( F \) and \( F' \) of a functor \( G: A \to X \) are naturally isomorphic.

The proof is just an application of the fact that a universal arrow, like an initial object, is unique up to isomorphism. Explicitly, adjunctions \( \langle F, G, \phi \rangle \) and \( \langle F', G, \phi' \rangle \) give to each \( x \) two universal arrows \( x \to GFx \) and \( x \to GF'x \); hence there is a unique isomorphism \( \theta_x: Fx \to F'x \) with \( G\theta_x \cdot \eta_x = \eta_x' \); it is easy to verify that \( \theta: F \to F' \) is natural.

**Corollary 2.** A functor \( G: A \to X \) has a left adjoint if and only if, for each \( x \in X \), the functor \( X(x, GA) \) is representable as a functor of \( A \). If \( \phi: A(F_0, x, a) \cong X(x, Ga) \) is a representation of this functor, then \( F_0 \) is the object function of a left-adjoint of \( G \) for which the bijection \( \phi \) is natural in \( a \) and gives the adjunction.

This is just a restatement of part (ii) of the theorem. Equivalently, \( G \) has a left-adjoint if and only if there is a universal arrow to \( G \) from every \( x \in X \).

We leave the reader to state the duals.

Adjoints of additive functors are additive.

**Theorem 3.** If the additive functor \( G: A \to M \) between Ab-categories \( A \) and \( M \) has a left adjoint \( F: M \to A \), then \( F \) is additive and the adjunction bijections

\[ \phi: A(Fm, a) \cong M(m, Ga) \]

are isomorphisms of abelian groups (for all \( m \in M, a \in A \)).

**Proof.** If \( \eta: I \to GF \) is the unit of the adjunction, then \( \phi \) may be written as \( \phi = \gamma \varepsilon \eta_m \) for any \( Fm \to a \). If also \( F' : Fm \to a \), the additivity of \( G \) gives

\[ \phi(f + f') = GF(f + f') \eta_m = (Gf + GFf') \eta_m = Gf \eta_m + GFf' \eta_m = \phi f + \phi f'. \]

Therefore \( \phi \) is a morphism of abelian groups. Next take \( g, g': m \to n \) in \( M \). Since \( \eta \) is natural,

\[ GF(g + g') \eta_m = \eta_n(g + g') = \eta_n g + \eta_n g' \]

On the other hand, since \( G \) is additive,

\[ G(Fg + Fg') \eta_m = (G_GFg + GFg') \eta_m = GGFg \eta_m + GGFg' \eta_m = \eta_m g + \eta_m g'. \]

The equality of these two results and the universal property of \( \eta_m \) show that \( F(g + g') = Fg + Fg' \). Hence \( F \) is additive.

Dually, any right adjoint of an additive functor is additive.

**Exercises**

1. Show that Theorem 2 can have an added clause (and its dual):

   (vi) A functor \( G: A \to X \) and for each \( x \in X \) a representation \( \phi_x \) of the functor \( X(x, GA) \to A \).

2. (Lawvere). Given functors \( G: A \to X \) and \( F: X \to A \), show that each adjunction \( \langle F, G, \phi \rangle \) can be described as an isomorphism \( \theta \) of comma categories such that the following diagram commutes

\[ \begin{array}{ccc} \theta: (F \downarrow G) & \cong & (L \downarrow G) \\ \downarrow \downarrow \downarrow \downarrow \downarrow & & \downarrow \downarrow \downarrow \downarrow \downarrow \\ X \times A & \to & X \times A \end{array} \]

Here the vertical maps have components the projection functors \( P \) and \( Q \).

3. For the adjunction \( \langle \Delta, \phi \rangle \) - product right adjoint to diagonal - show that the unit \( \delta_c: c \to \Delta c \) for each object \( c \in C \) is the unique arrow such that the diagram

\[ \begin{array}{ccc} C & \xrightarrow{\delta} & \Delta C \\ \downarrow{\cong} & & \downarrow{\cong} \\ C \times C & \xrightarrow{\beta \times \beta} & C \times C \end{array} \]

commutes. (This arrow \( \delta \) is often called the diagonal arrow of \( c \).) If \( C = \text{Set} \), show that \( \delta_c = \langle x, x \rangle \) for \( x \in c \).

4. (Pare). Given functors \( G: A \to X \) and \( K: X \to A \) and natural transformations \( \varepsilon: KG \Rightarrow id_A, \eta: id_G \Rightarrow KG \) such that \( G \varepsilon \eta = G \eta \varepsilon = G \eta \zeta = \eta \zeta G \), prove that \( \varepsilon K - K \varepsilon: K \Rightarrow K \) is an idempotent in \( A^2 \) and that \( G \) has a left adjoint if and only if this idempotent splits; explicitly if \( \varepsilon K = K \varepsilon \) splits as \( \alpha \beta = 1 \) and \( \beta K = K \alpha \), then \( F \) is a left adjoint of \( G \) with unit \( G \beta \cdot \eta \) and counit \( \varepsilon \cdot \alpha G \).

2. Examples of Adjoints

We now summarize a number of examples of adjoints, beginning with a table of left-adjoints of typical forgetful functors.
Examples of Adjoints

### Table of Adjoints

<table>
<thead>
<tr>
<th>Diagonal functor</th>
<th>Adjoint</th>
<th>Unit</th>
<th>Count</th>
</tr>
</thead>
</table>
| \( 
\begin{align*}
\Delta : C &\to C \times C \\
I : C \times C &\to C \\
\Pi : C \times C &\to C \\
\end{align*}
\) | Left: Coproduct | (pair of injections) | "folding" map |
| Right: Product | Diagonal arrow | (pair of projections) | |
| \( \phi : \text{Vet}(V, \text{Vet}(V, W)) \to \text{Vet}(W, \text{Vet}(V, K)) \) | | | |

In the case of limits, the form of the unit depends on the number of connected components of \( J \). Here a category \( J \) is called connected when to any two objects \( j, k \in J \) there is a finite sequence of arrows

\[
j = j_0 \rightarrow j_1 \rightarrow \cdots \rightarrow j_{n-1} \rightarrow j_n = k \quad \text{(both directions possible)}
\]

joining \( j \) to \( k \) (see Exercises 7, 8).

Duality functors provide further examples. For vector spaces \( V, W \) over a field \( K \), the dual \( \tilde{D} \) is a contravariant functor on \( \text{Vet} \) to \( \text{Vet} \), given on objects by \( \tilde{D}V = \text{Vet}(V, K) \) with the usual vector space structure and on arrows \( h : V \to W \) as \( \tilde{D}h : \tilde{D}W \to \tilde{D}V \), where \( \tilde{D}h)j = h \) for each \( f : W \to K \). A function

\[
\varphi = \varphi_{V, W} : \text{Vet}(V, \text{Vet}(W, K)) \to \text{Vet}(W, \text{Vet}(V, K))
\]

is defined for \( h : V \to \tilde{D}W \) by \( ([w]h)v = (hv)w \) for all \( v \in V, w \in W \).

Since \( \varphi_{V, W} \varphi_{V, W} = \text{id} \) for all \( V, W \), each \( \varphi \) is a bijection. The contravariant functor \( \tilde{D} \) leads to two different (covariant) functors with the same object function,

\[
\tilde{D} : \text{Vet} \to \text{Vet}, \quad \tilde{D}^\circ : \text{Vet} \to \text{Vet}^\circ,
\]

defined (as usual) for arrows \( h^\circ : W \to V \) and \( h : V \to W \) by

\[
\tilde{D}^\circ h = \tilde{D}h : \tilde{D}W \to \tilde{D}V, \quad \tilde{D}^\circ h = (\tilde{D}h)^\circ : \tilde{D}V \to \tilde{D}W.
\]

The bijection \( \varphi \) of (1) above may now be written as

\[
\text{Vet}^\circ(\tilde{D}^\circ W, V) \cong \text{Vet}(W, D V),
\]

defined as usual for arrows \( h^\circ : W \to V \) and \( h : V \to W \) by

\[
\tilde{D}h^\circ = \tilde{D}h : \tilde{D}W \to \tilde{D}V, \quad \tilde{D}^\circ h = (\tilde{D}h)^\circ : \tilde{D}V \to \tilde{D}W.
\]

The bijection \( \varphi \) of (1) above may now be written as

\[
\text{Vet}^\circ(\tilde{D}^\circ W, V) \cong \text{Vet}(W, D V),
\]

natural in \( V \) and \( W \). Therefore \( \tilde{D}^\circ \) is the left adjoint of \( D \). (Warning: It is not a right adjoint of \( D \), see § V.5, Exercise 2.)
Examples of Adjoints

(set \( V = D^{op} W \) in (2) is this map \( \eta_W = \kappa_W : W \to D^{op} W \), and the counit \( \varepsilon \) is an arrow \( \varepsilon_V : D^{op} D V \to V \) in \( \text{Vec}^{op} \) which turns out to be \( \varepsilon_V = (\kappa_V)^{op} \) for the same \( \kappa \).

This example illustrates the way in which adjunctions may replace isomorphisms of categories. For finite dimensional vector spaces, \( D \) and \( D^{op} \) are isomorphisms; for the general case, this is not true, but \( D \) is the right adjoint of \( D^{op} \).

This example also bears on adjoints for other contravariant functors.

Two contravariant functors \( \mathcal{F} \) from \( X \to Y \) and \( \mathcal{T} \) from \( Y \to Z \) are “adjoint on the right” (Freyd) when there is a bijection \( A(a, \mathcal{T} x) \cong X(x, \mathcal{F} a) \), natural in \( a \) and \( x \). We shall not need this terminology, because we can replace \( \mathcal{F} \) and \( \mathcal{T} \) by the covariant functors \( \mathcal{S} : A^{op} \to X \) and \( \mathcal{T} : X^{op} \to A \) and form the dual \( S^{op} \) : \( A \to X^{op} \), also covariant; thus the natural bijection above becomes \( X^{op}(S^{op} a, x) \cong A(a, \mathcal{T} x) \), and so states that \( S^{op} \) is left adjoint (in our usual sense) to \( T \) or, equivalently, that \( T^{op} \) is left adjoint to \( S \). It is not necessarily equivalent to say that \( T \) and \( S \) are adjoint “on the left.”

The next three sections will be concerned with three other types of adjoints: A left adjoint to an inclusion functor (of a full subcategory) is called a reflection; certain other special sorts of adjoints are “equivalences” of categories. Some other amusing examples of adjoints are given in the exercises to follow, some of which require knowledge of the subject matter involved. (Goguen [1971] shows for finite state machines that the functor “minimal realization” is left adjoint to the functor “behavior”. The reader is urged to find his own examples as well.

Exercises

1. For \( K \) a field and \( V \) a vector space over \( K \), there is an “exterior algebra” \( E(V) \), which is a graded, anticommutative algebra. Show that \( E \) is the left adjoint of a suitable forgetful functor (one which is not faithful).

2. Show that the functor \( U : R\text{-Mod} \to \text{Ab} \) has not only a left adjoint \( A \to R \otimes A \) but also a right adjoint \( A \to \text{hom}_R(A, R) \).

3. For \( K \) a field, let \( \text{Lie}_K \) be the category of all (small) Lie algebras \( L \) over \( K \), with arrows the morphisms of \( K \)-modules which also preserve the Lie bracket operation \( [a, b] \to [a, b] \). Let \( V : \text{Alg}_K \to \text{Lie}_K \) be the functor which assigns to each (associative) algebra \( A \) the Lie algebra \( VA \) on the same vector space, with bracket \( [a, b] = ab + ba \) for \( a, b \in A \). Using the Poincaré-Birkhoff-Witt Theorem show that the functor \( E \), where \( EL \) is the enveloping associative algebra of \( L \), is a left adjoint for \( V \).

4. Let \( \text{Ring} \) denote the category of rings \( R \) which do not necessarily have an identity element for multiplication. Show that the standard process of adding an identity to \( R \) provides a left adjoint for the forgetful functor \( \text{Ring} \to \text{Ring} \) (forget the presence of the identity).

5. If a monoid \( M \) is regarded as a discrete category, with objects the elements \( x \in M \), then the multiplication of \( M \) is a bifunctor \( \mu : M \times M \to M \). If \( M \) is a group, show that \( M/(x, x^2) = M/(x, x^2) \) provides right adjoints for both functors \( \mu(x, -) \) and \( \mu(-, y) : M \to M \). Conversely, does the presence of such adjoints make a monoid into a group?

6. Describe units and counits for pullback and products.

7. If the category \( J \) is a disjoint union (coproduct); \( \text{E}_I \) of categories \( \text{E}_I \) for index \( i \) in some set \( K \). With \( I \to J \) the injections of the coproduct, then each functor \( F : J \to C \) determines functors \( F_I : F : J \to C \).

(a) Show that \( \lim F \) is \( \lim F^I \) if the limits on the right exist.

(b) Show that every category \( J \) is a disjoint union of connected categories (called the connected components of \( J \)).

(c) Conclude that all limits can be obtained from products and limits of connected categories.

8. (a) Show that the category \( J \) is connected; prove for any \( j \in C \) that \( \lim F \) is equal to \( \text{Colim} \).

(b) Describe the unit for the right adjoint to \( A : C \to C' \).

(c) Smythe. Show that the functor \( O : \text{Cat} \to \text{Set} \) assigning to each category \( C \) the set of objects of \( C \) has a left adjoint \( D \) which assigns to each set \( X \), the discrete category on \( X \), and that \( D \) in turn has a left adjoint assigning to each category \( C \) the set of its connected components. Also show that \( O \) has a right adjoint which assigns to each set \( X \) a category with objects \( X \) and exactly one arrow in every hom-set.

9. If a category \( C \) has both cokernels and equalizers, show that the functor \( \text{Ker} : \text{C} \to \text{C} \) with \( \text{Ker} \) assigns to each arrow \( C \) of its cokernel pair has as right adjoint the functor which assigns to each parallel pair of arrows its equalizing arrow.

10. Show that if \( C \) has discrete coproducts and \( a \in C \), then the projection \( Q : (a \sqcup C) \to C \) of the comma category \( (Q(a, c)) \) has a left adjoint, with \( c \mapsto (a \to \sqcup C) \).

11. If \( X \) is a set and \( C \) a category with powers and copowers, prove that the copower \( c \mapsto X \cdot c \) is left adjoint to the power \( c \mapsto c^X \).

3. Reflective Subcategories

For many of the forgetful functors \( U : A \to X \) listed in §2, the counit \( \varepsilon : F \to \text{Id}_A \) of the adjunction assigns to each \( a \in A \) the epimorphism \( \varepsilon_a : F(Ua) \to a \) which gives the standard representation of \( a \) as a quotient of a free object. This is a general fact: Whenever a right adjoint \( G \) is faithful, every \( \varepsilon_a \) of the adjunction is epi.

Theorem 1. For an adjunction \( (F, G, \varepsilon, \eta) : X \to A \): (i) \( G \) is faithful if and only if every component \( \varepsilon_a \) of the counit \( \varepsilon \) is epi. (ii) \( G \) is full if and only if every \( \varepsilon_a \) is split monic. Hence \( G \) is full and faithful if and only if each \( \varepsilon_a \) is an isomorphism \( F \_a \cong a \).

The proof depends on a lemma.
Lemma. Let \( f^* : A(1, -) \rightarrow A(b, -) \) be the natural transformation induced by an arrow \( f : b \rightarrow a \) of \( A \). Then \( f^* \) is monic if and only if \( f \) is epi, while \( f^* \) is epi if and only if \( f \) is a split monic (i.e., if and only if \( f \) has a left inverse).

Note that \( f^* \Rightarrow f \) is the bijection \( \text{Nat}(A(a, -), A(b, -)) \cong A(b, a) \) given by the Yoneda lemma.

Observe also, that for functors \( S, T : C \rightarrow B \), a natural transformation \( \tau : S \Rightarrow T \) is epi (respectively, monic) in \( B \) if and only if every component \( \tau_x : S_x \rightarrow T_x \) is epi (respectively, monic) in \( B \) for \( B = \text{Set} \); this follows by Exercise III.4.4, computing the product pointwise as in Exercise III.5.5.

Proof. For \( h \in A(c, a) \), \( f^* h = hf \). Hence the first result is just the definition of an epi \( f \). If \( f^* \) is epi, there is an \( h_0 : a \rightarrow b \) with \( f^* h_0 = h_0 f = 1 : b \rightarrow b \), so \( f \) has a left inverse. The converse is immediate.

Now we prove the theorem. Apply the Yoneda Lemma to the natural transformation (arrow function of \( G \) followed by the adjunction)

\[
A(a, c) \overset{\text{can}}{\longrightarrow} X(Ga, Ga) \overset{\eta^{-1}}{\longrightarrow} A(FGa, c).
\]

It is determined (set \( c = a \)) by the image of \( 1 : a \rightarrow a \), which is exactly the definition of the counit \( \epsilon_G : FGa \rightarrow a \). But \( \eta^{-1} \) is an isomorphism, hence this natural transformation is monic or epi, respectively, when every \( G, \), is injective or surjective, respectively; that is, when \( G \) is faithful or full, respectively. The result now follows by the lemma.

A subcategory \( A \subseteq B \) is called reflective in \( B \) when the inclusion functor \( K : A \rightarrow B \) has a left adjoint \( F : B \rightarrow A \). This functor \( F \) may be called a reflector and the adjunction \( \langle F, \eta, \epsilon \rangle : B \rightarrow A \) a reflection of \( B \) in its subcategory \( A \). Since the inclusion functor \( K \) is always faithful, the counit \( \epsilon \) of a reflection is always epi. A reflection can be described in terms of the composite functor \( R =KF : B \rightarrow B \); indeed, \( A \subseteq B \) is reflective in \( B \) if and only if there is a functor \( R : B \rightarrow B \) with values in the subcategory \( A \) and a bijection of sets

\[
A(Rb, a) \cong B(b, a).
\]

natural in \( b \in B \) and \( a \in A \). A reflection may be described in terms of universal arrows: \( A \subseteq B \) is reflective if and only if to each \( b \in B \) there is an object \( Rb \) of the subcategory \( A \) and an arrow \( \eta_b : Rb \rightarrow b \) such that every arrow \( g : b \rightarrow a \) has the form \( g = f \eta_b \) for a unique arrow \( f : Rb \rightarrow a \) of \( A \). As usual, \( R \) is then (the object function of) a functor \( B \rightarrow B \) with values in \( A \).

If a subcategory \( A \subseteq B \) is reflective in \( B \), then by Theorem 1 each object \( a \in A \) is isomorphic to \( FKa \), and hence \( R \cong F \) for all \( a \).

Dually, \( A \subseteq B \) is coreflective in \( B \) when the inclusion functor \( A \rightarrow B \) has a right adjoint. (Warning: Mitchell [1965] has interchanged the meanings of "reflection" and "coreflection".)

Exercises

1. Show that the table of dual statements (§11.1) extends as follows:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Dual statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S, T : C \rightarrow B ) are functors</td>
<td>( S, T : C \rightarrow B ) are functors</td>
</tr>
<tr>
<td>( T ) is full</td>
<td>( T ) is full</td>
</tr>
<tr>
<td>( T ) is faithful</td>
<td>( T ) is faithful</td>
</tr>
<tr>
<td>( \eta : S \Rightarrow T ) is a natural transformation</td>
<td>( \eta : T \Rightarrow S ) is a natural transformation</td>
</tr>
<tr>
<td>( \langle F, G, \varphi \rangle : X \rightarrow A ) is an adjunction</td>
<td>( \langle G, F, \varphi^{-1} \rangle : A \rightarrow X ) is an adjunction</td>
</tr>
<tr>
<td>( \eta ) is the unit of ( \langle F, G, \varphi \rangle )</td>
<td>( \eta ) is the counit of ( \langle G, F, \varphi^{-1} \rangle )</td>
</tr>
</tbody>
</table>

2. Show that the torsion-free abelian groups form a full reflective subcategory of \( Ab \).

3. If \( \langle G, F, \varphi \rangle : X \rightarrow A \) is an adjunction with \( G \) full and every unit \( \eta \) a monic, then every \( \eta \) is also epi.

4. Show the following subcategories to be reflective:

(a) The full subcategory of all partial orders in the category \( \text{Preord} \) of all preorders, with arrows all monotone functions.

(b) The full subcategory of \( T_0 \)-spaces in \( \text{Top} \).

5. Given an adjunction \( \langle F, G, \varphi \rangle : X \rightarrow A \), prove that \( G \) is faithful if and only if \( \varphi^{-1} \) carries epis to epis.

6. Given an adjunction \( \langle F, G, \eta, \epsilon \rangle \) with either \( F \) or \( G \) full, prove that \( G \eta : Grp \rightarrow Grp \) is invertible with inverse \( \eta G : Grp \rightarrow Grp \).

7. If \( A \) is a full and reflective subcategory of \( B \), prove that every functor \( S : J \rightarrow A \) with a limit in \( B \) has a limit in \( A \).

4. Equivalence of Categories

A functor \( S : A \rightarrow C \) is an isomorphism of categories when there is a functor \( T : C \rightarrow A \) (backwards) such that \( ST = 1 : C \rightarrow C \) and \( TS = 1 : A \rightarrow A \). In this case, the identity natural transformations
Equivalence of Categories

η : 1 → ST and ε : TS → 1 make ⟨T, S; η, ε⟩ : C → A an adjunction. In other words, a two-sided inverse T of a functor S is a left-adjoint of S — and for that matter, T is also a right-adjoint of S.

There is a more general (and more useful) notion:

A functor S : A → C is an equivalence of categories (and the categories A and C are equivalent) when there is a functor T : C → A (backwards) and natural isomorphisms ST ≅ 1 : C → C and TS ≅ 1 : A → A. In this case T : C → A is also an equivalence of categories. We shall soon see that T is then both a left adjoint and a right adjoint of S.

Here is an example. In any category C a skeleton of C is any full subcategory A such that each object of C is isomorphic (in C) to exactly one object of A. Then A is equivalent to C and the inclusion K : A → C is an equivalence of categories. For, select to each c ∈ C an isomorphism θc : c ≅ Tc with Tc an object of A. Then we can make TA a functor T : C → A in exactly one way so that θ will become a natural isomorphism θ : 1 ≅ KT.

Moreover TK ≅ 1, so K is indeed an equivalence: A category is equivalent to (any one of) its skeletons. For example, the category of all finite sets has as a skeleton the full subcategory with all finite ordinal numbers 0, 1, 2, ..., n, .... (Here 0 is the empty set and each n = {0, 1, ..., n − 1}.)

An adjoint equivalence of categories is an adjunction ⟨T, S; η, ε⟩ : C → A in which both the unit η : 1 → ST and the counit ε : TS → 1 are natural isomorphisms: I ≅ ST, TS ≅ 1. Then η −1 and ε −1 are also natural isomorphisms, and the triangular identities εT • Tη = 1, Sε • ηS = 1 can be written as Tη −1 • ε −1 T = 1, η −1 • S • ε −1 = 1, respectively. These identities then state that ⟨S, T, ε −1, η −1⟩ : A → C is an adjunction with ε −1 : I → TS as unit and η −1 : ST → I as counit. Thus in an adjoint equivalence ⟨T, S, ε, η⟩ the functor T : C → A is the left adjoint of S : A → C with unit η and at the same time T is the right adjoint of S, with unit ε −1.

We can now state the main facts about equivalence.

Theorem 1. The following properties of a functor S : A → C are logically equivalent:

i) S is an equivalence of categories,

ii) S is part of an adjoint equivalence ⟨T, S; η, ε⟩ : C → A,

iii) S is full and faithful, and each object c ∈ C is isomorphic to Sa for some object a ∈ A.

Proof. Trivially, (ii) implies (i). To prove that (i) implies (iii), note that ST ≅ 1 means that εT : TS → 1 for each a ∈ A. The natural isomorphism θ : TS ≅ 1 gives for each f : a → a' the com-
Duality theorems in functional analysis are often instances of equivalences. For example, let $\mathbf{CAb}$ be the category of compact topological abelian groups, and let $P$ assign to each such group $G$ its character group $P(G)$, consisting of all continuous homomorphisms $G \to \mathbb{R}/\mathbb{Z}$. The Pontryagin duality theorem asserts that $P : \mathbf{CAb} \to \mathbf{Ab}^{op}$ is an equivalence of categories. Similarly, the Gelfand-Naimark theorem states that the functor $C$ which assigns to each compact Hausdorff space $X$ its abelian $C^*$-algebra of continuous complex-valued functions is an equivalence of categories (see Negrepontis [1971]).

Exercises

1. Prove: (a) Any two skeletons of a category $C$ are isomorphic.
   (b) If $A_0$ is a skeleton of $A$ and $C_0$ a skeleton of $C$, then $A$ and $C$ are equivalent if and only if $A_0$ and $C_0$ are isomorphic.
   (c) Prove: the composite of two equivalences $D \to C, C \to A$ is an equivalence.
   (d) State and prove the corresponding fact for adjoint equivalences.
2. (a) If $S : A \to C$ is full, faithful, and surjective on objects (each $c \in C$ is $c = Sa$ for some $a \in A$), prove that there is an adjoint equivalence $\langle T, S, 1, e \rangle : C \to A$ with unit the identity (and hence that $T$ is a left-adjoint-right-inverse of $S$).
3. Given a functor $G : A \to X$, prove the three following conditions logically equivalent:
   (a) $G$ has a left-adjoint-left-inverse,
   (b) $G$ has a left adjoint, and is full, faithful, and injective on objects.
   (c) There is a full reflective subcategory $Y$ of $X$ and an isomorphism $H : A \cong Y$ such that $G = KH$, where $K : Y \to X$ is the insertion.
4. If $J$ is a connected category and $\Delta : C \to C'$ has a left adjoint (cofinal), show that this left adjoint is a left-adjoint-left-inverse.

5. Adjoint functors for Preorders

Recall that a preorder $P$ is a set $P = \{p, p', \ldots\}$ equipped with a reflexive and transitive binary relation $p \leq p'$, and that preorders may be regarded as categories so that order-preserving functions become functors. An order-reversing function $L$ on $P$ to $Q$ is then a functor $L : P \to Q^{op}$.

**Theorem 1.** (Galois connections are adjoint pairs). Let $P, Q$ be two preorders and $L : P \to Q^{op}$, $R : Q^{op} \to P$ two order-preserving functions. Then $L$ (regarded as a functor) is a left adjoint to $R$ if and only if, for all $p \in P$ and $q \in Q$,

$$Lp \geq q \text{ in } Q \text{ if and only if } p \leq Rq \text{ in } P. \quad (1)$$

When this is the case, there is exactly one adjunction $\varphi$ making $L$ the left adjoint of $R$. For all $p$ and $q, p \leq RLp$ and $LRq \geq q$; hence also

$$Lp \geq LRLp \geq Lp, \quad Rq \leq RLRq \leq Rq. \quad (2)$$

**Proof.** Recall that $P$ becomes a category in which there is (exactly) one arrow $p \to p'$ whenever $p \leq p'$. Thus the condition (1) states precisely that there is a bijection $\text{hom}_P(Lp, q) \cong \text{hom}_P(p, Rq)$; since each hom-set has at most one element, this bijection is automatically natural. The unit of the adjunction is the inequality $p \leq RLp$ for all $p$, while the counit is $Lp \geq q$ for all $q$. The two Eqs. (2) are the triangular identities connecting unit and counit. In the convenient case when both $P$ and $Q$ are posets (i.e., when both the relations $\leq$ are antisymmetric) these conditions become $L = LRL$, and $R = RLR$ (each of these three passages reduce to one!).

A pair of order-preserving functions $L$ and $R$ which satisfy (1) is called a Galois connection from $P$ to $Q$. Here is the fundamental example, for a group $G$ acting on a set $U$, by $\langle x, y \rangle \mapsto y \cdot x$ for $x \in G$, $y \in U$. Take $P = \mathcal{P}(U)$, the set of all subsets $X \subseteq U$, ordered by inclusion, while $Q = \mathcal{P}(G)$ is the set of subgroups $S \subseteq G$ also ordered by inclusion ($S \subseteq S'$ if and only if $S' \subseteq S$). Let $LX = \{ x \mid x \in X \text{ implies } x \cdot x = x \}$, $RS = \{ x \mid x \in S \text{ implies } x \cdot x = x \}$; in other words, $LX$ is the subgroup of $G$ which fixes all points $x \in X$ and $RS$ is the set of fixed points of the automorphisms of $S$. Then $LX \geq S$ if and only if $x \cdot x = x$ for all $x \in S$ and $x \in X$, which in turn holds if and only if $X \subseteq RS$ in $P$. Therefore, $L$ and $R$ form an adjoint pair (a Galois connection). The original instance is that with $G$ a group of automorphisms of a field $U$, as in the classical Galois theory.

If $U$ and $V$ are sets, the set $\mathcal{P}(U)$ of all subsets of $U$ is a preorder under inclusion. For each function $f : U \to V$ the direct image functor $f_*$ defined by $f_*(X) = \{ f(x) \mid x \in X \}$ is an order-preserving function and hence a functor $f_* : \mathcal{P}(U) \to \mathcal{P}(V)$. The inverse image function $f^*(Y) = \{ x \mid f(x) \in y \}$ for some $y \in Y$ defines a functor $f^* : \mathcal{P}(V) \to \mathcal{P}(U)$ in the opposite direction. Since $f_ Xu \subseteq Y$ if and only if $U \subseteq f^* Y$, the direct image functor $f_*$ is left adjoint to the inverse image function $f^*$.

Certain adjoints for Boolean algebras are closely related to the basic connectives in logic. We again regard $\mathcal{P}(U)$ as a preorder, and hence as a category. The diagonal functor $\Delta : \mathcal{P}(U) \to \mathcal{P}(U) \times \mathcal{P}(U)$ has (as we have already noted) a right adjoint $\land$, sending subsets $X$, $Y$ to their intersection $X \cap Y$, and a left adjoint $\lor$, with $\langle X, Y \rangle \mapsto X \cup Y$, the union. If $X$ is a fixed subset of $U$, then intersection with $X$ is a functor $X \cap - : \mathcal{P}(U) \to \mathcal{P}(U)$. Since $X \cap Y \subseteq Z$ if and only if $Y \subseteq X' \cup Z$, where $X'$ is the complement of $X$ in $U$, the right adjoint of $X \cap -$ is $X' \cup -$.

Thus the construction of suitable adjoints yields the Boolean operations $\land, \lor$, and $\neg$ corresponding to "and", "or", and "not". Now consider the first projection $P : U \times V \to U$ from the product of two sets $U$ and $V$. Each subset $S \subseteq U \times V$ defines two corresponding subsets of $U$ by

$$P_+ S = \{ x \mid \exists y \quad \text{and} \quad \langle x, y \rangle \in S \};$$

$$P_* S = \{ x \mid \forall y, y \in V \quad \text{implies} \quad \langle x, y \rangle \in S \};$$
they arise from \( \langle x, y \rangle \in S \) by applying the existential quantifier \( \exists y \), 
"there exists a \( y \)" and the universal quantifier \( \forall y \), 
"for all \( y \)", respectively. 
Also, \( P^* S \) is the direct image of \( S \) under the projection \( P \). Now for all 
subsets \( X \subseteq U \) one has 
\[
S \subseteq P^* X \Rightarrow P^* S \subseteq X ; \quad P^* X \subseteq S \Rightarrow X \subseteq P^* S .
\]
where \( \Rightarrow \) means "if and only if". These state that \( P^* \), the inverse image operation, 
have both a left adjoint \( P_* \) and a right adjoint \( P^* \). In this sense, 
both quantifiers \( \exists \) and \( \forall \) can be interpreted as adjoints.

There is also a geometric interpretation: \( P^* X \) is the cylinder 
\( X \times V \subset U \times V \) over the base \( X \subset U \), 
\( P_* S \) is the projection of \( S \subset U \times V \) on the base \( U \), and \( P^* S \) is the largest subset \( X \) of \( U \) such that the cylinder on 
\( X \) is wholly contained in \( S \).

**Exercises**

1. Let \( H \) be a space with an inner product (e.g., Hilbert space). If \( P = Q \) is the set 
of all subsets of \( H \), ordered by inclusion, show that \( LS = RS = \) the orthogonal 
complement of \( S \) gives a Galois connection.

2. In a Galois connection between posets, show that the subset \( \{ p | p \in RLp \} \) of 
\( P \) equals \( \{ q | q \in LRq \} \) for some \( q \) and give a bijection from this set to the subset 
\( \{ q | q \in LRq \} \) of \( Q \). What are these sets in the case of a group of automorphisms 
of a field? Does this generalize to an arbitrary adjunction?

3. For \( C \) a category with pullbacks, each arrow \( f: a \rightarrow a' \) defines a functor 
\( (C f) = f_s^* : (C a) \rightarrow (C a') \) which carries each object \( x \mapsto a \) of \( (C a) \) to the 
composite \( x \mapsto a \mapsto a' \). Show that \( f_s \) has a right adjoint \( f^* \) with \( f^*(x \mapsto a) = y \mapsto a \), 
where \( y \) is the vertex of the pullback of \( a \rightarrow a' \rightarrow x \).

### 6. Cartesian Closed Categories

Much of the force of category theory will be seen to reside in using 
categories with specified additional structures. One basic example will 
be the closed categories (§ VII. 7); at present we can define readily one 
useful special case: "cartesian closed".

To assert that a category \( C \) has all finite products and coproducts is to 
assert that the functors \( C \rightarrow \mathbf{1} \) and \( \Delta : C \rightarrow C \times C \) have both left and right 
adjoints. Indeed, the left adjoints give initial object and coproduct, 
respectively, while the right adjoints give terminal object and product, 
respectively.

Using just adjoints we will now define "cartesian closed category". 
A category \( C \) with all finite products specifically given is called cartesian 
closed when each of the following functors 
\[
C \rightarrow \mathbf{1} , \quad C \rightarrow C \times C , \quad C \rightarrow C , \quad c \rightarrow 0 , \quad c \rightarrow \langle c , c \rangle , \quad a \rightarrow a \times b ,
\]
has a specified right adjoint (with a specified adjunction). These adjoints 
are written as follows
\[
t \leftarrow 0 , \quad a \times b \rightarrow \langle a , b \rangle , \quad c^b \rightarrow c .
\]

Thus to specify the first is to specify a terminal object \( t \) in \( C \), and specifying 
the second is specifying for each pair of objects \( a , b \in C \) a product object 
\( a \times b \) together with its projections \( a \leftarrow a \times b \rightarrow b \). These projections 
determine the adjunction (they constitute the counit of the adjunction); 
as already noted, \( x \) is then a bifunctor. The third required adjoint 
specifies for each functor \( - \times b : C \rightarrow C \) a right adjoint, with the corresponding 
bijection
\[
\text{hom}(a \times b , c) \cong \text{hom}(a , c^b)
\]
natural in \( a \) and in \( c \). By the parameter theorem (to be proved in the next 
section), \( \langle b , c \rangle \mapsto c^b \) is then the (object function of a bifunctor \( C^a \times C \rightarrow C \). 
Specifying the adjunction amounts to specifying for each \( c \) and \( b \) an arrow \( e \)
\[
e : c^b \rightarrow b 
\]
which is natural in \( c \) and universal from \( - \times b \) to \( c \). We call \( e = e_{c^b} \) the 
evaluation map. It amounts to the ordinary evaluation \( f(x) \mapsto f(x) \) of a 
function \( f \) at an argument \( x \) in both of the following cases:

- **Set** is a cartesian closed category, with \( c^b = \text{hom}(b , c) \).
- **Cat** is cartesian closed, with \( C^b \) the functor category.

A closely related example is the functor 
\[
- \otimes_k B : K-\text{Mod} \rightarrow K-\text{Mod}
\]
which has a right adjoint \( \text{hom}_k(B , -) \); the adjunction is determined by 
a counit \( \text{hom}_k(B , A) \otimes_k B \rightarrow A \) given by evaluation.

**Exercises**

1. (a) If \( U \) is any set, show that the preorder \( \mathcal{P}(U) \) of all subsets of \( U \) is a cartesian 
closed category.

(b) Show that any Boolean algebra, regarded as a preorder, is cartesian closed.

2. In some elementary theory \( T \), consider the set \( S = \{ p , q , \ldots \} \) of sentences of \( T \) 
as a preorder, with \( p \subseteq q \) meaning "\( p \) entails \( q \" (i.e., \( q \) is a consequence of \( p \) 
on the basis of the axioms of \( T \)). Prove that \( S \) is a cartesian closed category, 
with product given by conjunction and exponential \( q^p \) given by "\( p \) implies \( q \".

3. In any cartesian closed category, prove \( e^c \cong c \) and \( e^{p^c} \cong (p^c)^c \).

4. In any cartesian closed category obtain a natural transformation \( c^b \rightarrow c \), 
which agrees in \( \text{Set} \) with composition of functions. Prove it (like composition) 
associative.

5. Show that a cartesian closed need not imply \( A^c \) cartesian closed.


7. Transformations of Adjoints

We next study maps comparing different adjunctions. Given two adjunctions

\[ \left\langle F, G, \phi, \eta, \epsilon \right\rangle : X \to A, \quad \left\langle F', G', \phi', \eta', \epsilon' \right\rangle : X' \to A' \]  

we define a map of adjunctions (from the first to the second adjunction) to be a pair of functors \( K : A \to A' \) and \( L : X \to X' \) such that both squares

\[ \begin{array}{c}
A \xrightarrow{G} X \xrightarrow{F} A \\
\downarrow \quad \downarrow \\
A' \xrightarrow{G'} X' \xrightarrow{F'} A'
\end{array} \]  

commute, and such that the diagram of hom-sets and adjunctions

\[ \begin{array}{c}
A(Fx, a) \xrightarrow{\eta} X(x, Ga) \\
\downarrow \quad \downarrow \\
A'(KFx, Ka) \xrightarrow{\eta'} X'(Lx, LGa)
\end{array} \]  

commutes for all objects \( x \in X \) and \( a \in A \). Here \( K_{Fx,a} \) is the map \( f \mapsto Kf \) given by the functor \( K \).

**Proposition 1.** Given adjunctions (1) and functors \( K \) and \( L \) satisfying (2), the condition (3) on hom-sets is equivalent to \( L \eta = \eta' L \) and also to \( \epsilon' K = K \epsilon \).

**Proof.** Given (3) commutative, set \( a = Fx \) and chase the identity arrow \( 1 : Fx \to Fx \) to get the units \( \eta, \eta' \) and the equality

\[ \langle L \eta : L \to LGF \rangle = \langle \eta' L : L \to G' F'L \rangle, \]

where \( LGF = G' F' L \) by (2). Conversely, given the equality \( L \eta = \eta' L \) of natural transformations, the definition of the adjunctions \( \phi \) and \( \phi' \) by their units gives (3). The case of the counits is dual to this one.

Next, given two adjunctions

\[ \langle F, G, \phi, \eta, \epsilon \rangle, \quad \langle F', G', \phi', \eta', \epsilon' \rangle : X \to A \]  

between the same categories, two natural transformations

\( \sigma : F \to F' \), \( \tau : G' \to G \)

are said to be conjugate (for the given adjunctions) when the diagram

\[ \begin{array}{c}
A(Fx, a) \xrightarrow{\sigma} X(x, Ga) \\
\downarrow (\sigma \eta) \quad \downarrow \epsilon \\
A(Fx, a) \xrightarrow{\eta} X(x, Ga)
\end{array} \]

commutes for every pair of objects \( x \in X \), \( a \in A \).

**Theorem 2.** Given the two adjunctions (4), the natural transformations \( \sigma \) and \( \tau \) are conjugate if and only if any one of the four following diagrams (of natural transformations) commutes

\[ \begin{array}{c}
G \xrightarrow{\epsilon} G' \\
F \xrightarrow{\eta} F'
\end{array} \]

\[ \begin{array}{c}
FG \xrightarrow{\epsilon' F'} FG' \xrightarrow{FG} FG' \\
I_A \xrightarrow{\eta'} I_A
\end{array} \]

\[ \begin{array}{c}
F \xrightarrow{\eta} F' \\
G \xrightarrow{\epsilon} G'
\end{array} \]

Also, given the adjunctions (4) and the natural transformation \( \sigma : F \to F' \), there is a unique \( \tau : G' \to G \) such that the pair \( (\sigma, \tau) \) is conjugate. Dually, given (4) and \( \tau \), there is a unique \( \sigma \) with \( (\sigma, \tau) \) conjugate.

**Proof.** First, (5) implies (6) and (5) implies (7). For, put \( x = G'a \) in (5), start with the identity arrow \( 1 : G'a \to G'a \) in the upper right and use the description of \( \phi \) and \( \phi' \) by unit and counit to chase this element 1 around the diagram as follows

\[ \begin{array}{c}
\epsilon' \circ \sigma G \xrightarrow{1} 1 \\
\downarrow \quad \downarrow \\
\sigma G \circ \epsilon \xrightarrow{1} 1
\end{array} \]

where \( \sigma G \) is the map \( f \mapsto \sigma f \) given by the natural transformation \( \sigma \).

The result (lower right) is the first equality of (6). A slightly different chase yields

\[ \begin{array}{c}
\epsilon' \circ \sigma G \xrightarrow{1} 1 \\
\downarrow \quad \downarrow \\
\sigma \circ \epsilon' \xrightarrow{1} 1
\end{array} \]

The resulting equality is the first diagram of (7). The second halves of (6) and (7) are duals.
Next, suppose \( \sigma \) but not \( \tau \) given. Then the Yoneda Lemma applied to the composite transformation \( \varphi (\sigma) \cdot \varphi^{-1} \) (three legs of (5)) shows that there is a unique family of arrows \( \tau_a \) for which (5) commutes, and this family is a natural transformation. Since each \( \epsilon_a : F G a \to a \) is universal from \( F \to a \), there is also a unique family of arrows \( \tau_a \) for which the first of (7) commutes. Since (5) implies (7), \( \tau_a = \tau_a \). In other words, if \( \tau = \tau' \) makes the first square of (7) commute, it also makes (5) commute. Therefore the first square of (7) implies (5). Given \( \sigma \), there is immediately a unique natural transformation \( \tau : G \to G' \) for which the first of (6) commutes; since (5) implies (6), \( \tau_a = \tau_a \), and hence the solutions \( \tau_a \) of (5) are necessarily natural; moreover (6) implies (5).

The reader may also show that (6) implies (5) or (7) by constructing suitable diagrams of natural transformations.

We now regard a conjugate pair \( \langle \sigma, \tau \rangle \) of natural transformations as a transformation (or morphism) from the first to the second adjunction. The "vertical" composite of two such

\[
\langle F, G, \eta, \varepsilon \rangle \circ \langle F', G', \eta', \varepsilon' \rangle = \langle F', G', \eta', \varepsilon' \rangle \circ \langle F, G, \eta, \varepsilon \rangle
\]

is evidently (say by condition (5)) a transformation \( \langle \sigma', \tau' \rangle \circ \langle \sigma, \tau \rangle \) from the first to the third adjunction. For the two given categories \( X \) and \( A \) we thus have a new category \( A^{adj} X \), the category of adjunctions from \( X \) to \( A \); its objects are the adjunctions \( \langle F, G, \eta, \varepsilon \rangle \); its arrows are the transformations (conjugate pairs) \( \langle \sigma, \tau \rangle \), with the composition just noted. Also there are two evident "forgetful" functors to the ordinary functor categories, as follows

\[ A^X \leftarrow A^{adj} X, \quad A^{adj} X \rightarrow X^A, \]

\[ F \leftarrow \langle F, G, \eta, \varepsilon \rangle \rightarrow G \]

\[ \sigma \]

\[ F \leftarrow \langle F, G', \eta', \varepsilon' \rangle \rightarrow G' \]

A typical example for \( Set \) is the bijection

\[
\text{hom}(S \times T, R) \cong \text{hom}(S, \text{hom}(T, R))
\]

discussed in §1 as an example of an adjunction (for each fixed set \( T \)). If \( t : T \to T' \) is a function between two such sets, then \( - \times t \) is a natural transformation of functors \( - \times T \to - \times T' \). Its conjugate is the natural transformation \( \text{hom}(t, -) : \text{hom}(T', -) \to \text{hom}(T, -) \); this is, as it should be, in the reverse direction, corresponding to the fact that \( S \times T \) is covariant and \( \text{hom}(T, R) \) contravariant in the argument \( T \). We may call (9) an adjunction with a "parameter" \( T \in \text{Set} \). For a commutative ring

K then the adjunction \( \text{Mod}_R(A \otimes_R B, C) \cong \text{Mod}_R(A, \text{Hom}_R(B, C)) \) has a parameter \( B \in \text{Mod}_R \). Here is general statement:

**Theorem 3 (Adjunctions with a parameter).** Given a bifunctor \( F : X \times P \to A \), assume for each object \( p \in P \) that \( F(-, p) : X \to A \) has a right adjoint \( G(p, -) : A \to X \), via an adjunction

\[
\text{hom}(F(x, p), a) \cong \text{hom}(x, G(p, a)),
\]

natural in \( x \) and \( a \). There is then a unique way to assign to each arrow \( h : p \to p' \) of \( P \) and each object \( a \in A \) an arrow \( G(h, a) : G(p, a) \to G(p', a) \) of \( X \) so that \( G \) becomes a bifunctor \( P^{op} \times A \to X \) for which the bijection of the adjunction (10) is natural in all three variables \( x, p, \) and \( a \). This assignment of arrows \( G(h, a) \) to \( \langle h, a \rangle \) may also be described as the unique way to make \( G(h, -) \) a natural transformation conjugate to \( F(-, h) \).

*Proof.* The condition that the adjunction (10) be natural in \( p \in P \) is the commutativity of the square

\[
\text{hom}(F(x, p), a) \cong \text{hom}(x, G(p, a))
\]

\[
\text{hom}(F(x, p'), a) \cong \text{hom}(x, G(p', a)).
\]

This commutativity (for all \( a \)) states precisely that \( G(h, -) : G(p, -) \to G(p', -) \) must be chosen as the conjugate to \( F(-, h) : F(-, p) \to F(-, p') \). By the previous theorem, there exists a unique choice of \( G(h, -) \) to realize this — and the condition of conjugacy may be expressed in any of the five equivalent ways stated there. For a second arrow \( h' : p' \to p'' \), the uniqueness of the choice of conjugates shows for \( h' h \) that \( G(h' h, -) = G(h', -) G(h, -) \), so that \( G(-, a) \) is a functor and \( G \) a bifunctor, as required.

Dually, given a bifunctor \( G : P^{op} \times A \to X \) where each \( G(p, -) \) has a right adjoint \( F(-, p) \), there is a unique way to make \( F \) a bifunctor \( X \times P \to A \).

**Exercises**

1. Interpret the definition \( C(X \cdot a, c) \cong \text{Set}(X, C(c, a)) \) of copowers \( X \cdot a \) in \( C \) as an adjunction with parameter \( a \).

2. Let \( \eta : X \to G(p, F(x, p)) \) be the unit of an adjunction with parameter. It is natural in \( X \), but what property of \( \eta \) corresponds to the naturality of the adjunction (10) in \( p \)?

3. In the functor category \( A^X \) let \( S \) be that full subcategory with objects those functors \( F : X \to A \) which have a right adjoint \( RF : A \to X \). Make \( R \) a functor \( S^{op} \to A^X \) by choosing one \( RF \) for each \( F \), with \( R \sigma \) the conjugate of \( \sigma \).
Composition of Adjoint Functors

4. (Kelly) An adjoint square is an array of categories, functors, adjunctions, and natural transformations

\[
\begin{array}{ccc}
X & \xrightarrow{(f, g, \varepsilon)} & A \\
\downarrow H & & \downarrow \kappa \\
 X' & \xrightarrow{(f', g', \varepsilon')} & A'.
\end{array}
\]

such that the following diagram of hom-sets always commutes

\[
\begin{array}{ccc}
A(Fx, a) & \xrightarrow{\kappa} & A(KFx, Ka) \\
\downarrow \phi & & \downarrow \phi' \\
X(x, Ga) & \xrightarrow{H} & X'(Hx, HGa).
\end{array}
\]

Express this last condition variously in terms of unit and counit of the adjunctions and prove that each of \(\alpha, \tau\) determines the other. (The case \(H = K\) = identity functor is that treated in the text above.)

5. (Palmquist) Given \(H, K\), and the two adjunctions as in Exercise 4. establish a bijection between natural transformations \(\alpha : FH \rightarrow K\) and natural transformations \(\beta : H \rightarrow G\).

8. Composition of Adjoint Functors

Two successive adjunctions compose to give a single adjunction, in the following sense:

**Theorem 1.** Given two adjunctions

\[\langle F, G, \eta, \varepsilon \rangle : X \rightarrow A, \quad \langle F', G', \eta', \varepsilon' \rangle : A \rightarrow D\]

the composite functors yield an adjunction

\[\langle F'F, G'G, G'F \cdot \eta, \varepsilon : F \circ G \rangle : X \rightarrow D.\]

**Proof.** With hom-sets, the two given adjunctions yield a composite isomorphism, natural in \(x \in X\) and \(d \in D\):

\[D(FFx, d) \cong A(Fx, Gd) \cong X(x, Gd).\]

This makes the composite \(\overline{F}\) left adjoint to \(G\). Setting \(d = FFx\) and applying these two isomorphisms to the identity \(1 : FFx \rightarrow FFx\), we find that the unit of the composite adjunction is \(x \xrightarrow{\eta x} G x \xrightarrow{G \eta x} GFFx\), so is \(G\eta F \cdot \eta\), as asserted. By the dual argument, the counit is \(\varepsilon \cdot F \varepsilon G\), q.e.d. One can also calculate directly that these last formulas give natural transformations \(1 \rightarrow G\overline{F}F\) and \(F\overline{F}G \rightarrow 1\) which satisfy the triangular identities.

Using this composition, we may form a category \(\text{Adj}\) whose objects are all (small) categories \(X, A, D, \ldots\) and whose arrows are the adjunctions \(\langle F, G, \eta, \varepsilon \rangle : X \rightarrow A\), composed as above; the identity arrow for each category \(A\) is the identity adjunction \(A \rightarrow A\).

This category has additional structure. Each hom-set \(\text{Adj}(X, A)\) may be regarded as a category; to wit, the category \(\text{Adj}(X, A)\) of adjunctions from \(X\) to \(A\) as described in the last section. Its objects are these adjunctions and its arrows are the conjugate pairs \(\langle \alpha, \tau \rangle\), under "vertical" composition defined in (7.8).

**Theorem 2.** Given two conjugate pairs

\[\langle \alpha, \tau \rangle : \langle F, G, \eta, \varepsilon \rangle \rightarrow \langle F', G', \eta', \varepsilon' \rangle : X \rightarrow A,\]

\[\langle \tilde{\alpha}, \tilde{\tau} \rangle : \langle \overline{F}, \overline{G}, \overline{\eta}, \overline{\varepsilon} \rangle \rightarrow \langle \overline{F'}, \overline{G'}, \overline{\eta'}, \overline{\varepsilon'} \rangle : A \rightarrow D\]

the (horizontal) composite natural transformations \(\overline{\alpha} = \tau \tilde{\alpha}\) and \(\tilde{\tau} = \alpha \tilde{\alpha}\) yield a composite adjunction \(\overline{\alpha} : \overline{F} \rightarrow \overline{F'}\) for \(\overline{G} \rightarrow \overline{G'}\) of natural transformations for the composite adjunctions.

The proof may be visualized by the diagram of hom-sets

\[D(FFx, d) \cong A(Fx, Gd) \cong X(x, Gd)\]

\[\begin{array}{ccc}
D(FFx, d) & \xrightarrow{\alpha} & A(Fx, Gd) \\
\downarrow \phi & & \downarrow \phi' \\
D(FFx, d) & \xrightarrow{\tau} & X(x, Gd).
\end{array}\]

Moreover, this operation of (horizontal) composition is a bifunctor

\[\text{Adj}(A, D) \times \text{Adj}(X, A) \rightarrow \text{Adj}(X, D).\]

This means that \(\text{Adj}\) is a "two-dimensional" category, as is \(\text{Cat}\) (see § II.5).

**Exercises**

1. Prove that horizontal composition is a bifunctor, as in (1), and that this implies an interchange law between horizontal and vertical composition of conjugate pairs.

2. Show that the adjunction with right adjoint the forgetful functor \(\text{Rng} \rightarrow \text{Set}\) can be obtained as a composite adjunction in two ways, \(\text{Rng} \rightarrow \text{Ab} \rightarrow \text{Set}\) and \(\text{Rng} \rightarrow \text{Mon} \rightarrow \text{Set}\).

3. Let \(R, S, T\) be rings.

(a) For a bimodule \(\varepsilon E\), show that \(- \otimes \varepsilon E : \text{Mod}_R \rightarrow \text{Mod}_T\) has a right adjoint \(\text{hom}_R(E, -)\).

(b) Show that this is an adjunction with parameter \(E \in R\text{-Mod-S}\).

(c) Describe the composite of this adjunction with a similar adjunction \(\text{Mod}_R \rightarrow \text{Mod}_T\).
Composition of Adjoints

Notes

The multiple examples, here and elsewhere, of adjoint functors tend to show that adjoints occur almost everywhere in many branches of Mathematics. It is the thesis of this book that a systematic use of all these adjunctions illuminates and clarifies these subjects. Nevertheless, the notion of an adjoint pair of functors was developed only very recently. The word "adjoint" seems to have arisen first (and long ago) to describe linear differential operators. About 1930 the concept was carried over to a Hilbert space $H$, where the adjoint $T^*$ of a given linear transformation $T$ on $H$ is defined by equality of the inner products

$$(T^*x, y) = (x, Ty)$$

for all vectors $x, y \in H$. Clearly, there is a formal analogy to the definition of adjoint functor.

Daniel Kan in [1958] was the first to recognize and study adjoint functors. He needed them for the study of simplicial objects, and he developed the basic properties such as units and counits, limits as adjoints, adjunctions with a parameter, and conjugate transformations, as well as an important existence theorem (the Kan extension — see Chapter X). Note that his discovery came ten years after the exact formulation of universal constructions. Initially, the idea of adjunctions took on slowly, and the relation to universal arrows was not clear. Freyd in his 1960 Princeton thesis (unpublished but widely circulated) and in his book [1964] and Lawvere [1963, 1964] emphasized the dominant position of adjunctions. One must pause to ask if there are other basic general notions still to be discovered.

One may also speculate as to why the discovery of adjoint functors was so delayed. Ideas about Hilbert space or universal constructions in general topology might have suggested adjoints, but they did not; perhaps the 1939–1945 war interrupted this development. During the next decade 1945–55 there were very few studies of categories, category theory was just a language, and possible workers may have been discouraged by the widespread pragmatic distrust of "general abstract nonsense" (category theory). Bourbaki just missed ([1948], Appendix III). His definition of universal construction was clumsy, because it avoided categorical language, but it amounted to studying a bifunctor $W : X \times A \to \text{Set}$ and asking for a universal element of $W(x, -)$ for each $x$. This amounts to asking for objects $Fx \in A$ and a natural isomorphism $W(x, a) \cong A(Fx, a)$; it includes the problem of finding a left adjoint $F$ to a functor $G : A \to X$, with $W(x, a) = \hom(x, Ga)$. It also includes the problem of finding a tensor product for two modules $A$ and $B$, with $W((A, B), C)$ taken to be the set of bilinear functions $A \times B \to C$. Moreover, the tensor product $A \otimes B$ is not in this way an example of a left adjoint (though it is an example of our universal arrows). In other words Bourbaki's idea of universal construction was devised to be so general as to include more—and in particular, to include the ideas of multilinear algebra which were important to French Mathematical traditions. In retrospect, this added significance was missed; Bourbaki's construction problem emphasized representable functors, and asked "Find $Fx$ so that $W(x, a) \cong A(Fx, a)$". This formulation lacks the symmetry of the adjunction problem, "Find $Fx$ so that $X(x, Ga) \cong A(Fx, a')" and so missed a basic discovery; this discovery was left to a younger man, perhaps one less beholden to tradition or to fashion. Put differently, good general theory does not search for the maximum generality, but for the right generality.

V. Limits

This chapter examines the construction and properties of limits, as well as the relation of limits to adjoints. This relation is then used in the basic existence theorems for adjoint functors, which give universals and adjoints in a wide variety of cases. The chapter closes with some indications of the uses of adjoint functors in topology.

1. Creation of Limits

A category $C$ is called small-complete (sometimes just complete) if all small diagrams in $C$ have limits in $C$; that is, if every functor $F : J \to C$ to a small category $J$ has a limit. We shall show that $\text{Set}$, $\text{Grp}$, $\text{Ab}$, and many other categories of algebras are small-complete.

The construction of limits in $\text{Set}$ may be illustrated by considering the limit of a functor $F : \omega \to \text{Set}$; here $\omega$, the linearly ordered set of all finite ordinals, is the free category generated by the graph

$$\{0 \to 1 \to 2 \to 3 \to \ldots \}.$$

The functor $F : \omega \to \text{Set}$ is just a list of sets $F_n$ and of functions $f_n$, as in the first row of the diagram below

$$\begin{array}{ccccccccc}
F_0 & \longrightarrow & F_1 & \longrightarrow & F_2 & \longrightarrow & \cdots & \longrightarrow & F_n & \longrightarrow & \cdots \\
\uparrow & & \uparrow & & \uparrow & & \cdots & & \uparrow & & \uparrow \\
\Pi_{F_0} & \longrightarrow & \Pi_{F_1} & \longrightarrow & \Pi_{F_2} & \longrightarrow & \cdots & \longrightarrow & \Pi_{F_n} & \longrightarrow & \cdots \\
\langle x_0, x_1, \ldots, x_n \in F_n \rangle & \longrightarrow & \langle x_0, x_1, \ldots, f_n x_{n+1} = x_{n+1} \in F_{n+1} \rangle.
\end{array}$$

Given $F$, form first the product set $\Pi_{F_0}$; it consists of all strings $x = \langle x_0, x_1, x_2, \ldots, \rangle$ of elements, with each $x_n \in F_n$, and it has projections $p_n : \Pi_{F_0} \to F_n$, but the triangles formed by these projections do not commute $(f_n p_{n+1} + p_n)$. A limit must be at least a vertex of a set of commuting triangles (a cone). So take the subset $L$ of those strings $x$ which "match" under $f$, in that $f_n x_{n+1} = x_{n+1}$ for all $n$. Then functions $\mu_n : L \to F_n$ are defined by $\mu_n x = x_n$, since the string $x$ matches, $f_n \mu_{n+1} = \mu_n$.
for all $n$, so $\mu : L \to F$ is a cone from the vertex $L \in \text{Set}$ to the base $F$. If $\tau : M \to F$ is any other cone from a set $M$ as vertex, each $m \in M$ determines a string $(\tau, m)$ which matches and hence a function $g : M \to L$, with $gm = (\tau, m)$, so with $\mu g = \tau$. Since $g$ is the unique such function this shows that $\mu$ is a universal cone to $F$, and so that $L$ is the limit set of $F$.

A string $x$ which "matches" is the same thing as a cone $x : \to F$ to $F$ from the one point set $\ast$. Hence the limit $L$ above can be described as the set $L = \text{Cone}(\ast, F)$ of all such cones. The same construction applies for any domain category (in place of $\omega^\mu$).

**Theorem 1 (Completeness of Set).** If the category $J$ is small, any functor $F : J \to \text{Set}$ has a limit which is the set $\text{Cone}(\ast, F)$ of all cones $\sigma : \to F$ from the one point set $\ast$ to $F$, while the limiting cone $\nu$, with

$$v_j : \text{Cone}(\ast, F) \to F_j, \quad \sigma \mapsto \sigma_j,$$

(2)

is that function sending each cone $\sigma$ to the element $\sigma_j \in F_j$.

For example, if $J$ is discrete, the set $\text{Cone}(\ast, F)$ is just the cartesian product $\Pi J F_j$.

**Proof.** Since $J$ is small, $\text{Cone}(\ast, F)$ is a small set, hence an object of $\text{Set}$. If $u_j \to k$ is any arrow of $J$, then $F_k \sigma_j = \sigma_k$ because $\sigma$ is a cone; hence $\nu$ as defined (2) is a cone to the base $F$. To prove it universal, consider any other cone $\tau : X \to F$ to $F$ from some set $X$. Then for each $x \in X$, $\tau x$ is a cone to $F$ from one point, so there is a unique function $h : X \to \text{Cone}(\ast, F)$ sending each $x$ to $\tau x$.

The crux of this proof is the (natural) bijection

$$\text{Cone}(X, F) \cong \text{Set}(X, \text{Cone}(\ast, F))$$

(3)

given by $\tau \mapsto h$, as above. Since a cone is just a natural transformation, this may be rewritten as an adjunction

$$\text{Nat}(\Delta X, F) \cong \text{Set}(X, \text{Cone}(\ast, F)).$$

By the very definition of limit, this proves that $\text{Lim} F \cong \text{Cone}(\ast, F)$.

Limits in $\text{Grp}$ and other categories may be constructed from the set of all cones in much the same way. For example, if $F : \omega^a \to \text{Grp}$, as displayed in (1), then each $F_n$ is a group, the set $L$ of all cones (all matching strings $x$) is also a group under pointwise multiplication $(\langle x \rangle)_n = x_n x_n$, and, the projection $\mu_n : L \to F_n$ with $x \mapsto x_n$ is a group homomorphism, so that $\mu : \to F$ is a limiting cone in $\text{Grp}$.

The $p$-adic integers $\mathbb{Z}_p$ (with $p$ a prime) illustrate this construction. Take $F : \omega^p \to \text{Rng}$ with $F_n = \mathbb{Z}/p^n\mathbb{Z}$, the ring of integers modulo $p^n$; then $F_n = \text{Cone}(\ast, F)$.

The canonical projection $Z/p^n\mathbb{Z} \to Z/p^{n+1}\mathbb{Z}$ then $\mathbb{Z}_p = \text{lim} F$ exists. An element $\lambda$ of $\mathbb{Z}_p$ is a cone from $\ast$ to $F$; that is, $\lambda$ can be written as a sequence $\lambda = \{\lambda_0, \lambda_1, \ldots\}$ of integers with $\lambda_{n+1} = \lambda_n (\text{mod } p^n)$ for all $n$, where $\lambda = \lambda^p$ holds when $\lambda_n = \lambda_n (\text{mod } p^n)$ for all $n$. Two $p$-adic integers $\lambda$ and $\mu$ can be added and multiplied "termwise", by the formulas

$$(\lambda + \mu)_n = \lambda_n + \mu_n, \quad (\lambda \mu)_n = \lambda_n \mu_n.$$}

These operations make $\mathbb{Z}_p = \text{Lim} F$ a ring, the ring of $p$-adic integers, and this description completely determines $\mathbb{Z}_p$. This description is quicker than the classical one, which first defines a $p$-adic valuation (and thus a topology) in $\mathbb{Z}$, and then observes that each $p$-adic integer $\lambda$ is represented by a Cauchy sequence in that topology.

Formal power series rings also can be described as limits (Ex. 7).

Again, in $\text{Top}$, take each object $F_n$ to be a circle $S^1$, and each arrow $f_n : S^1 \to S^1$ to be the continuous map wrapping the domain circle $S^1$ uniformly $p$ times around the codomain circle. The inverse limit set $L$ then becomes a topological space when we introduce just those open sets in $L$ necessary to make all the functions $\mu_n : L \to S^1$ continuous. This $L$ is the limit space in $\text{Top}$; it is known as the $p$-adic solenoid.

Here is the general construction for groups.

**Theorem 2.** Let $U : \text{Grp} \to \text{Set}$ be the forgetful functor. If $H : J \to \text{Grp}$ is such that the composite $U H$ has a limit $L$ and a limiting cone $\nu : L \to U H$ in $\text{Set}$, then there is exactly one group structure on the set $L$ for which each arrow $v_j : L \to U H_j$ of the cone $\nu$ is a morphism of groups; moreover, this group $L$ is a limit of $H$ with $\nu$ as limiting cone.

**Proof.** By Theorem 1, take $L = \text{Cone}(\ast, U H)$; define the product of two such cones $\sigma, \tau \in \text{Cone}(\ast, U H)$ by $\sigma \cdot \tau = \sigma \tau_j$ (the product in the group $H_j$) and the inverse by $\sigma^{-1} = \tau^{-1}$ (the inverse in $H_j$). These definitions make $L$ a group and each component of $\nu$ a morphism of groups; conversely, if $\nu$ given by $\nu_j$ is to be a group of $\sigma_j$ for each $j$, then the product of $\sigma_j \in L$ must be given by this formula.

Now if $G$ is any group and $\lambda : G \to H_j$ any cone in $\text{Grp}$ (consisting of group morphisms $\lambda_j : G \to H_j$ for $j \in J$), then $U \lambda : U G \to U H$ is a cone in $\text{Set}$, so by universality $U \lambda = (U \nu) h$ for a unique function $h : U G \to L$.

For any two group elements $g_1, g_2$ in $G$,

$$(h(g_1 g_2))_j = \lambda_j(g_1 g_2) = (\lambda_j g_1)(\lambda_j g_2) = (h g_1)(h g_2) = ((h g_1)(h g_2))_j,$$

because $\lambda$ is a morphism of groups, so is $h$, and therefore $L$ is indeed the limit in $\text{Grp}$.

This argument is just a formalization of the familiar termwise construction of the multiplication in cartesian products of groups, in the $p$-adic numbers, etc. The conclusion of the Theorem constructs limits in $\text{Grp}$ from the limits in $\text{Set}$ in a unique way, using $U$. The same argument will construct all small limits in $\text{Rng}$, $\text{Ab}$, $\text{R-Mod}$ and similar algebraic categories, using the forgetful functors $U$ to $\text{Set}$. In other words, each forgetful functor "creates" limits in the sense of the following definition:
Definition. A functor \( V : A \to X \) creates limits for a functor \( F : J \to A \) if

(i) To every limiting cone \( \tau : x \to V F \) in \( X \) there is exactly one pair \((a, \sigma)\) consisting of an object \( a \in A \) with \( V a = x \) and a cone \( \sigma : a \to F \) with \( V \sigma = \tau \), and if, moreover,

(ii) This cone \( \sigma : a \to F \) is a limiting cone in \( A \).

Similarly, we may define "\( V \) creates products" (the above, with \( J \)
restricted to be discrete); "\( V \) creates finite limits" (the above, with \( J \)
finite), or "\( V \) creates colimits" (the above with the arrows in all cones
reversed). Note especially that "\( V \) creates limits" means only that \( V \)
produces limits for functors \( F \) whose composite \( VF \) already has a limit.

In this terminology, Theorem 2 now reads

**Theorem 3.** The forgetful functor \( U : \text{Grp} \to \text{Set} \) creates limits.

**Exercises**

1. Prove that the projections \((x \upharpoonright C) \to C\) of the comma category create limits.
2. If \( \text{Comp Haus} \to \text{Top} \) is the full subcategory of all compact Hausdorff spaces, show
   that the forgetful functor \( \text{Comp Haus} \to \text{Set} \) creates limits.
3. For any category \( X \), show that the projection \( X^I \to X \times X \) which sends each arrow \( f : x \to y \) in \( X \) to the pair \((x, y)\), creates limits.
4. Prove that the category of all small finite sets is **finitely complete** (i.e., has all
   finite limits).
5. Prove that \( \text{Cat} \) is small-complete.
6. Show that each \( p \)-adic integer \( J \) is determined by a string of integers \( a_i \), with
   all \( a_i \in \{0, 1, \ldots, p-1\} \), each \( J_n \equiv a_0 + a_1 p + \cdots + a_n p^n \) (mod \( p^n \)). Show that
   addition and multiplication of \( p \)-adic integers correspond to the usual operations
   of addition and multiplication of infinite "decimals" \( a_0 a_1 \ldots a_n \) with base \( p \),
   the decimals extending infinitely to the left.
7. Let \( K[x] \) be the usual ring of polynomials in \( x \) with coefficients in the
   commutative ring \( K \), while \( F : \mathbb{F}_p^a \to \text{Rng} \) is defined by \( F_a = K[x]/(x^n) \), with the evident
   projections, and \( (x^n) \) the usual principal ideal. Prove that \( \text{Lim} F \) is the ring of
   formal power series in \( x \), coefficients in \( K \).

2. **Limits by Products and Equalizers**

   The construction of the limit of \( F : J \to \text{Set} \) as the set of all cones
   
   \[
   \text{Cone}(\ast, F) \subset \Pi_{j \in J} F_j
   \]

   can be made in two steps: Each cone \( \sigma \) is an element \( x \) of the product
   \( \Pi_{j \in J} F_j \) with projections \( p_j \); to require that an element \( x \) of the product be a
   cone is to require that \( (Fu)_j = x_j \) for every arrow \( u : j \to k \) in \( J \); this
   amounts to requiring that \( x \) lie in the equalizer of \( (Fu) \) \( p_j \) and \( p_k : \Pi_{j \in J} F_j \to F_k \).
   Here is the general formulation:

**Theorem 1.** For categories \( C \) and \( J \), if \( C \) has equalizers of all pairs
of arrows and all products indexed by the sets \( \text{obj}(J) \) and \( \text{arr}(J) \), then
\( C \) has a limit for every functor \( F : J \to C \).

The proof constructs the following diagram in stages, with \( i \) denoting
an object and \( u : j \to k \) an arrow of the index category \( J \). By assumption,
the products \( \Pi_{j \in J} F_j \) and \( \Pi_{j \in J} F_k \) and their projections exist, where the second
product is taken over all arrows \( u \) of \( J \), with argument at each arrow \( u \) the
value \( F_u = F_{\text{cod} u} \) of \( F \) at the codomain object of \( u \). Since \( \Pi_{j \in J} F_j \) is a product,
there is a unique arrow \( f \) such that the upper square commutes for every \( u \) and
a unique arrow \( g \) such that the lower square commutes for every \( u \). By hypothesis,

\[
\Pi_{j \in J} F_k = \Pi_{j \in J} F_{\text{cod} u} \quad \text{(1)}
\]

there exists an equalizer \( e \) for \( f \) and \( g \). Its composite with the projections
\( p_j \) gives arrows \( \mu_j = p_j e : d \to F_j \) for each \( i \). Since \( e \) equalizes \( f \) and \( g \), one
has \( F_{\mu_j} = \mu_k \) for every \( u : j \to k \); hence \( \mu : d \to F \) is a cone from the
vertex \( d \) to the base \( F \). If \( \tau \) is any other such cone, of vertex \( e \), its maps
\( \tau \), combine to yield a unique map \( h : e \to \Pi_{j \in J} F_j \) to the product; \( \tau \) a cone implies \( fh = gh \). Hence \( h \) factors uniquely through \( e \) and therefore the
cone \( \tau \) factors uniquely through the cone \( \mu \). This proves that \( d \) and the
cone \( \mu \) provide a limit for \( F \). For the record:

**Theorem 2 (Limits by product and equalizers, continued).** The limit
of \( F : J \to C \) is the equalizer \( e \) of \( f, g : \Pi_{j \in J} F_j \to \Pi_{j \in J} F_{\text{cod} u} \) (\( u \in \text{arr} J, \ i \in J \)),
where \( p_j f = p_{\text{cod} u} \) \( p_j g = p_{\text{dom} u} \); the limiting cone \( \mu \) is \( \mu_j = p_j e \), for
\( j \in J \), all as in (1).

This theorem has several useful consequences and special cases.

**Corollary 1.** If a category \( C \) has a terminal object, equalizers of all
pairs of arrows, and products of all pairs of objects, then \( C \) has all finite
limits.

Here \( a \) finite limit is a limit of \( J \to C \), with the category \( J \) finite.

**Corollary 2.** If \( C \) has equalizers of all pairs of arrows and all small
products, then \( C \) is small-complete.

For example, this gives another proof that \( \text{Set} \) is small-complete.

The concept of completeness is useful chiefly for large categories and
for preorders. In a preorder \( P \), a product of objects \( a_j \in F \), is an object \( d \)
with \( d \leq a_j \) for all \( j \) and such that \( c \leq a_j \), for all \( j \) implies \( c \leq d \); in other
words a product is just a greatest lower bound or meet of the factors $a_j$ (dually, a coproduct is a least upper bound or join).

**Proposition 3.** (Freyd). A small category $C$ which is small-complete is simply a preorder which has a greatest lower bound for every small set of its elements.

**Proof.** Suppose $C$ is not a preorder. Then there are objects $a, b \in C$ with arrows $f : g : a \to b$. For any small set $J$ form the product $\prod_J b$ of factors $b_j$ all equal to $b$. Then an arrow $h : a \to \prod_J b$ is determined by its components, which can be $f$ or $g$. There are thus at least $2^J$ arrows $a \to \prod_J b$. If the small set $J$ has cardinal larger than $\text{Arr} C$, this is a contradiction.

**Exercises**

1. (Manes) A parallel pair of arrows $f, g : a \to b$ in $C$ has a common left inverse $h$ when there is an arrow $h : b \to a$ with $hf = 1 = hg$.

   (a) Prove that a category $C$ with all small products and with equalizers for all parallel pairs with a common left inverse is small complete. (Hint: The parallel pair used in the proof of Theorem 1 does in fact have a common left inverse.)

   (b) In $\text{Set}$, show that a parallel pair of arrows $f, g : X \to Y$ has a common right inverse if and only if the corresponding function $(f, g) : X \times Y \to Y$ has image containing the diagonal $\{(x, y) : x = y \} \subseteq Y$.

2. Prove that $C_1, C_2$ complete (or cocomplete) imply the same for the product category $C_1 \times C_2$.

3. ($\text{Lim}$ and $\text{Lim}$ as functors). If $F, F' : J \to C$ have limiting cones $\mu, \mu'$ (or colimiting cones $v, v'$), show that each natural transformation $\beta : F \to F'$ determines uniquely arrows $\lim \beta$ or $\lim \beta^\prime$ such that the following diagram commutes, where $\Delta : C \to C'$ is the diagonal functor:

   $\begin{array}{ccc} 
   \Delta \text{Lim} F & \longrightarrow & \Delta \text{Lim} F' \\
   \downarrow \text{Lim} F & f & \downarrow \text{Lim} F' \\
   \Delta \text{Lim} F' & \longrightarrow & \Delta \text{Lim} F'. 
   \end{array}$

   Conclude: If $C$ is complete, $\text{Lim}$ (or $\text{Lim}$) is a functor $C' \to C$.

4. (Limits of composites). Given composable functors $J : \mathcal{J} \to \mathcal{J}, \mathcal{J}, \mathcal{C}$

and limiting cones $v, v'$ for $F, F'$ for $\text{HFW}$, observe that $\Delta_J (H C) = H \times \Delta_J C = W : J \to C'$, and show that there is a unique "canonical" arrow $\gamma : H \times \text{Lim} F \to \text{Lim} \text{HFW}$ such that the following diagram commutes

$\begin{array}{ccc} 
   \Delta_J (H \times \text{Lim} F) & \longrightarrow & \Delta_J (\text{Lim} \text{HFW}) \\
   \downarrow H \times \text{Lim} F & \gamma & \downarrow \text{Lim} \text{HFW} \\
   \Delta_J (\text{Lim} \text{HFW}) & \longrightarrow & \Delta_J (H \times \text{Lim} F) 
   \end{array}$

Dually, construct $\delta : \text{Lim} \text{HFW} \to H \times \text{Lim} F$ as indicated at the right.

5. (Limit as a functor on the comma category of all diagrams in $C$)

   (a) Interpret $W$ of Ex. 4 as an arrow in $(\text{Cat} \downarrow C)$ to show (for $C$ complete) that $\text{Lim}$ is a functor $(\text{Cat} \downarrow C)^{op} \to C$.

   (b) Let $(\text{Cat} \downarrow C)$ be the ("super-comma") category with objects $F : J \to C$, arrows $\langle \beta, W \rangle : F \to F'$ those pairs consisting of a functor $W : J \to J$ and a natural transformation $\beta : F W \to F'$. Combine Exercise 3 and Exercise 4 to show (for $C$ complete) that $\text{Lim}$ is a functor $(\text{Cat} \downarrow C)^{op} \to C$. Dualize.

3. **Limits with Parameters**

Let $T : J \times P \to X$ be a bifunctor, and suppose for each value $p \in P$ of the "parameter" $p$ that $T(\cdot, p) : J \to X$ has a limit. Then these limits for all $p$ form the object function $p \mapsto \text{Lim} T(p, \cdot)$ of a functor $P \to X$.

Instead of proving this directly, we replace functors $P \to X$ by objects of the functor category $X^P$. This replaces $T : J \times P \to X$ by its adjunct $S : J \to X^P$, under the adjunction $\text{Cat} (J \times P, X) \cong \text{Cat} (J, X^P)$. Recall that for each $p \in P$ there is a functor $E_p : X^P \to X$, "evaluate at $p$", given for arrows (natural transformations) $\sigma : H \to H'$ of $X^P$

$E_p H = H_p, \quad E_p \sigma = \sigma_p : H_p \to H_p'.$

**Theorem 1.** If $S : J \to X^P$ is such that for each object $p \in P$ the composite $E_p S : J \to X$ has a limit $L_p$ with a limiting cone $\tau_p : L_p \to E_p S$, then there is a unique functor $L : P \to X$ with object function $p \mapsto L_p$ such that $p \mapsto \tau_p$ is a natural transformation $\tau : \Delta L \to \Delta L_p : S$; moreover, $\tau$ is a limiting cone from the vertex $L \in X^P$ to the base $S : J \to X^P$.

**Proof.** Let $h : p \to q$ be any arrow of $P$. Then, writing $E_p S$ as $S_p$, the given cones $\tau_p$ and $\tau_q$ for a typical arrow $u : j \to k$ of $J$ have the form

$\begin{array}{ccc}
   L_p & \downarrow & L_q \\
   \downarrow \tau_p & \searrow & \tau_q \\
   S_p & \longrightarrow & S_q \\
   \quad \downarrow \tau_p & \quad \downarrow \tau_q \\
   S_{j,k} & \rightarrow & S_{k,j} \\
   \quad \downarrow \tau_p & \quad \downarrow \tau_q \\
   S_p & \rightarrow & S_q \\
\end{array}$

The triangles commute because $\tau_p$ and $\tau_q$ are cones and the parallelograms because $S$ is a functor. Since the inside cone is universal there is a unique arrow $L_j : L_p \to L_q$ such that $\tau_j = L_j \tau_p$ for all $j \in J$. The assignment $h \mapsto L_h$ makes $L$ a functor (Proof: put another cone outside) and $\tau$ a natural transformation $\Delta L \to S$ (a cone from the object $L \in X^P$ to the functor $S : J \to X^P$). It is a limiting cone; for if $\sigma : M \to S$ is any
other cone there are unique arrows \( M_p \rightarrow L_p \) because \( L_p \) is a limit; they combine to give a unique natural transformation \( M \rightarrow L \).

The conclusion may be written
\[
E_p(\text{Lim } S) = \text{Lim } (E_p S).
\]

In a functor category, limits may be calculated pointwise (provided the pointwise limits exist).

**Corollary.** If \( X \) is small-complete, so is every functor category \( X^p \).

This theorem becomes a case of “creation” of limits, if we write \([P] \) for the discrete subcategory consisting of all objects and identity arrows of \( P \).

**Theorem 2.** For any categories \( X \) and \( P \), the inclusion functor \( i : [P] \rightarrow P \) induces a functor \( * = X^i : X^P \rightarrow X^{[P]} \) which creates limits.

### 4. Preservation of Limits

A functor \( H : C \rightarrow D \) is said to **preserve the limits** of functors \( F : J \rightarrow C \) when every limiting cone \( v : b \rightarrow F \) in \( C \) for a functor \( F \) yields by composition with \( H \) a limiting cone \( Hv : Hb \rightarrow HF \) in \( D \); this requires not only that \( H \) take each limit object which exists in \( C \) to a limit object in \( D \) but also that \( H \) take limiting cones to limiting cones. A functor is called **continuous** when it preserves all small limits.

**Theorem 1.** For any category \( C \) with small hom-sets, each hom-functor \( C(c, -) : C \rightarrow \text{Set} \) preserves all limits; in particular, all small limits.

The same proof will give a more general result: If \( C \) has hom-sets in \( \text{Ens} \), any category of sets in which \( \text{Ens}(X, Y) \) consists of all functions on \( X \) to \( Y \), then each hom-functor \( C(c, -) : C \rightarrow \text{Ens} \) preserves all limits which exist in \( C \).

**Proof.** Let \( J \) be any category and \( F : J \rightarrow C \) a functor with a limiting cone \( v : \text{Lim } F \rightarrow F \) in \( C \). Apply the hom-functor \( C(c, -) \); there results a cone \( C(c, v) \), as in the diagram
\[
\begin{array}{ccc}
C(c, \text{Lim } F) & \xrightarrow{\psi} & C(c, F_i), \\
\downarrow & & \downarrow \\
X & \xrightarrow{\phi} & C(c, F_i)
\end{array}
\]

in \( \text{Set} \). For any other cone \( \tau \) to the same base from a vertex set \( X \), each element \( x \in X \) gives a cone \( \tau_x : c \rightarrow F_i \) in \( C \) and hence, because \( v \) is universal, a unique arrow \( h_x : \text{Lim } F \rightarrow \text{Lim } F \) with \( v h_x = \tau_x \). Then setting \( k_x = h_x \) for each \( x \) defines a function, and hence an arrow \( k \) in \( \text{Ens} \) as

\[
\text{shown, with } v_x k = \tau_x \text{ for all } x. \text{ Since } k \text{ is clearly unique with this property, }
\]

\( v_x \) is a limiting cone in \( \text{Set} \), as required.

The same proof, differently stated, might start by noting that the definition of the functor \( C(c, F) : J \rightarrow \text{Set} \) shows that a cone \( \lambda : c \rightarrow F \) in \( C \) is the same thing as a cone \( \lambda : J \rightarrow C(c, F) \) in \( \text{Set} \), with vertex a point \( \star \). Then, because \( \text{Cone}(X, -) \cong \text{Set}(X, \text{Cone}(\star, -)) \) as in (1.3),

\[
\text{Cone}(X, C(c, F)) \cong \text{Set}(X, \text{Cone}(\star, C(c, F)))
\]

\[
= \text{Set}(X, \text{Cone}(c, F)) \subseteq \text{Set}(X, C(c, \text{Lim } F)),
\]

where the last step uses the definition of \( \text{Lim } F \). But \( \text{Lim } S \), for each \( S : J \rightarrow \text{Set} \), is defined by the adjunction \( \text{Cone}(X, S) \cong \text{Set}(X, \text{Lim } S) \). Therefore the above equations determine \( \text{Lim } S \) (together with the correct limiting cone)

\[
\text{Lim } C(c, F) \cong C(c, \text{Lim } F).
\]

Some authors use this equation to define limits in \( C \) in terms of limits in \( \text{Set} \); for example, the product of objects \( a_i \) is defined by

\[
\prod C(c, a_i) \cong C(c, \prod a_i).
\]

The contravariant hom-functor may be written as

\[
C(-, c) = C^{op}(c, -) : C^{op} \rightarrow \text{Set};
\]

hence the theorem shows that this functor \( C(-, c) \) carries small colimits (and their colimiting cones) in \( C \) to the corresponding limits and limiting cones in \( \text{Set} \). For example, the definition of a small coproduct provides an isomorphism (coproduct to product):

\[
C(\bigcup a_i, c) \cong \prod C(a_i, c).
\]

More generally, the colimit of any \( F : J \rightarrow C \) is determined by

\[
C(\text{Colim } F, c) \cong \text{Lim } C(F, -). \quad (3)
\]

Creation and preservation are related:

**Theorem 2.** If \( V : A \rightarrow X \) creates limits for \( F : J \rightarrow A \) and the composite \( VF : J \rightarrow X \) has a limit, then \( V \) preserves the limit of \( F \).

In particular, if \( V \) creates all small limits and \( X \) is small-complete, then \( A \) is also small-complete, and \( V \) is continuous.

**Proof.** Let \( \tau : a \rightarrow F \) and \( a : x \rightarrow VF \) be limiting cones in \( A \) and \( X \), respectively. Since \( V \) creates limits, there is a unique cone \( q : b \rightarrow F \) in \( A \) with \( Vq : Vb \rightarrow VF \) equal to \( \sigma : x \rightarrow VF \); moreover, \( q \) is a limiting cone. But limits are unique up to isomorphism, so there is an isomorphism \( \theta : b \cong a \) with \( \tau \theta = q \). Thus \( V \theta : Vb \cong x \cong Va \), with \( V\tau \circ V\theta = Vq = a \), so \( Va \) is a limit and \( V \) preserves limits, as desired.
In any category an object \( p \) is called *projective* if every arrow \( h : p \to c \) from \( p \) factors through every epi \( g : b \to c \), as \( h = gh' \) for some \( h' \).

\[
\begin{array}{ccc}
  p & \xrightarrow{h} & c \\
  \downarrow & & \downarrow \\
  b & \xrightarrow{g} & c
\end{array}
\]

It is equivalent to require that \( g \) epi implies \( \text{hom}(p, g) : \text{hom}(p, b) \to \text{hom}(p, c) \) epi in \( \text{Set} \). In other words, \( p \) is projective exactly when \( \text{hom}(p, -) \) preserves epis. Dually, an object \( q \) is *injective* when \( \text{hom}(-, q) \) carries monics to epis. These notions are especially useful in \( \text{R-Mod} \) and other Ab-categories; in \( \text{R-Mod} \) the projectives are the direct summands of the free modules.

**Exercises**

1. Prove that the composite of continuous functors is continuous.
2. If \( C \) is complete, and \( H : C \to D \) preserves all small products and all equalizers (of parallel pairs) prove that \( H \) is continuous.
3. Show that the functor \( F : \text{Set} \to \text{Ab} \) sending each set \( X \) to the free abelian group generated by the set \( X \) is not continuous.
4. For any small set \( X \), show that the functor (product with \( X \)) \( X \times - : \text{Set} \to \text{Set} \) preserves all colimits.
5. (Preservation of Limits) Given \( H : C \to C' \) and a functor \( F : J \to C \) such that \( F \) and \( HF \) have limits, prove that \( H \) preserves the limits of \( F \) if and only if the canonical arrow \( H \cdot \text{Lim} F \to \text{Lim} HF \) of Exercise 2.4 is an isomorphism (This is a natural way to describe the preservation of limits when both categories \( C \) and \( C' \) are given with specified limits).

5. Adjoints on Limits

One of the most useful properties of adjoints is this: A functor which is a right adjoint preserves all the limits which exist in its domain.

**Theorem 1.** If the functor \( G : A \to X \) has a left adjoint, while the functor \( T : J \to A \) has a limiting cone \( \tau : a \to T \) in \( A \), then \( GT \) has the limiting cone \( G\tau : Ga \to GT \) in \( X \).

**Proof.** By composition, \( Gr \) is indeed a cone from the vertex \( Ga \) in \( X \). If \( F \) is a left adjoint to \( G \), and if we apply the adjunction isomorphism to every arrow of a cone \( \sigma : x \to GT \), we get arrows \( \langle \sigma \rangle : Fx \to T \) for \( i \in J \) which form a cone \( \sigma : Fx \to T \) in \( A \). But \( \tau : a \to T \) is universal among cones to \( T \) in \( A \), so there is a unique arrow \( h : Fx \to a \) with \( h\tau = \sigma \). Taking adjuncts again, this gives a unique arrow \( h' : x \to Ga \) with \( (h')\tau = Gr \cdot h' = (\sigma')T = \sigma \). The uniqueness of the arrow \( h' \) states precisely that \( Gr : Ga \to T \) is universal, q.e.d.

The proof may be illustrated by the following diagrams (where \( u : i \to j \) is any arrow of \( J \)).

\[
\begin{array}{ccc}
  \text{in } A & & \text{in } X \\
  T_i & \xrightarrow{\tau_i} & a \\
  u & \xrightarrow{\gamma u} & Ga \\
  T_j & \xrightarrow{\tau_j} & Fx, \\
  Gr & \xrightarrow{G\tau_j} & GT_j & \xrightarrow{\gamma_j} & x.
\end{array}
\]

This proof can also be cast in a more sophisticated form by using the fact that \( \text{Lim} \) is right adjoint to the diagonal functor \( \Delta \). In fact, given an adjunction

\[
\langle F, G, \eta, \epsilon \rangle : X \to A
\]

and any index category \( J \), one may form the functor categories (from \( J \)) and hence the diagram

\[
\langle F^J, G^J, \eta^J, \epsilon^J \rangle : X^J \to A^J,
\]

where \( F^J = FS \) for each functor \( S : J \to X \), and \( \eta^J S = \eta S : S \to GFS \), etc. The triangular identities for \( \eta \) and \( \epsilon \) yield the same identities for \( \eta^J \) and \( \epsilon^J \), so the second diagram is indeed an adjunction (in brief, adjunctions pass to the functor category). Now we have the diagram of adjoint pairs

\[
\begin{array}{ccc}
  X^J & \xrightarrow{\text{Lim}} & A^J \\
  X & \xrightarrow{\text{Lim}} & A
\end{array}
\]

The definitions of the diagonal functors \( \Delta \) show at once that \( F^J \Delta = \Delta F \), so the diagram of left adjoints commutes in this square. Since compositions of adjoints give adjoints, it follows that the composites \( \text{Lim} \circ G^J \) and \( G \circ \text{Lim} \) are both right adjoints to \( F^J \circ \Delta = \Delta \circ F \). Since the right adjoint of a given functor is unique up to natural isomorphism, it now follows that \( \text{Lim} \circ G^J \cong G \circ \text{Lim} \). This proves again for each functor \( T : J \to A \) with limit \( a \) (and limiting cone \( \tau : a \to T \) in \( A \)) that \( Ga = G\text{Lim} T = \text{Lim} G(T) = \text{Lim} GT \). The reader should show that the same argument proves that \( G \) preserves limiting cones (put units and counits in the square diagram above, and recall that the limiting cone \( \tau : a \to T \) is just the value of the counit of the adjunction \( \langle \Delta, \text{Lim}, \ldots \rangle : A \to A^J \) on the functor \( T \)).

The dual of the theorem is equally useful: Any functor \( P \) which has a right adjoint (i.e., which is a left adjoint) must preserve colimits (coproducts, coequalizers, etc.). This explains why the coproduct (free product) of two free groups is again a free group (on the disjoint union of the sets of generators).
Similarly (by the original theorem) all the typical forgetful functors in algebra preserve products, kernels, equalizers, and other types of limits. Typically, the product of two algebraic systems (groups, rings, etc.) has as underlying set just the (cartesian) product of the two underlying sets. This, and other similar facts, are immediate consequences of this one (easy) theorem. The theorem can also be used to show that certain functors do not have adjoints.

Exercises

1. Show that, for a fixed set \( X \), the functor \( X \times - : \text{Set} \to \text{Set} \) cannot have a left adjoint, unless \( X \) is a one-point set.

2. For the functor \( D : \text{Vct}^{op} \to \text{Vct} \) of (IV.2) show that \( D \) has no right adjoint (and hence, in particular, is not the left adjoint of \( D^{op} \)).

3. If \( C \) is a full and reflective subcategory of a small-cocomplete category \( D \), prove that \( C \) is small-cocomplete.

4. Prove that \( \text{Set}^{op} \) is not cartesian closed.

6. Freyd's Adjoint Functor Theorem

To formulate the basic theorem for the existence of a left adjoint to a given functor, we first treat the case of the existence of an initial object in a category and then use the fact that each universal arrow defined by the unit of a left adjoint is an initial object in a suitable comma category.

**Theorem 1. (Existence of an initial object).** Let \( D \) be a small-complete category with small hom-sets. Then \( D \) has an initial object if and only if it satisfies the following

**Solution Set Condition.** There exists a small set \( I \) and an \( I \)-indexed family \( k_i \) of objects of \( D \) such that for every \( d \in D \) there is an \( i \in I \) and an arrow \( k_i \to d \) of \( D \).

**Proof.** This solution set condition is necessary: If \( D \) has an initial object \( k \), then \( k \) indexed by the one-point set realizes the condition, since there is always a (unique) arrow \( k \to d \).

Conversely, assume the solution set condition. Since \( D \) is small-complete, it contains a product object \( w = \prod k_i \) of the given \( I \)-indexed family. For each \( d \in D \), there is at least one arrow \( w \to d \), for example, a composite \( w = \prod k_i \to k_i \to d \), where the first arrow is a projection of the product. By hypothesis, the set of endomorphisms \( D(w,w) \) of \( w \) is small and \( D \) is complete, so we can construct the equalizer \( e : w \to w \) of the set of all the endomorphisms of \( w \). For each \( d \in D \), there is by \( v \to w \to d \) at least one arrow \( v \\to d \). Suppose there were two, \( f, g : v \to d \), and take

\[
\begin{array}{c}
\text{Exercise 6. Freyd's Adjoint Functor Theorem}
\end{array}
\]

their equalizer \( e \) as in the figure below

\[
\begin{array}{ccc}
u & \xrightarrow{e} & v \\
\downarrow & & \downarrow f \\
d & \xrightarrow{g} & d
\end{array}
\]

By the construction of \( w \), there is an arrow \( s : w \to u \), so the composite \( e e_s \) is, like \( 1_u \), an endomorphism of \( w \). But \( e \) was defined as the equalizer of all endomorphisms of \( w \), so

\[
e e_s = 1_w e = e_1 u.
\]

Now \( e \) is an equalizer, hence is monic; cancelling \( e \) on the left gives \( e_1 s e = 1_u \). This states that the equalizer \( e_1 \) of \( f \) and \( g \) has a right inverse. Like any equalizer, \( e_1 \) is monic, hence is an isomorphism. Therefore, \( f = g \); this conclusion means that \( u \) is initial in \( D \).

This proof will be reformulated in § X.2.

**Theorem 2. (The Freyd Adjoint Functor Theorem).** Given a small-complete category \( A \) with small hom-sets, a functor \( G : A \to X \) has a left adjoint if and only if it preserves all small limits and satisfies the following

**Solution Set Condition.** For each object \( x \in X \) there is a small set \( I \) and an \( I \)-indexed family of arrows \( f_i : x \to G a_i \) such that every arrow \( h : x \to G a \) can be written as a composite \( h = G i \circ f_i \) for some index \( i \) and some \( a_i \to a \).

**Proof.** If \( G \) has a left adjoint \( F \), then it must preserve all the limits which exist in its domain \( A \); in particular, all the small ones. Moreover, the universal arrow \( \eta_x : x \to GFx \) which is the unit of the adjunction satisfies the solution set condition for \( x \), with \( I \) the one-point set.

Conversely, given these conditions, it will suffice to construct a universal arrow \( x \to G a \) from each \( x \in X \) to \( G \); then \( G \) has a left adjoint by the pointwise construction of adjoints. This universal arrow is an initial object in the comma category \( (x \downarrow G) = D \), so we need only verify the conditions of the previous theorem for this category. The solution set condition for \( G \) clearly gives the condition of the same name for \( (x \downarrow G) = D \). Since \( A \) has small hom-sets, so does \( D \). To show \( D \) small-complete we need only arbitrary small products and equalizers of parallel pairs. They may be created as follows:

**Lemma.** If \( G : A \to X \) preserves all small products (or, all equalizers) then for each \( x \in X \) the projection

\[
Q : (x \downarrow G) \to A, \quad (x \to G a) \mapsto a
\]

of the comma category creates all small products (or, all equalizers).
Proof. Let \( J \) be a set (a discrete category) and \( f_j: x \to G a_j \) a \( J \)-indexed family of objects of \((x \downarrow G)\) such that the product diagram \( \Pi a_j \to a_j \) exists in \( A \). Since \( G \) preserves products, \( G \Pi a_j \to G a_j \) is a product diagram in \( X \), so there is a unique arrow \( f: x \to G \Pi a_j \) in \( X \) with \((G f_j) = f_j \) for all \( j \):

\[
\begin{array}{ccc}
\Pi a_j & \xrightarrow{f} & G \Pi a_j \\
\text{ } & \downarrow{f_j} & \text{ } \\
\text{ } & G a_j & \text{ }
\end{array}
\]

This equation states that \( f_j: f \to f_j \) is a cone of arrows in \((x \downarrow G)\); indeed, it is the unique cone there which projects under \( Q \) to the given cone \( f_j: \Pi a_j \to a_j \). One then verifies that this cone \( f \) is a product diagram in \((x \downarrow G)\); these two results show that \( Q \) creates products.

Similarly, we “create” the equalizer of two arrows \( s, t: f \to g \) in \((x \downarrow G)\). As in the figure below, we are given the equalizer \( e \) of \( Q s, Q t \); that is, of \( s \) and \( t \) as arrows in \( A \). Since \( G \) preserves equalizers, \( Ge \) is then the equalizer of \( G s \) and \( G t \). But \( G s \cdot f = G t \cdot f \), so there is a unique arrow \( h: x \to G a \) making \( G e \cdot h = f \), as below. In other words \( e: h \to f \) in \((x \downarrow G)\) is the unique arrow of \((x \downarrow G)\) with \( Q \)-projection \( e: a \to b \).

\[
\begin{array}{ccc}
x & \xrightarrow{h} & Ga \\
\text{ } & \downarrow{ge} & \text{ } \\
\text{ } & Gd & \text{ }
\end{array}
\]

\[
\begin{array}{ccc}
x & \xrightarrow{f} & Gb \\
\text{ } & \downarrow{gr} & \text{ } \\
\text{ } & Gc & \text{ }
\end{array}
\]

It remains to show that the arrow \( e \) is an equalizer in \((x \downarrow G)\). So consider another object \( k: x \to Gd \) of \((x \downarrow G)\) and an arrow \( r: k \to f \) of \((x \downarrow G)\) with \( sr = tr \) in \((x \downarrow G)\). Then \( sr = tr \) in \( A \), so there is a unique \( r' \) in \( A \) with \( r = er' \). It remains only to show \( r' \) an arrow \( k \to h \) of \((x \downarrow G)\); but \( Ge(r') \cdot k = G(e) \cdot k = Gr \cdot k = f \), so by the unique choice of \( h \), \( Gr \cdot k = h \), which states that \( r' \) is an arrow of \((x \downarrow G)\).

This line of argument applies not just to products or equalizers, but to the creation of any limit (Exercise 1).

Theorem 3. (The Representability Theorem). Let the category \( D \) be small complete with small hom-sets. A functor \( K: D \to \text{Set} \) is representable if and only if \( K \) preserves all small limits and satisfies the following Solution Set Condition. There exists a small set \( S \) of objects of \( D \) such that for any object \( d \in D \) and any element \( x \in Kd \) there exist an \( s \in S \), an element \( y \in Ks \) and an arrow \( f: s \to d \) with \((Kf)(y) = x \).

Proof. This is another reformulation of the existence Theorem 1 for initial objects. Indeed, a representation of \( K \) is a universal arrow from the one-point set \( * \) to \( K \) (Proposition III.2.2), hence an initial object in the comma category \((* \downarrow K)\), which is small-complete because \( K \) is assumed continuous. Conversely, if \( K \) is representable, it is necessarily continuous.

The solution set condition (or something like it) is requisite in all three theorems. For an example, let \( \text{Ord} \) be the ordered set of all small ordinal numbers \( \alpha, \beta, \ldots \); it is a category with hom-set \( \text{Ord}(\alpha, \beta) \) empty or the one-point set according as \( \alpha > \beta \) or \( \alpha \leq \beta \). The category \( \text{Ord} \) is small-complete, because the product of any small set of ordinals is their least upper bound. The functor \( K: \text{Ord} \to \text{Set} \) with \( Kx = * \) the one-point set for every \( x \) is clearly continuous. However \( K \) is not representable: Were \( Kx \simeq \text{Ord}(\beta, x) \) for some \( \beta \), then \( x \leq \beta \) for all \( x \), so \( \beta \) would be a largest small ordinal, which is known to be impossible.

Complete Boolean algebras provide another example to show that some solution set condition is requisite. For a given denumerable set \( D \) one can construct an arbitrarily large complete Boolean algebra generated by \( D \) (Solovay [1966]); this implies that there is no free complete Boolean algebra generated by \( D \), and hence that the forgetful functor \( \text{Comp Boo!} \to \text{Set} \) has no left adjoint—though it is continuous and \( \text{Comp Boo!} \) is small-complete.

The adjoint functor theorem has many applications.

For example, it gives a left adjoint to the forgetful functor \( U: \text{Grp} \to \text{Set} \). Indeed, we already know that \( U \) creates all limits (Theorem 1.3), hence that \( \text{Grp} \) is small-complete and \( U \) continuous. It remains to find a solution-set for each \( X \in \text{Set} \). Consider any function \( f: X \to UG \) for \( G \) a group, and take the subgroup \( S \) of \( G \) generated by all elements \( f x \), for \( x \in X \). Every element of \( S \) is then a finite product, say \((f x_1)^{a_1} (f x_2)^{a_2} \cdots (f x_n)^{a_n}\), of these generators and their inverses so the cardinal number of \( S \) is bounded, given \( X \). Taking one copy of each isomorphism class of such groups \( S \) then gives a small set of groups, and the set of all functions \( X \to US \) is then a solution set.

This left adjoint \( F: \text{Set} \to \text{Grp} \) assigns to each set \( X \) the free group \( FX \) generated by \( X \), so our theorem has produced this free group without entering into the usual (rather fussy) explicit construction of the elements of \( FX \) as equivalence classes of words in letters of \( X \). To be sure, the usual construction also shows that the universal arrow \( X \to UXF \) is injective (different elements of \( X \) are different as generators of the free group). However, we can also obtain this fact by general arguments and the observation that there does exist a group \( H \) with two different elements \( h \neq k \). Indeed, for any two elements \( x \neq y \) in \( X \) we then take a function \( f: X \to UH \) with \( f x = h \) and \( f y = k \). Since \( f \) must factor through the universal \( X \to UXF \), it follows that this universal must be an injection.
This construction applies not just to \textbf{Grp} but to the category of all small algebraic systems of a given type \( \tau \). The type \( \tau \) of an algebraic system is given by a set \( \Omega \) of operators and a set \( E \) of identities. The set \( \Omega \) of operators is a graded set; that is, a set \( \Omega \) with a function which assigns to each element \( \omega \in \Omega \) a natural number \( n \), called the arity of \( \omega \). Thus an operator \( \omega \) of arity 2 is a binary operator, one of arity 3 a ternary operator, and so on. If \( \Omega \) is any set, an action of \( \Omega \) on \( S \) is a function \( A \) which assigns to each operator \( \omega \) of arity \( n \) an \( n \)-ary operation \( \omega_A : S^n \to S \) (Here \( S^n = S \times \cdots \times S \), with \( n \) factors). From the given operators \( \Omega \) one forms the set \( A \) of all "derived" operators; given \( \omega \) of arity \( n \) and \( n \) derived operators \( \lambda_1, \ldots, \lambda_n \) of arities \( m_1, \ldots, m_n \), the evident "composite" \( \omega \lambda_1 \cdots \lambda_n \) is a derived operator of arity \( m_1 + \cdots + m_n \); also, given \( \lambda \) of arity \( m \) and \( f : n \to m \) any function from \( \{1, \ldots, n\} \) to \( \{1, \ldots, m\} \), "substitution" of \( f \) in \( \lambda \) gives an derived operator \( \theta \) of arity \( m \), described in terms of variables \( x_i \) as \( \theta(x_{f(1)}, \ldots, x_{f(n)}) \). (This description by variables refers implicitly to the action of \( \Omega \) on a set; for the abstract formulation of this and of composition, we refer to the standard treatments of universal algebra: Cohn [1965], or Grätzer [1968].) At any rate, each action \( A \) of \( \Omega \) on a set \( S \) extends uniquely to an action of the set \( A \) of derived operators on \( S \).

The set \( E \) of identities for algebraic systems of type \( \tau \) is a set of ordered pairs \( \langle \lambda, \mu \rangle \) of derived operators, where \( \lambda \) and \( \mu \) have the same arity \( n \). An action \( A \) of \( \Omega \) on \( S \) satisfies the identity \( \langle \lambda, \mu \rangle \) if \( \lambda_A \circ \mu_A = \mu_A \circ \lambda_A : S^n \to S \). An algebra \( A \) of type \( \tau \) is an \( \langle \Omega, E \rangle \)-algebra is a set \( S \) together with an action \( A \) of \( \Omega \) on \( S \) which satisfies all the identities of \( E \); so we call \( S \) the underlying set of the algebra and often write \( |A| = S \). A morphism \( g : A \to A' \) of \( \langle \Omega, E \rangle \)-algebras is a function \( g : S \to S' \) on the underlying sets which preserve all the operators of \( \Omega \) (and hence of \( A \)) in the sense that

\[ g \omega_A(a_1, \ldots, a_n) = \omega_{A'}(g(a_1), \ldots, g(a_n)) \]

for all \( a_i \in A \). The collection of all small \( \langle \Omega, E \rangle \)-algebras, with these morphisms as arrows, is a category \( \langle \Omega, E \rangle \to \text{Alg} \), often called a variety or an equational class of algebras. This description includes the familiar cases such as \textbf{Grp}, \textbf{Rng}, \textbf{Ab} and many others less familiar (e.g. nilpotent groups of specified class). For example, to describe \textbf{Grp}, take three operators in \( \Omega \), the product, the inverse, and the assignment of the identity element \( e \), of arities 2, 1, and 0, respectively, and take in \( E \) the axioms for the identity (\( ex = x = xe \)), the axioms for the inverse (\( xx^{-1} = e = x^{-1}x \)), and the associative law.

For any variety of algebras, the adjoint functor theorem will yield a left adjoint for the forgetful functor \( \langle \Omega, E \rangle \to \text{Alg} \to \text{Set} \); the solution set is obtained just as in the case of groups (see also §7 below). Thus this theorem produces for any set \( X \) the free ring, the free abelian group, the free \( R \)-module, etc. generated by the elements of the given set \( X \).

It does not produce free fields: In defining a field, the inverse to multiplication is not everywhere defined, so fields are not algebraic systems in the sense considered (and, for that matter, free fields do not exist).

Another illustration of the adjoint functor theorem is the construction of the left adjoint to

\[ V: \text{Comp Haus} \to \text{Set} \]

the forgetful functor which sends each compact Hausdorff space to the set of all its points. Given compact Hausdorff spaces \( X_i \), the usual product topology on the cartesian product set \( Y = \prod_i V X_i \) is Hausdorff and compact (the latter by the Tychonoff theorem); hence \textbf{Comp Haus} has all small products and \( V \) preserves them. For that matter, \( V \) creates these products: The product topology is chosen with the fewest open sets to make all the projections \( p_i : Y \to X_i \), continuous, so any other compact topology \( Y' \) with all \( p_i \) continuous would be the same set \( Y \) topologized with more open sets; then \( id: Y \to Y' \) is a continuous injection from a compact to a Hausdorff space, hence an isomorphism. By a similar argument, \( V \) creates all equalizers, hence all small limits. It remains to find for each set \( S \) a solution set of arrows \( f : S \to V X \) where each \( X \) is compact Hausdorff. Since \( X \) may be replaced by the closure \( fS \subseteq X \), it is enough to assume \( fS \) dense in \( X \). To each point \( x \in X \), consider the set \( Lx = \{ D | D \subseteq S \text{ and } x \in fD \} \); thus \( Lx \) is a non-void set of subsets of \( S \). If \( x \neq x' \) are separated in \( X \) by disjoint open sets \( U \) and \( U' \), then \( f^{-1}U \cap Lx = \emptyset \) and \( f^{-1}U \cap Lx' = \emptyset \). Thus \( Lx \) is an injection \( X \to \text{Pow} S \) from \( X \) to the double power set of \( S \). If we take all subsets \( X \subseteq \text{Pow} S \), all topologies on each set \( X \) and all functions \( f : S \to V X \) we obtain a small solution set for \( S \). The adjoint functor theorem then provides a left adjoint to \( V \); it assigns to each set \( S \) the Stone-Cech compactification of the discrete topology on \( S \).

\[ \text{Exercises} \]

1. For \( G : A \to X \) continuous, show that the projection \( (x \downarrow) G \to A \) creates all small limits.

2. Use the adjoint functor theorem to find a left adjoint to each of the forgetful functors \( \text{Rng} \to \text{Set}, \text{Rng} \to \text{Ab}, \text{Cat} \to \text{Grp} \). Compare with the standard explicit construction of these adjoints.

3. Given a pullback diagram in \( \text{Cat} \),

\[ \begin{array}{ccc}
A' & \longrightarrow & A \\
\downarrow & & \downarrow \downarrow \\
X' & \rightarrow & X
\end{array} \]

if \( H \) creates limits and \( G \) preserves them prove that \( H' \) creates them.

4. Use Exercise 3 and the fact that \( (x \downarrow) X \to X \) creates limits to give a new proof of the result of Exercise 1.
7. Subobjects and Generators

Concepts such as subring, subspace, and subfield will now be treated categorically, using arrows instead of elements. For instance we will regard a subgroup \( S \) of a group \( G \) not as a set of elements of \( G \), but as the monomorphism \( S \to G \) given by insertion.

Let \( A \) be any category. If \( u : s \to a \) and \( v : t \to a \) are two monics with a common codomain \( a \), write \( u \leq v \) when \( u \) factors through \( v \); that is, when \( u = uv' \) for some arrow \( u' \) (which is then necessarily also monic). When both \( u \leq v \) and \( v \leq u \), write \( u \equiv v \); this defines an equivalence relation \( \equiv \) among the monics with codomain \( a \), and the corresponding equivalence classes of these monics are called the subobjects of \( a \). It is often convenient to say that a monic \( u : s \to a \) is a subobject of \( a \) – that is, to identify \( u \) with the equivalence class of all \( v = u \theta \), for \( \theta : s' \to s \) an invertible arrow. These subobjects do correspond to the usual subobjects (defined via elements) in familiar large categories such as \( \text{Rng} \), \( \text{Grp} \), \( \text{Ab} \), and \( \text{R-Mod} \), but not in \( \text{Top} \).

**Lemma.** In any square pullback diagram

\[
\begin{array}{ccc}
\ast & \xrightarrow{k} & p \\
\downarrow{s} & & \downarrow{f} \\
\ast & \xrightarrow{s} & a
\end{array}
\]

\( f \) monic implies \( f' \) monic (and \( g \) monic implies \( g' \) monic).

Briefly, pullbacks of monics are monic.

**Proof.** Consider a parallel pair \( h, k \), as shown, with \( f'h = f'k \). Then \( g f'k = g f'k \), so \( f g'k \). Since \( f \) is monic, this gives \( g h = g k \). But we also have \( f' h = f'k \); these two equations, since \( p \) is a pullback, imply \( h = k \).

The set of all subobjects of each \( a \in A \) is partly ordered by the binary relation \( u \leq v \). If \( u : s \to a \) and \( v : t \to a \) are two subobjects of \( a \), and \( A \) has pullbacks, the pullback of these two arrows gives (Lemma above) another monic \( w : p \to a \) with codomain \( a \) and with \( w \leq u, w \leq v \); it is the intersection (= meet or greatest lower bound) of the subobjects \( u \) and \( v \) in the partly ordered set of all subobjects of \( a \in A \). Similarly, if \( J \) is any set and \( u : s \to a \) for \( i \in J \) any \( J \)-indexed set of subobjects of \( a \in A \), the pullback of all these arrows, if it exists, gives the intersection of the subobjects \( u_i \) of \( a \). The union (= join or least upper bound) of subobjects can be found under added hypotheses.

Dually, two epis \( r, s \) with domain \( a \) are equivalent when \( r = \theta s \) for some invertible \( \theta \). The equivalence classes of such epis are the quotient objects of \( a \), partly ordered by the relation \( r \leq s \), which holds when \( r \) factors through \( s \) as \( r = r s \). This definition of quotients by duality is simpler than the usual definition of quotient algebras by equivalence classes, and agrees with the usual definition in those categories where epis are onto. This latter is the case, for example, in \( \text{Grp} \). Hence every quotient object of a group \( G \) in \( \text{Grp} \) is represented by the projection \( p : G \to G/N \) of \( G \) onto the factor group \( G/N \) of \( G \) by some normal subgroup \( N \) of \( G \), and \( G/M \leq G/N \) holds if and only if \( M \subseteq N \) (in general, the relation \( r \leq s \) for quotients means that in \( r \) "more" is divided out!).

A set \( S \) of objects of the category \( C \) is said to generate \( C \) when to any parallel pair \( h, h' : c \to d \) of arrows of \( C \), \( h + h' \) implies that there is an \( s \in S \) and an arrow \( f : s \to c \) with \( h f + h' f \) (The term "generates" is well established but poorly chosen; "separates" would have been better). This definition includes the case of a single object \( s \) generating a category \( C \). For example, any one-point set generates \( \text{Set} \), \( \text{Z} \) generates \( \text{Ab} \) and \( \text{Grp} \), and \( R \) generates \( \text{R-Mod} \). The set of finite cyclic groups is a generator for the category of all finite abelian groups (or, of all torsion abelian groups).

Dually, a set \( Q \) of objects is a cogenerating set for the category \( C \) when to every parallel pair \( h, h' : b \to c \) of arrows of \( C \) there is an object \( q \in Q \) and an arrow \( g : b \to q \) with \( gh = gh' \). A single object \( q \) is a cogenerator when \( \{ q \} \) is a cogenerating set. For example, any two-point set is a cogenerator in \( \text{Set} \).

In terms of subobjects we can examine further the construction of solution sets. Given any functor \( G : A \to X \) an arrow \( f : x \to Ga \) is said to span \( a \) when there is no proper monomorphism \( s \to a \) in \( A \) such that \( f \) factors through \( G s \to Ga \).

**Lemma.** In the category \( A \), suppose that every set of subobjects of an object \( a \in A \) has a pullback. Then if \( G : A \to X \) preserves all these pullbacks, every arrow \( h : x \to Ga \) factors through an arrow \( f : x \to Gb \) which spans \( b \).

**Proof.** Consider the set of all those subobjects \( u_j : s_j \to a \) such that \( h \) factors through \( Gu_j \) as \( h = Gu_j \cdot h_j \). Take the pullback \( v : b \to a \) of all the \( u_j \).

Then, as in the diagrams

\[\begin{array}{ccc}
b & \xrightarrow{h} & Gb \\
\downarrow{v} & & \downarrow{g} \\
\ast & \xrightarrow{a} & \ast
\end{array}\]

\( Gv : Gb \to Ga \) is still a pullback (for the \( Gu_j \)), so \( h \) factors through \( Gv \) via \( f \), as shown. It follows from the construction that \( f \) spans \( b \).

This lemma states that a solution set for \( x \) can be the set of all arrows from \( x \) which span.
As an application consider the category of algebras of given type \( r \). Given an arrow \( f: S \to G A \), the algebra \( A \) has a subalgebra consisting of all elements obtained from elements of \( f(S) \) by iterated applications of operators \( u \in \Omega \). The cardinal number of this subalgebra \( A_f \) is then bounded by the cardinal of \( S \) and that of \( \Omega \). Since \( f \) factors through \( S \to G A_f \), these latter arrows from the set \( S \) form a small set which is a solution set for \( G: \text{Alg} \to \text{Set} \). They are spanning arrows in the sense of the lemma, provided a subobject of \( a \) is redefined to be a morphism \( u: S \to a \) for which \( G u \) is injective in \( \text{Set} \).

Another example of the use of \( \text{this} \) lemma with the adjoint functor theorem is the proof of the existence of tensor products of modules. Given modules \( A \) and \( B \) over a commutative ring \( K \), a tensor product is a universal element of the set \( \text{Bilin}(A, B; C) \) of bilinear functions \( \beta: A \times B \to C \) to a some third \( K \)-module \( C \). This is the (object function of) a functor of \( C \). To get a solution set for given \( A \) and \( B \), if \( A \) is small, then \( C \) is a small-complete category (small-complete category with arrows). Then \( C \) consists of all finite sums \( \sum \beta(a, b) \), so the solution set condition holds; since \( K \)-Mod is small-complete and \( \text{Bilin}: K \text{-Mod} \to \text{Set} \) is continuous, a tensor product \( \otimes: A \times B \to A \otimes B \) exists. The usual (more explicit) construction is wholly needless, since all the properties of the tensor product follow directly from the universality.

Exercises

1. Use the adjoint functor theorem to construct the coproduct in \( \text{Grp} \) (the coproduct \( \text{GLH} \) in \( \text{Grp} \) is usually called the free product). Using the product \( G \times H \), show also that the injections \( G \to \text{GLH} \) and \( H \to \text{GLH} \) of the coproduct are both monic, and that their images intersect in the identity subgroup.

2. Make a similar construction for the coproduct of rings.

3. If \( R \) is a ring, \( A \) a right \( R \)-module and \( B \) a left \( R \)-module, use the adjoint functor theorem to construct \( A \otimes R B \) (this tensor product is an abelian group, with a function \( (a, b) \to a \otimes b \in A \otimes R B \) which is bivariant, has \( a \otimes b = a \otimes q r b \) for all \( a \in A, r \in R, \) and \( b \in B \), and is universal with these properties). Prove that \( A \otimes R B \) is spanned (as an abelian group) by the elements \( a \otimes b \). If \( S \to R \) is a morphism of rings, examine the relation of \( A \otimes R B \) to \( A \otimes S B \).

4. Construct coequalizers in \( \text{Alg} \) by the adjoint functor theorem.

8. The Special Adjoint Functor Theorem

We now consider another existence theorem for adjoints which avoids the solution set condition by assuming a small set of objects which cogenerates.

**Theorem 1.** (Special Initial-Object Theorem). If the category \( D \) is small-complete, has small hom-sets, and a small cogenerating set \( Q \), then \( D \) has an initial object provided every set of subobjects of each \( d \in D \) has an intersection.

**Proof.** Form the product \( q_0 = \prod_{a \in D} q_a \) of all the objects in the small cogenerating set \( Q \) and take the intersection \( \cap \) of all subobjects of \( q_0 \). For any object \( d \in D \), there is at most one arrow \( r \to d \), for if there were two different arrows, their equalizer would be a proper monic to \( r \), hence a subobject of \( q_0 \) smaller than the intersection \( r \).

To show \( r \) initial in \( D \), we thus need only construct an arrow \( r \to d \) for each \( d \). Do consider the set \( H \) of all arrows \( h: d \to q \in Q \) and the (small) product \( \prod_{d \in D} q_d \). Take the arrow \( j: d \to \prod_{d \in D} q_d \) with components \( h \) (i.e., with \( p_h: j = h \) for each projection \( p_h \)). Since the \( Q \) cogenerates, \( j \) is monic. Form the pullback

\[
\begin{array}{ccc}
\ast & \to & \prod_{d \in D} q_d = q_0 \\
\downarrow & & \downarrow \\
& k & \to \\
\end{array}
\]

where \( k \) is the arrow with components \( p_h \circ k = p_h \) for each \( h: d \to q \). Then \( j \) is pullback of a monic \( j \), so \( c \) is a subobject of \( q_0 \). But \( r \) was the intersection of all subobjects of \( q_0 \), so there is an arrow \( r \to c \). The composite \( r \to c \to d \) is the desired arrow.

**Theorem 2.** (The Special Adjoint Functor Theorem). Let the category \( A \) be small-complete, with small hom-sets, and a small cogenerating set \( Q \), while every set of subobjects of an object \( a \in A \) has a pullback (and hence has an intersection). Let the category \( X \) have small hom-sets. Then a functor \( G: A \to X \) has a left-adjoint if and only if \( G \) preserves all small limits and all pullbacks of families of monics.

**Proof.** The conditions are necessary, since any right adjoint functor must indeed preserve all limits (in particular, all pullbacks). Conversey, it suffices as usual to construct for each \( x \in X \) an initial object in the comma category \( D = (x \downarrow G) \). We shall show that this category satisfies the hypotheses of the previous theorem for the construction of an initial object. First we verify that subobjects in \( (x \downarrow G) \) have the expected form.

**Lemma.** An arrow \( h: (f: x \to Ga, a) \to (f': x \to Ga', a') \) in the comma category \( (x \downarrow G) \) is monic if and only if \( h: a \to a' \) is monic in \( A \).

**Proof.** Trivially, \( h: a \to a' \) monic implies \( h: f \to f' \) monic. For the converse, observe that \( h \) monic means exactly that its kernel pair (the pullback of \( h \) with \( h \)) is \( 1_a, 1_a: a \to a \). On the other hand, by the lemma of \( \S 6 \) the projection

\[
(x \downarrow G) \to A, \quad (f: x \to Ga, a) \to a
\]

of the comma category creates all limits, and in particular, creates kernel pairs. Moreover, \( A \) has all kernel pairs. Therefore (Theorem 4.2), the
projection of the comma category preserves all kernel pairs, in particular, the kernel pair \(1_a, 1_a\), and consequently carries monics (in \((x \downarrow G)\)) to monics in \(A\), as desired.

Now return to the theorem. We are given a small cogenerating set \(Q\) in \(A\). Since \(X\) has small hom-sets, the set \(Q\) of all objects \(k: x \to Gq\) with \(q \in Q\) is small. It is, moreover, cogenerating in \((x \downarrow G)\). Given \(s \neq t\), \(f: x \to Ga, a) \mapsto (f: x \to Ga', a')\) in \((x \downarrow G)\), there is a \(q_0 \in Q\) and an arrow \(h: a' \to q_0\) with \(hs \neq ht\), and this \(h\) can be regarded as an arrow

\[
h: (f': x \to Ga', a') \mapsto (f_0: x \to Gq_0, q_0),
\]

where \(f_0 = Gh \circ f'\), with \(hs \neq ht\) in \((x \downarrow G)\). Therefore \(Q\) cogenerates \((x \downarrow G)\).

Since \(A\) is small-complete and \(G\) is continuous, \((x \downarrow G)\) small-complete remains only to construct an intersection in \((x \downarrow G)\) for every set of subobjects \(h_i: \langle f_i: x \to Ga, a_i \rangle \mapsto \langle f: x \to Ga, a \rangle\), where \(i \in J\). By the lemma, the corresponding arrows \(h_i: a_i \to a\) are monics in \(A\). By hypothesis, they then have a pullback \(h: b \to a\) in \(A\)

\[
\begin{array}{ccc}
a & \xrightarrow{f} & b \\
\downarrow{h} & & \downarrow{h} \\
G & \xrightarrow{Gf} & Ga
\end{array}
\]

The functor \(G\) preserves pullbacks, so \(Gh: Gb \to Ga\) with \(Gh = Gh \circ Gs\) is a pullback of the \(Gh\) in \(X\). Since also \(Gh_i = f_i \quad \text{for all } i \in J\), there is a unique \(f_0: x \to Gb\) with \(f_0 = Gs \circ f_0\); the resulting arrow \(h_i: (x, f_0, b) \mapsto (Gf, a)\) is then a pullback in \((x \downarrow G)\) of the given \(h_i\) (again, because the projection of the comma category creates pullbacks). This pullback is the required intersection of the \(h_i\).

There is another form of this theorem. Define a category to be \textit{well-powered} when the subobjects of each object \(a \in A\) can be indexed by a small set; that is, when there is to each \(a\) a small set \(J_a\) and a bijection from \(J_a\) to the set of all subobjects of \(a\). Many familiar large categories — \(\text{Top}, \text{Grp}, \text{R-Mod}, \text{etc.}\) — are well-powered; the dual notion is called \textit{co-well-powered}. If \(A\) is well-powered and small-complete, then any set of subobjects of an \(a \in A\) has an intersection, formed by the usual pullback. Therefore the special adjoint functor theorem specializes as follows:

\textbf{Corollary.} If \(A\) is small complete, well-powered, with small hom-sets, and a small cogenerating set, while \(X\) has small hom-sets, then a functor \(G: A \to X\) has a left adjoint if and only if it is continuous. In particular, any continuous \(K: A \to \text{Set}\) is representable.

This classical form of the special adjoint functor theorem (sometimes called \(SAFT\)) often appears without an explicit "small hom-set" hypothesis — in sources which consider only categories with small hom-sets. Some authors use "locally small category" to mean "well-powered"; others use it to mean "has small hom-sets", so we avoid this term!

The classical form of \(SAFT\) can be deduced directly from the adjoint functor theorem by constructing a solution set (as in Freyd [1964, p. 89], or Schubert [1970, p. 88]).

A typical example is the inclusion functor

\[
G: \text{Comp Haus} \subseteq \text{Top}
\]

of the full subcategory of compact Hausdorff spaces in \(\text{Top}\). As already noted, \(\text{Comp Haus}\) is small complete; it also has small hom-sets. The Urysohn lemma states that to any two points \(x \neq y\) in a compact Hausdorff space \(X\) there is a continuous function \(f: X \to I\) to the unit interval \(I\) with \(fx = 0, fy = 1\). It follows that \(I\) is a cogenerator for \(\text{Comp Haus}\).

Hence the special adjoint functor theorem gives a left adjoint for the inclusion \(G\) above. This left-adjoint (or sometimes, its restriction to the full subcategory of completely regular spaces) is called the \textit{Stone-Čech compactification}. This includes the case of a discrete space, as done in §6.

Watt's Theorem [1960] is another example. Any ring \(R\) is a generator in the category \(\text{R-Mod}\), hence a cogenerator in \((\text{R-Mod})^\text{op}\). It follows that any contravariant additive functor \(T\) on \(\text{R-Mod}\) to \(\text{Ab}\) which takes small colimits to limits is representable by a group isomorphism \(T \cong \text{hom}_A(\_ , C)\) for some \(\text{R-module} C\). Indeed, by the special adjoin functor theorem \(T: (\text{R-Mod})^\text{op} \to \text{Ab}\) has a left adjoint \(F\); since \(T\) is additive, the adjunction

\[
\text{Ab}(G, TA) \cong \text{hom}_A(A, FG), \quad G \in \text{Ab}, \quad A \in \text{R-Mod},
\]

is an isomorphism of additive groups; set \(G = \text{Z}\) to get

\[
TA \cong \text{Ab}(Z, TA) \cong \text{hom}_A(A, FZ).
\]

\textbf{Exercises}

1. Let \(K: A \to \text{Set}\) be any functor. If \(K\) has a left adjoint, prove that it is representable. Conversely, if \(A\) has all small copowers and \(K\) is representable as \(K \cong A(a, \_ )\) for some \(a \in A\) prove that \(K\) has a left adjoint (which assigns to each set \(X\) the small copower \(X \cdot a\)).

2. For \(A\) a left \(\text{R-module}, B\) a right \(\text{R-module}\) and \(G\) an abelian group, establish adjunctions

(a) \(\text{hom}_B(A, \text{hom}_G(B, G)) \cong \text{hom}_B(B \otimes_G A, G) \cong \text{hom}_G(B, \text{hom}_G(A, G))\), where \(\text{hom}_G(B, G)\) has a suitable (left or right) \(\text{R-module}\) structure, and where hom\(_G\) denotes the hom-set in \(\text{R-Mod}\), hom\(_G\) that in \(\text{Ab}\).

(b) The additive group \(\text{Q/Z}\) of rational numbers modulo 1 is known to be an injective cogenerator of \(\text{Ab}\). Use \(\text{a)\) to prove that \(\text{hom}_G(R, \text{Q/Z})\) is an injective cogenerator of \(\text{R-Mod}\) ("injective" object as defined in §4).
3. Use Exercise 2(b) and the special adjoint functor theorem to prove that any continuous additive functor \( T: R\text{-Mod} \to Ab \) is representable. (Watt's theorem).

4. (Stone-Cech compactification.) If \( X \) is a completely regular topological space, show that the universal arrow \( X \to G\!X \) for the left adjoint to (1) is an injection. (Use the Urysohn lemma: For \( x \neq y \) in \( X \) completely regular there exists a function \( f: X \to t \) with \( f(x) \neq f(y) \) and \( t \) the unit interval.)

Classical sources describe this compactification only when \( X \) is completely regular. This restriction is needless: it arose from the idea of considering just universal injections, not universal arrows.

9. Adjoins in Topology

\( \text{Top} \) is the category with objects all (small) topological spaces \( X, Y, \ldots \) and arrows all continuous maps \( f: X \to Y \). The standard forgetful functor (usually a nameless orphan!)

\[ G: \text{Top} \to \text{Set} \]

sends \( X \) to \( G\!X \), the set of points in \( X \), is faithful, and has a left adjoint \( D \) which assigns to each set \( S \) the discrete topology on \( S \) (i.e., all subsets of \( S \) are open). Therefore \( G \) preserves all limits which may exist in \( \text{Top} \) (this is why the underlying set of the product of spaces is the cartesian product of their underlying sets). The forgetful functor \( G \) also has a right adjoint \( D' \), which assigns to each set \( S \) the indiscrete topology on \( S \) (with only \( S \) and \( \emptyset \) open). Therefore \( G \) preserves all colimits which may exist in \( \text{Top} \) — and this is why the coproduct of two spaces is formed by putting a topology on the disjoint union of the underlying sets.

Next consider the subspace topology on a set \( S \subseteq G\!X \).

If \( X \) is a fixed topological space, \( G \) induces a functor

\[ G \downarrow X : (\text{Top} \downarrow X) \to (\text{Set} \downarrow G\!X) \]

\[ Y \xrightarrow{f} X \xrightarrow{G} G\!X \xrightarrow{Gf} G\!X \]

\[ \downarrow \quad \downarrow \quad \downarrow \]

\[ Y' \xrightarrow{f'} X \xrightarrow{G} G\!X \xrightarrow{Gf'} G\!X \]

Here \( f \) and \( f' \) are objects and \( h \) an arrow of the comma category \((\text{Top} \downarrow X)\). This functor \( G \downarrow X \) has a right adjoint \( L \). Indeed, an object \( t: S \to G\!X \) in \((\text{Set} \downarrow G\!X)\) is a set \( S \) and a function \( t \) on \( S \) to \( G\!X \). Put on \( S \) the topology with open sets all \( t^{-1} U \) for \( U \) open in \( X \), and call the resulting space \( L\!S \); then \( t \) is a continuous map \( L\!t: L\!S \to X \). (For example, if \( S \) is a subset of \( G\!X \), then \( L\!S \) is just \( S \) with the usual "subspace topology").

This topology on \( L\!S \) has the familiar universal property: Any continuous map \( f: Y \to X \) which factors through \( t \) as \( G\!f = t \circ s \), in \( \text{Set} \),

\[ G\!Y \xrightarrow{Gf} G\!X \]

\[ \downarrow \quad \downarrow \quad \downarrow \]

\[ G\!Y \xrightarrow{Gf} G\!X \]

has \( s: Y \to L\!S \) continuous. This property just restates the desired adjunction: \( \text{hom}(G\!f, t) \cong \text{hom}(f, L\!t) \). Observe that \((G \downarrow X) \cdot L = \text{Id}: L \) is a "right-adjoint-right-inverse" to \((G \downarrow X)\).

Note especially that the universal property of the subspace topology on a subset \( S \subseteq G\!X \) refers not only to the other subspaces of \( X \), but to other spaces \( Y \) and any continuous \( f: Y \to X \) which factors through the inclusion \( t: S \to G\!X \) (i.e., has image contained in the subset \( S \)).

This adjoint may be used to construct (the usual) equalizers in \( \text{Top} \) by the following general process:

**Proposition 1.** If \( G: C \to D \) is a faithful functor, if \( D \) has equalizers, and \( f, f' \), for each \( x \in C \), \((G \downarrow X) \cdot (C \downarrow x) \to (D \downarrow G\!x) \) has a right-adjoint-right-inverse \( L \), then \( C \) has equalizers.

**Proof.** To get the equalizer of a parallel pair \( f, f' : x \to y \), apply \( G \), take the equalizer \( t: s \to G\!x \) of \( G\!f, G\!f' \) in \( D \) and apply \( L \); the universal property of the adjunction shows \( L\!t: L\!s \to x \) an equalizer in \( C \).

This argument is just an element-free version of the usual definition of the equalizer: Given two continuous maps \( f, f': X \to Y \), take the set \( S \) of points \( x \) of \( X \) with \( f(x) = f'(x) \) and impose the subspace topology. The adjunction explains why the subspace topology.

Now \( \text{Top} \) is well known to be complete: To prove this one needs only equalizers (of parallel pairs) and products. The product of any family \( X_i, i \in J \), of spaces is constructed by taking the product \( \Pi G\!X_i \) of the underlying sets and putting on it the (universal) topology in which all projections \( p_i: \Pi G\!X_i \to G\!X_i, i \in J \), are continuous. The general fact that to spaces \( X_i \), a set \( S \), and functions \( t_i: S \to G\!X_i \), there is a "universal" topology with exactly those open sets on \( S \) required to make all \( t_i \) continuous can be expressed categorically (Exercise 3).

Colimits may be treated in dual fashion. For any space \( X \) the functor

\[ (X \downarrow G): (X \downarrow \text{Top}) \to (G\!X \downarrow \text{Set}) \]

has a left adjoint \( M \). Indeed, an object of \((G\!X \downarrow \text{Set})\) is a function \( t: G\!X \to S \) to a set \( S \). Put on \( S \) the topology with open sets all subsets \( U \subseteq S \) with \( t^{-1} U \) open in \( X \) and call the resulting space \( M\!S \). (If \( t: G\!X \to S \) is a surjection, this is the familiar "quotient topology" or "identification topology" on \( S \).) Then the function \( t \) is a continuous map \( M\!t: X \to M\!S \).
Moreover, \( f : X \to Y \) continuous and \( Gf = k \circ t \) for some function \( k \),

\[
\begin{array}{ccc}
\text{Top} & \xymatrix{
X \ar[r]^f & Y, & GX \ar[r]^{Gf} & Gy, \\
M & \ar[l]_M S & \ar[l]_S \\
} \\
\end{array}
\]

implies that \( k : Ms \to Y \) is continuous. Thus \( k = k \) is an adjunction

\[
(X \downarrow \text{Top}) (M, f) \cong (GX \downarrow \text{Set}) (t, Gf)
\]

with unit the identity map, so \( M \) is left-adjoint-right-inverse to \( X \downarrow G \).

Now Proposition 1 was proved just from the axioms for a category, so its dual is also true. This dual proposition and the above adjunction prove that \( \text{Top} \) has coequalizers.

Similar constructions yield coproducts (= disjoint unions) and general colimits in \( \text{Top} \). Such colimits appear often, usually under other names, as for instance in the basic process of constructing spaces by glueing pieces together. For example, let \( \{U_i \mid i \in I\} \) be an open cover of a space \( X \). Each continuous \( f : X \to Y \) determines a \( I \)-indexed family of restrictions \( f_i : U_i \to Y \); conversely, a familiar result states that a \( I \)-indexed family of continuous maps \( f_i : U_i \to Y \) determines a map \( f \) continuous on all of \( X \) if and only if \( f_i ([U_i \cap U_j] = f_j ([U_i \cap U_j]) \) for all \( i \) and \( j \). This result may be expressed by the statement that the following diagram is an equalizer

\[
\begin{array}{ccc}
\text{Top}(X, Y) & \to & \coprod_{i} \text{Top}(U_i, Y) \\
& \ar[l] \ar[r] & \coprod_{i,j} \text{Top}(U_i \cap U_j, Y) \\
\end{array}
\]

where the arrows are given by restriction, as above. This result may equally well be expressed by the statement that \( X \) is the colimit in \( \text{Top} \), with colimiting cone the inclusion maps \( U_i \to X \), of the functor \( U : J \to \text{Top} \), where \( J \) is the category with objects the pairs of indices \( \langle i, j \rangle \), the single indices \( \langle i \rangle \), and the (non-identity) maps \( \langle i,j \rangle \to \langle i \rangle, \langle i,j \rangle \to \langle j \rangle \), while \( U \) is the functor with \( U \langle i, j \rangle = U_i \cap U_j, U \langle i \rangle = U_i \) with \( U \) on (non-identity) arrows the inclusion maps.

Another coequalizer is the space \( X/A \) obtained from the space \( X \) by **collapsing the subset \( A \) to a point.** It is the coequalizer

\[
\begin{array}{ccc}
\ast & \to & X \\
\downarrow & & \downarrow \\
\ast & \to & X/A \\
\end{array}
\]

of the set of all the arrows sending the one point space \( \ast \) to one of the points \( a \in A \). It is used in homotopy theory. If we consider the category \( \text{Top}^{(1)} \) whose objects are pairs \( \langle X, A \rangle \) (a space \( X \) with a subset \( A \)) and whose arrows \( \langle X, A \rangle \to \langle X', A' \rangle \) are continuous maps \( X \to X' \) sending

\[
A \to A',
\]

then the definition of \( X/A \), for \( A \) a pointed topological space, reads:

\[
\text{Top}_c(X/A, Y) = \text{Top}^2((X, A), (Y, \ast)).
\]

Thus \( \langle X, A \rangle \to X/A \) is left adjoint to the functor \( Y \to (Y, \ast) \) which sends each pointed space to the pair \( (Y, \ast) \).

There are many familiar subcategories of \( \text{Top} \).

**Proposition 2. Haus.** The full subcategory of all Hausdorff spaces in \( \text{Top} \) is complete and cocomplete. The inclusion functor \( \text{Haus} \to \text{Top} \) has a left adjoint \( H \), as does the forgetful functor \( \text{Haus} \to \text{Set} \).

**Proof.** The left adjoint \( H \) will be obtained by the adjunction theorem. First, any product of Hausdorff spaces or subspace of a Hausdorff space is also Hausdorff, hence \( \text{Haus} \) is complete and the inclusion functor is continuous (i.e., it preserves small limits). It remains only to verify the solution set condition for every topological space \( X \). But any continuous map of \( X \) to a Hausdorff space \( Y \) factors through the image, a subspace of \( Y \), hence Hausdorff. This image is a quotient set of \( X \) with some topology, so there is at most a small set of (non-isomorphic) surjections \( X \to Y \) to a Hausdorff \( Y \). This is the solution set condition.

The resulting left adjoint \( H \) assigns to each space \( X \) a Hausdorff space \( HX \) and a continuous map \( \eta : X \to HX \), universal from \( X \) to a Hausdorff space. Now \( \eta \) universal implies that \( \eta \) is a surjection, so \( HX \) may be described as the "largest Hausdorff quotient" of \( X \). If \( X \) is already Hausdorff, we may take \( HX = X \) and \( \eta = 1 \), so \( H \) is a left-adjoint-left-inverse to the inclusion.

Since \( H \) is a left adjoint, it preserves colimits. It follows that \( \text{Haus} \) has all small colimits (is cocomplete). In particular, the coproduct in \( \text{Haus} \) is the coproduct in \( \text{Top} \) (because a coproduct of Hausdorff spaces is Hausdorff), while a coequalizer in \( \text{Haus} \) is the largest Hausdorff quotient of the coequalizer in \( \text{Top} \).

The full subcategory of compactly generated Hausdorff spaces is especially convenient because it is cartesian closed (§ VII.8).

**Exercises**

1. For the full subcategory \( \text{L con} \) of locally connected spaces in \( \text{Top} \), prove that \( D : \text{Set} \to \text{L con} \) has a left adjoint \( C \), assigning to each space \( X \) the set \( X \) of its connected components, but show that this functor \( C \) can have no left adjoint (because of misbehavior on equalizers).

2. Show that the right adjoint \( D : \text{Set} \to \text{Top} \) for the forgetful functor \( D \) has no right adjoint (misbehavior on coproducts).

3. (Categorical construction of the usual products in \( \text{Top} \).)
   (a) For diagonal functors \( C : C \to C', D : C \to D', \) and \( T \in C \), each \( G : C \to D \) defines \( G_* : (C \downarrow T) \to (D' \downarrow GT) \) by \( (c : T \to g) \mapsto (G_* : Gc \to GT) \). If \( G_* \) has a left adjoint and \( GT \) a limit in \( D \), prove that \( T \) has a limit in \( C \).
VI. Monads and Algebras

In this chapter we will examine more closely the relation between universal algebra and adjoint functors. For each type $\tau$ of algebras, we have the category $\text{Alg}$, of all algebras of the given type, the forgetful functor $G: \text{Alg} \rightarrow \text{Set}$, and its left adjoint $F$, which assigns to each set $S$ the free algebra $FS$ of type $\tau$ generated by elements of $S$. A trace of this adjunction $\langle F, G, \phi \rangle: \text{Set} \rightarrow \text{Alg}$ resides in the category $\text{Set}$; indeed, the composite $T = GF$ is a functor $\text{Set} \rightarrow \text{Set}$, which assigns to each set $S$ the set of all elements of $S$ and the corresponding free algebra. Moreover, this functor $T$ is equipped with certain natural transformations which give it a monoid-like structure, called a "monad". The remarkable part is that then the whole category $\text{Alg}$, can be reconstructed from this monad in $\text{Set}$.

Another principal result is a theorem due to Beck, which describes exactly those categories $A$ with adjunctions $\langle F, G, \phi \rangle: X \rightarrow A$ which can be reconstructed from a monad $T$ in the base category $X$. It then turns out that algebras in this last sense are so general as to include the compact Hausdorff spaces (§9).

1. Monads in a Category

Any endofunctor $T: X \rightarrow X$ has composites $T^2 = T \circ T: X \rightarrow X$ and $T^3 = T^2 \circ T: X \rightarrow X$. If $\mu: T^2 \Rightarrow T$ is a natural transformation, with components $\mu_x: T^2 x \rightarrow T x$ for each $x \in X$, then $T \mu: T^3 \Rightarrow T^2$ denotes the natural transformation with components $(T\mu)_x = T(\mu_x): T^3 x \rightarrow T^2 x$ while $\mu T: T^2 \Rightarrow T^3$ has components $\mu T_x = \mu_{T x}$. Indeed, $T \mu$ and $\mu T$ are "horizontal" composites in the sense of §11.5.

**Definition.** A monad $T = \langle T, \eta, \mu \rangle$ in a category $X$ consists of a functor $T: X \rightarrow X$ and two natural transformations

\[ \eta: 1_X \Rightarrow T \quad \mu: T^2 \Rightarrow T \]

which make the following diagrams commute

\[ \begin{array}{ccc}
T^3 & \xrightarrow{T \eta} & T^2 \\
\mu T & \Downarrow \mu & \Downarrow \mu \\
T^2 & \xrightarrow{\mu} & T
\end{array} \quad
\begin{array}{ccc}
T^2 & \xrightarrow{T \eta} & T^2 \\
\mu T & \Downarrow \mu & \Downarrow \mu \\
T & \xrightarrow{\mu} & T
\end{array} \]

This completes the definition.
Formally, the definition of a monad is like that of a monoid \( M \) in sets, as described in the introduction. The set \( M \) of elements of the monoid is replaced by the endofunctor \( T: X \to X \), while the cartesian product \( \times \) of two sets is replaced by composite of two functors, the binary operation \( \mu: M \times M \to M \) of multiplication by the transformation \( \mu: T \to T \) and the unit (identity) element \( \eta: 1 \to M \) by \( \eta: I \to T \). We shall thus call \( \eta \) the unit and \( \mu \) the multiplication of the monad \( T \); the first commutative diagram of (2) is then the associative law for the monad, while the second and third diagrams express the left and right unit laws, respectively. All told, a monad in \( X \) is just a monoid in the category of endofunctors of \( X \), with product \( \times \) replaced by composition of endofunctors and unit set by the identity endofunctor.

**Terminology.** These objects \( \langle X, T, \eta, \mu \rangle \) have been variously called “dual standard construction”, “triple”, “monoid”, and “triad”. The frequent but unfortunate use of the word “triple” in this sense has achieved a maximum of needless confusion, what with the conflict with ordered triple, plus the use of associated terms such as “triple derived functors” for functors which are not three times derived from anything in the world. Hence the term **monad**.

Every adjunction \( \langle F, G, \eta, \epsilon \rangle: X \to A \) gives rise to a monad in the category \( X \). Specifically, the two functors \( F: X \to A \) and \( G: A \to X \) have composite \( FGF \) an endofunctor, the unit \( \eta \) of the adjunction is a natural transformation \( \eta: Id \to T \) and the counit \( \epsilon: FG \to Id \) of the adjunction yields by horizontal composition a natural transformation \( \mu: G \circ F \to T \). The associative law of (2) above for this \( \mu \) becomes the commutativity of the first diagram below

\[
\begin{array}{ccc}
GF & GF & GF \\
\downarrow \eta & \downarrow \mu & \downarrow \mu \\
GF & F & F \\
\end{array}
\]

Dropping \( G \) in front and \( F \) behind, this amounts to the commutativity of the second diagram, which holds by the very definition (§ II.4) of the (horizontal) composite \( \epsilon \circ \eta = \epsilon \cdot (FG) = \epsilon \cdot (F \epsilon G) \) (i.e., by the “interchange law” for functors and natural transformations). Similarly, the left and right unit laws of (2) reduce to the diagrams

\[
\begin{array}{ccc}
I_x & GF & GF \\
\downarrow \epsilon & \downarrow \mu & \downarrow \mu \\
GF & F & F \\
\end{array}
\]

which are essentially just the two triangular identities

\[
1 = G \epsilon \cdot \eta G: G \to G \quad 1 = \epsilon F \cdot F \eta: F \to F
\]

for an adjunction. Therefore \( \langle GF, \eta, G \epsilon \rangle \) is indeed a monad in \( X \). Call it the **monad defined by the adjunction** \( \langle F, G, \eta, \epsilon \rangle \).

For example, the **free group monad** in \( \text{Set} \) is the monad defined by the adjunction \( \langle F, G, \eta, \epsilon \rangle: \text{Set} \to \text{Grp} \), with \( G: \text{Grp} \to \text{Set} \) the usual forgetful functor.

Dually, a **comonad** in a category consists of a functor \( L \) and transformations

\[
L: A \to A, \quad \epsilon: L \to I, \quad \delta: L \to L^2
\]

which render commutative the diagrams

\[
\begin{array}{ccc}
L & \xrightarrow{\delta} & L^2 \\
\downarrow L & \quad & \downarrow L \\
L & \xrightarrow{\epsilon} & I \\
\end{array}
\]

Each adjunction \( \langle F, G, \eta, \epsilon \rangle: X \to A \) defines a comonad \( \langle FG, \epsilon, \eta G \rangle \) in \( A \).

What is a monad in a preorder \( P \)? A functor \( T: P \to P \) is just a function \( T: P \to P \) which is monotonic \( (x \leq y \implies T(x) \leq T(y)) \); there are natural transformations \( \eta \) and \( \mu \) as in (1) precisely when

\[
x \leq Tx \implies T(Tx) \leq Tx
\]

(3) for all \( x \in P \); the diagrams (2) then necessarily commute because in a preorder there is at most one arrow from here to yonder. The first equation of (3) gives \( T(x) \leq T(Tx) \). Now suppose that the preorder \( T \) is a partial order \( (x \leq y \iff x = y) \). Then the Eqs. (3) imply that \( T(Tx) = T(x) \). Hence a monad \( T \) in a partial order \( P \) is just a closure operation \( T \) in \( P \); that is, a monotonic function \( T: P \to P \) with \( x \leq Tx \) and \( t(tx) = t(x) \) for all \( x \in P \).

We leave the reader to describe a morphism \( \langle T, \mu, \eta \rangle \to \langle T', \mu', \eta' \rangle \) of monads (a suitable natural transformation \( T \to T' \)) and the category of all monads in a given category \( X \).

### 2. Algebras for a Monad

The natural question, “Can every monad be defined by a suitable pair of adjoint functors?” has a positive answer, in fact there are two positive answers provided by two suitable pairs of adjoint functors. The first answer (due to Eilenberg-Moore [1965]) construct from a monad \( \langle T, \eta, \mu \rangle \) in \( X \) a category of \( T \)-**algebras** and an adjunction \( X \to X^T \) which defines \( \langle T, \eta, \mu \rangle \) in \( X \). Formally, the definition of a \( T \)-algebra is that of a set on which the “monoid” \( T \) acts (cf. the introduction).
Definition. If $T = \langle T, \eta, \mu \rangle$ is a monad in $X$, a $T$-algebra $\langle x, h \rangle$ is a pair consisting of an object $x \in X$ (the underlying object of the algebra) and an arrow $h : Tx \rightarrow x$ of $X$ (called the structure map of the algebra) which makes both the diagrams

\[
\begin{array}{ccc}
T^1 x & \xrightarrow{T} & Tx \\
\mu_x & \downarrow & \downarrow h \\
T x & \xrightarrow{h} & x
\end{array}
\]

commute. (The first diagram is the associative law, the second the unit law.)

A morphism $f : \langle x, h \rangle \rightarrow \langle x', h' \rangle$ of $T$-algebras is an arrow $f : x \rightarrow x'$ of $X$ which renders commutative the diagram

\[
\begin{array}{ccc}
x & \xrightarrow{h} & Tx \\
f & \downarrow & \downarrow Tf \\
x' & \xrightarrow{h'} & T x'
\end{array}
\]

Theorem 1. (Every monad is defined by its $T$-algebras.) If $\langle T, \eta, \mu \rangle$ is a monad in $X$, then the set of all $T$-algebras and their morphisms form a category $X^T$. There is an adjunction

\[
\langle F^T, G^T; \eta^T, \varepsilon^T \rangle : X \rightarrow X^T
\]

in which the functors $G^T$ and $F^T$ are given by the respective assignments

\[
\begin{array}{ccc}
\langle x, h \rangle & \xrightarrow{f} & \langle x', h' \rangle \\
F^T : & \downarrow & \downarrow F^T \\
\langle x', h' \rangle & \xrightarrow{f} & \langle T x', \mu_x \rangle
\end{array}
\]

while $\eta^T = \eta$ and $\varepsilon^T \langle x, h \rangle = h$ for each $T$-algebra $\langle x, h \rangle$. The monad defined in $X$ by this adjunction is the given monad $\langle T, \eta, \mu \rangle$.

The proof is straightforward verification. If $f = \langle x, h \rangle \rightarrow \langle x', h' \rangle$ and $g : \langle x', h' \rangle \rightarrow \langle x'', h'' \rangle$ are morphisms of $T$-algebras, so is their composite $g f$; with this composition of arrows, the $T$-algebras evidently form a category $X^T$, as asserted. The functor $G^T : X^T \rightarrow X$ is the evident functor which simply forgets the structure map of each $T$-algebra. On the other hand, for each $x \in X$ the pair $\langle Tx, \mu_x : T(Tx) \rightarrow Tx \rangle$ is a $T$-algebra (the free $T$-algebra on $x$), in view of the associative and (left) unit laws for the monad $T$. Hence $x \mapsto \langle Tx, \mu_x \rangle$ does indeed define a functor $F^T : X \rightarrow X^T$, as asserted. Then $G^T F^T x = G^T (Tx, \mu_x) = Tx$, so the unit $\eta$ of the given monad is a natural transformation $\eta = \eta^T : I_x \rightarrow G^T F^T$. On the other hand, $F^T G^T (x, h) = (Tx, \mu_x)$, while the first square in the definition (1) of a $T$-algebra $\langle x, h \rangle$ states that the structure map $h : Tx \rightarrow x$ is a morphism $\langle Tx, \mu_x \rangle \rightarrow \langle x, h \rangle$ of $T$-algebras. The resulting transformation

\[
e^T_{\langle x, h \rangle} = h : F^T G^T \langle x, h \rangle \rightarrow \langle x, h \rangle
\]

is natural, by the definition (above) of a morphism of $T$-algebras. The triangular identities for an adjunction read

\[
\begin{array}{ccc}
Tx & \xrightarrow{\mu_x} & Tx \\
\downarrow & \downarrow & \downarrow \\
X & \xrightarrow{h} & x
\end{array}
\]

The first holds by the (right) unit law for $T$, the second by the unit law (see (1)) for a $T$-algebra. Therefore $\eta^T$ and $\varepsilon^T$ define an adjunction, as stated.

This adjunction thus determines a monad in $X$. The endofunctor $G^T F^T$ is the original $T$, its unit $\eta^T$ is the original unit, and its multiplication $\mu^T = G^T \varepsilon^T F^T$ has $\mu^T x = G^T \varepsilon^T (Tx, \mu_x) = G^T \mu_x = \mu_x$, so is the original multiplication of $T$. The proof is complete.

We now give several examples which show that the $T$-algebras for familiar monads are the familiar algebras.

Closure. A closure operation $T$ on a preorder $P$ is a monad in $P$ (see §1); a $T$-algebra is then an $x \in P$ with $Tx \leq x$ (the structure map). Since $x \leq Tx$ for all $x$, a $T$-algebra is simply an element $x \in P$ with $x \leq Tx \leq x$. If $P$ is a partial order, this means that $x = Tx$, so that $T$-algebra is simply an element $x$ of the partial order which is closed, in the usual sense.

Group actions. If $G$ is a (small) group, then for every (small) set $X$ the definitions

\[
Tx = G \times X, \quad x \xrightarrow{\xi x} G \times X, \quad G \times (G \times X) \xrightarrow{G \times \xi} G \times X, \quad x \xrightarrow{(u, x)} \langle g_1, \langle g_2, x \rangle \rangle \xrightarrow{\langle g_1 g_2, x \rangle}
\]

for $x \in X, g_1, g_2 \in G$ and $u$ the unit element of $G$, define a monad $\langle T, \eta, \mu \rangle$ on $Set$. A $T$-algebra is then a set $X$ together with a function $h : G \times X \rightarrow X$ (the structure map) such that always

\[
h(g_1, g_2, x) = h(g_1, h(g_2, x)), \quad h(u, x) = x.
\]

If we write $g \cdot x$ for $h(g, x)$, these are just the usual conditions that $\langle g, x \rangle \mapsto g \cdot x$ defines an action of the group $G$ on the set $X$. That $T$-algebras for the monad $T$ are just the group actions is not a surprise, since our definition of $T$-algebras was constructed on the model of group actions.
Modules. If $R$ is a (small) ring, then for each (small) abelian group $A$ the definitions

$$TA = R \otimes A, \quad A \rightarrow R \otimes A, \quad R \otimes (R \otimes A) \rightarrow R \otimes A,$$

for $a \in A$, $r_1, r_2 \in R$, define a monad on $Ab$. Much as in the previous case, the $T$-algebras are exactly the left $R$-modules.

Exercises

1. Complete semi-lattices (E. Manes; thesis). Recall that a complete semi-lattice is a partial order $Q$ in which every subset $S \subseteq Q$ has a supremum (least upper bound) in $Q$. Let $\mathcal{P}$ be the covariant power set functor on $Set$ so that $\mathcal{P}X$ is the set of all subsets $S \subseteq X$, while for each function $f : X \rightarrow Y$, $(\mathcal{P}f)S$ is the direct image of $S$ under $f$. For each set $X$, let $\eta_x : X \rightarrow \mathcal{P}X$ send each $x \in X$ to the one point set $\{x\}$, while $\mu_x : \mathcal{P}\mathcal{P}X \rightarrow \mathcal{P}X$ sends each set of sets into its union.

(a) Prove that $\langle \mathcal{P}, \eta, \mu \rangle$ is a monad on $Set$.

(b) Prove that each $\mathcal{P}$-algebra $\langle X, h \rangle$ is a complete semi-lattice when $x \leq y$ is defined by $h(x, y) = y$, and $supS = hS$ for each $S \subseteq X$.

(c) Prove conversely that every (small) complete semi-lattice is a $\mathcal{P}$-algebra in this way.

(d) Conclude that the category of $\mathcal{P}$-algebras is the category of all (small) complete semi-lattices, with morphisms the sup-preservation functions.

2. Show that $G^T : X^T \rightarrow X$ creates limits.

3. (a) For monads $\langle T, \eta, \mu \rangle$ and $\langle T', \eta', \mu' \rangle$ on $X$, define a morphism $\theta$ of monads as a suitable natural transformation $\theta : T \rightarrow T'$, and construct the category of all monads in $X$.

(b) From $\theta$ construct a functor $\theta^* : X^T \rightarrow X^{T'}$ such that $G^T \cdot \theta^* = G^{T'}$ and a natural transformation $\theta^* \rightarrow \eta^{T'}$.

3. The Comparison with Algebras

Suppose we start with an adjunction $X \rightarrow A$, construct the monad $T$ defined in $X$ by the adjunction and then the category $X^T$ of $T$-algebras; we then ask: How is this related to the original category $A$? A full answer will relate not only the categories, but the adjunctions, and is provided by the following comparison theorem.

**Theorem 1.** (Comparison of adjunctions with algebras.) Let $\langle F, G, \eta, \epsilon \rangle : X \rightarrow A$ be an adjunction, $T = \langle GF, \eta, G\epsilon F \rangle$ the monad it defines in $X$. Then there is a unique functor $K : A \rightarrow X^T$ with $G^T K = G$ and $K F = F^T$. 

The Comparison with Algebras

Proof. The conclusion asserts that we can fill in the arrow $K$ in the following diagram so that both the $F$-square and the $G$-square commute

$$
\begin{array}{ccc}
A & \rightarrow & X^T \\
f \downarrow & & \downarrow f^T \\
X & \rightarrow & X^T
\end{array}
$$

Now the counit $\epsilon$ of the given adjunction defines for each $a \in A$ an arrow $G\epsilon a : GFa \rightarrow Ga$. This arrow may be considered as a structure map $h$ for a $T$-algebra structure on the object $Ga = x$, for the requisite diagrams (cases of (2.1)) are

$$
\begin{array}{c}
GF Ga \quad GGF a \\
\downarrow G\eta a \quad \downarrow G\epsilon a \\
Ga \quad Ga
\end{array}
$$

They commute (the first is the definition of $G\epsilon a$, the second is one of the triangular identities for the given adjunction). Therefore for any $f : a \rightarrow a'$ in $A$ we define $K$ by

$$
K a = \langle Ga, G\epsilon a \rangle, \quad K f = Gf : \langle Ga, G\epsilon a \rangle \rightarrow \langle Ga, G\epsilon a' \rangle;
$$

since $\epsilon$ is natural, the proposed arrow $Kf$ commutes with $Ge$ and so is a morphism of $T$-algebras. It is routine to verify that $K$ is a functor with $K F = F^T$, $G^T K = G$.

It remains to show $K$ unique. First, each $Ka$ must be a $T$-algebra, and the commutativity requirement $G^T K = G$ means that the underlying $X$-object of this $T$-algebra $Ka$ is $Ga$. Therefore $Ka$ must have the form $Ka = \langle Ga, h \rangle$ for some structure map $h$; moreover $G^T K = G$ means that the value of $K$ on an arrow $f$ in $A$ must be $Kf = Gf$, exactly as in (2) above. It remains only to determine the structure map $h$. Now (1) commutes, and the two adjunctions $\langle F, G, \eta \rangle$ and $\langle F, G^T, \eta \rangle$ have the same unit $\eta$, so the two functors $K : A \rightarrow X^T$ and the identity $I : X \rightarrow X$ define a map of the first adjunction to the second, in the sense considered in §4.7. Proposition IV.7.1 for this map then states that $K \epsilon = \epsilon^T K$. But $K$ on arrows is $G$, so $K \epsilon = G\epsilon a$ for each $a \in A$, while the definition of the counit $\epsilon^T$ of an algebra gives $\epsilon^T Ka = \epsilon^T \langle Ga, h \rangle = h$. Thus $K \epsilon = \epsilon^T K$ implies $G\epsilon a = h$, so the structure map $h$ is determined and $K$ is unique.

For many familiar adjunctions $\langle F, G, \eta \rangle$ this comparison functor $K$ will be an isomorphism; we then say that $G$ is monadic (tripleable).
However, here is an easy example when $K$ is not an isomorphism, and not even an equivalence. The forgetful functor $G: \text{Top} \to \text{Set}$ has a left adjoint $D$ which assigns to each set $X$ the discrete topological space (all subsets open in $X$), for the identity arrow $\eta_k: X \to GDX$ is trivially universal from the object $X$ to the functor $G$. This adjunction $\langle D, G, \eta, ... \rangle: \text{Set} \to \text{Top}$ defines on $\text{Set}$ the monad $I = \langle I, 1, 1 \rangle$ which is the identity (identity functor, identity as unit and as multiplication). The $I$-algebras in $\text{Set}$ are just the sets, so the comparison functor $\text{Top} \to \text{Top}^I = \text{Set}$ is in this case the given forgetful functor $G$.

4. Words and Free Semigroups

The comparison functor can be illustrated explicitly in the case of semigroups. A semigroup is a set $S$ equipped with an associative binary operation $\cdot: S \times S \to S$. The free semigroup $WX$ on a set $X$ is like the free monoid on $X$ (§11.7). It consists of all words $\langle x_1 \rangle \cdots \langle x_n \rangle$ of positive length $n$ spelled in letters $x_i \in X$, where we write $\langle \rangle$ to distinguish the word $\langle \rangle$ in $WX$ from the element $x \in X$. Words are multiplied by juxtaposition,

$$\langle x_1 \rangle \cdots \langle x_n \rangle \langle y_1 \rangle \cdots \langle y_m \rangle = \langle x_1 \rangle \cdots \langle x_n \rangle \langle y_1 \rangle \cdots \langle y_m \rangle;$$

this multiplication $\cdot$ is associative, so makes $FX = \langle WX, \cdot \rangle$ a semigroup, with the set $WX$ the disjoint union $\bigcup_n X^n$; $n = 1, 2, \ldots$. If $G: \text{Smgrp} \to \text{Set}$ is the forgetful functor from the category of all small semigroups (forget the multiplication), then the arrow $\eta_X: X \to GFX$ defined by $x \mapsto \langle x \rangle$ (send each $x$ to the one-letter word in $X$) is universal from $X$ to $G$. Therefore $F$ is a functor, left adjoint to $G$, and $\eta$ defines an adjunction

$$\langle F, G, \eta, \varepsilon \rangle: \text{Set} \to \text{Smgrp}.$$

If $S$ is any semigroup (set $S$ with an associative binary operation $S \times S \to S$, written as multiplication) the counit $\varepsilon$ of this adjunction is by definition the morphism $\varepsilon_S: FGS \to S$ of semigroups for which the composite $G \varepsilon_S G\eta_S: GS \to GFS \to GS$ is the identity; in other words, $\varepsilon_S$ is the unique morphism of semi-groups which sends each generator $\langle s \rangle$ to $s$. This means that

$$\varepsilon_S(\langle s_1 \rangle \cdots \langle s_n \rangle) = s_1 \cdots s_n \quad \text{(product in $S$)} \quad (1)$$

for all $s_1 \in S$: The counit $\varepsilon$ removes the "pointy bracket" $\langle \rangle$.

Proposition 1. The monad on $\text{Set}$ determined by the adjunction $\text{Set} \to \text{Smgrp}$ is

$$W = \langle W: \text{Set} \to \text{Set}, \eta: I \to W, \mu: W^2 \to W \rangle$$

where $WX = \bigcup_n X^n$, $\eta_X = \langle x \rangle$ for each $x \in X$, while $\mu$ is

$$\mu_S(\langle x_1 \rangle \cdots \langle x_{n_1} \rangle) \cdots \langle x_k \rangle) = \langle x_{11} \rangle \cdots \langle x_{1n_1} \rangle \cdots \langle x_{k1} \rangle \cdots \langle x_{kn_k} \rangle$$

for all positive integers $k$, all $k$-tuples $n_1, \ldots, n_k$ of positive integers, and all $x_{ij} \in X$.

Proof. By definition, $\eta_X = \langle x \rangle$, while $\mu = G \cdot F: W^2 \to W$ is determined by the formula above for $\varepsilon_S$, where we have written each element of $W^2 X$ at a word (of length $k$) in $k$ words of the respective lengths $n_1, \ldots, n_k$. More briefly, $\mu_X$ applied to a word of words, removes the outer pointy brackets.

Note that this description allows direct verification of the unit and associativity laws for the monad $W$, without overt reference to the notion of a semi-group. For example, the associativity law for $\mu$ amounts to an observation on three layers of pointy brackets, that removing first the middle brackets and then the outer brackets gives the same result as removing first the outer brackets and then the (newly) outer brackets.

Proposition 2. For the above word-monad $W$ in $\text{Set}$, the $W$-algebras have the form $(S, v_1, v_2, \ldots): A$ set $S$ equipped with an $n$-ary operation $v_n: S^n \to S$ for each positive integer $n$, such that $v_1 = 1$ while for every positive $k$ and each $k$-tuple of positive integers $n_1, \ldots, n_k$ one has the identity

$$v_k(v_{n_1} \cdots x = v_{n_k} \cdots x = \langle x_1 \rangle \cdots \langle x_k \rangle;$$

where $n$ runs over all $k$-tuples $(n_1, \ldots, n_k)$. With this notation, the associative law for the structure map $h$ takes the stated form (2).

The simplest case of this identity (2), for $n = 3$, is $\beta_1 (x, y, z) = v_3(x, y, z) = x \cdot y \cdot z$. If we write the binary operation $v_2$ as multiplication, this states that the ternary operation $v_3$ satisfies, for all elements $x, y, z \in S$,

$$(x \cdot y) \cdot z = v_3(x, y, z) = x \cdot y \cdot z.$$
Similarly, $v_n$ must be the $n$-fold product. An easy induction proves

**Corollary.** The system $\langle S, v_1, v_2, \ldots \rangle$ is a $W$-algebra, as above, if and only if $v_1 = 1$, $v_2 : S \times S \to S$ is an associative binary operation on $S$, and for all $n \geq 2$, $v_{n+1} = v_n(v_2 \times 1) : S^{n+1} \to S$.

Thus, if we start with semigroups, regarded as sets $\langle S, v \rangle$ with one associative binary operation, define the resulting monad $W$ on Set, and then construct the category of $W$-algebras, we get the same semigroups, now regarded as algebraic system $\langle S, v_1, v_2, \ldots \rangle$, where $v_1 = 1$, $v_2 = v$, and $v_{n+1}$ is $v_2$ iterated. The comparison functor $K : Smgrp \to \text{Set}^w$ is the evident map $\langle S, v \rangle \to \langle S, 1, v_2, \ldots, v_n, \ldots \rangle$ where $v_2$ is the iterate of the binary $v$. In other words, $K$ is an isomorphism, but it replaces the algebraic system $\langle S, v \rangle$ with one binary operation by the same set with all the iterated operations derived from this binary operation.

A similar description applies to algebras over other familiar monads (Exercises 1, 2).

**Exercises**

1. Let $W_0$ be the monad in Set defined by the forgetful functor $\text{Mon} \to \text{Set}$. Show that a $W_0$-algebra is a set $M$ with a string $v_0, v_1, \ldots$ of $n$-ary operations $v_n$, where $v_n : M^n \to M$ is the unit of the monoid $M$ and $v_0$ is the $n$-fold product.

2. For any ring $R$ with identity, the forgetful functor $G : R\text{-Mod} \to \text{Set}$ from the category of left $R$-modules has a left adjoint and so defines a monad $\langle T_R(\eta, \mu) \rangle$ in Set.

   a. Prove that this monad may be described as follows: For each set $X$, $T_R X$ is the set of all those functions $f : X \to R$ with only a finite number of non-zero values; for each function $t : X \to Y$ and each $y \in Y$, $\{[[T_R t] f]_x \} = \Sigma f_x$, with sum taken over all $x \in X$ with $hx = y$; for each $x \in X$, $\eta_x : x : X \to R$ is defined by $(\eta_x x) x = 1$, $((\eta_x x) y) y = 0$; for each $k \in \mu k : X \to X$ is defined for $x \in X$ by $(\mu_x k) x = \Sigma x f_x$, with sum taken over all $f \in T_R X$.

   b. From this description, verify directly that $\langle T_R(\eta, \mu) \rangle$ is a monad.

   c. Show that the $\langle T_R(\eta, \mu) \rangle$-algebras are the usual $R$-modules, described not via addition and scalar multiple, but via all operations of linear combination (The sum of the function $k$ assigns to each $f$ the “linear combination with coefficients $f_k$’).

3. Give a similar complete description of the adjunction defined by the forgetful functor $\text{CRng} \to \text{Set}$, noting that $TX$ is the ring of all polynomials with integral coefficients in letters (i.e., indeterminates) $x \in X$.

4. The adjunction $(F, G, \varphi) : A \to R\text{ng}$ with $G$ the functor “forget the multiplication in a ring” defines a monad $T$ in $A$.

   a. Give a direct description of this monad, like that in the text for $W$, with $X^*$ replaced by the $n$-fold tensor power and coproduct $1_\mathbb{I}$ by the (infinite) direct sum of abelian groups.

   b. Give the corresponding description of $T$-algebras and show that the comparison functor from rings to $T$-algebras is an isomorphism.

**5. Free Algebras for a Monad**

Given an adjunction $\langle F, G, \varphi \rangle : X \to A$,

any full subcategory $B \subseteq A$ which contains all the objects $Fx$ for $x \in X$ leads to another adjunction $\langle F_B, G_B, \varphi_B \rangle : X \to B$

where the functor $F_B$ is just $F$ with its codomain restricted from $A$ to $B$, $G_B$ is $G$ with domain restricted to $B$, while for $x \in X$ and $b \in B$ the given adjunction leads to a bijection $\varphi_B$

$$\hom_B(F_B x, b) = \hom_A(F x, b) \cong \hom_X(x, G b) = \hom_M(x, G_B b),$$

which is manifestly natural in $x$ and $b$. Moreover, this second adjunction $\varphi_B$ defines in $X$ the same monad as did the first. This observation shows that one and the same monad in $X$ can usually be defined by many different adjunctions. The "smallest" such adjunction will be the one where $B$ is $FX$, the full subcategory of $A$ with objects all the "free" objects $Fx \in A$. The familiar properties of arrows $Fx \to FY$ between such free objects do suggest a way of constructing this subcategory $FX$ and the adjunction $\varphi_B$ directly from the monad. Here is the suggested construction, which really gives this category directly and not as a subcategory (cf. Exercise 3).

**Theorem 1.** (The Kleisli category of a monad, [1965]). Given a monad $\langle T, \eta, \mu \rangle$ in a category $X$, consider to each object $x \in X$ a new object $x_T$ and to each arrow $f : x \to y$ in $X$ a new arrow $f^* : x_T \to y_T$. These new objects and arrows constitute a category when the composite of $f^*$ with $g^* : y_T \to z_T$ is defined by

$$g^* \circ f^* = (\mu_x \circ Tg \circ f)^*.$$  \hspace{1cm} (1)

Moreover, functors $F_T : X \to X_T$ and $G_T : X_T \to X$ are defined by

$$F_T : k : x \to y \mapsto (\eta_x k) : x_T \to y_T,$$ \hspace{1cm} (2)

$$G_T : f^* : x_T \to y_T \mapsto \mu_y : T f : T x \to T^2 y \to T y$$ \hspace{1cm} (3)

respectively, so that $G_T x_T = Tx$ on objects. Then the bijection $f^* \mapsto f$ gives an adjunction $\langle F_T, G_T, \varphi_T \rangle : X \to X_T$ which defines in $X$ precisely the given monad $\langle T, \eta, \mu \rangle$.

Sketch of proof: The definition of the arrows $f^*$ amounts to a bijection $X_f(x_T, y_T) \cong X(x, T y)$ on hom-sets, while the definition of the composite in $X_T$ refers to the composite

$$x \overset{f^*}{\to} T y \overset{Tg}{\to} T^2 z \overset{\mu_z}{\to} T z.$$
split coequalizers

We need certain special types of coequalizers. By a fork in a category $C$ we mean a diagram

\[ \begin{array}{ccc}
  a & \xrightarrow{\alpha} & b \\
  \downarrow{\varepsilon_1} & & \downarrow{\varepsilon_2} \\
  e & \xrightarrow{e} & c
\end{array} \]

in $C$ with $e \delta_0 = e \delta_1$. A fork is thus just a cone from the diagram $a \rightarrow b$ to the vertex $c$. Recall that an arrow $e$ is a coequalizer of the parallel pair of arrows $\delta_0$ and $\delta_1$ if it is a fork and if any $f: b \rightarrow d$ with $f \delta_0 = f \delta_1$ has the form $f = f' e$ for a unique $f': c \rightarrow d$. An arrow $e$ is called an absolute coequalizer of $\delta_0$ and $\delta_1$ in $C$ if for any functor $T: C \rightarrow X$ (to any category $X$ whatever) the resulting fork

\[ Ta \xrightarrow{T \delta_0} Tb \xrightarrow{T e} Tc \]

still has $Te$ a coequalizer (of $T \delta_0$ and $T \delta_1$). In particular, an absolute coequalizer is automatically a coequalizer. In the same way one can define absolute colimits (or absolute limits) of any other type.

A split fork in $C$ is a fork (1) with two more arrows

\[ \begin{array}{ccc}
  a & \xrightarrow{i} & b \\
  \downarrow{s} & & \downarrow{e} \\
  c & \xrightarrow{t} & d
\end{array} \]

which satisfy with the arrows (1) the conditions

\[ e \delta_0 = e \delta_1, \quad es = 1, \quad \delta_0 t = 1, \quad \delta_1 t = se. \]

We say that $s$ and $t$ split the fork (1). These conditions imply that $e$ is a split epi, with right inverse $s$. A split fork can also be represented as a pair of commutative squares

\[ \begin{array}{ccc}
  b & \xrightarrow{a} & d \\
  | & \downarrow{\varepsilon_1} & | \\
  c & \xrightarrow{f} & e
\end{array} \]

such that both horizontal composites are the identity. Put differently: The arrows $\delta_1$ and $e$ are objects in the functor category $C^2$ and $(\delta_0, e): \delta_1 e \rightarrow e$ is an arrow between them which has $(c, f): \delta_1 e \rightarrow \delta_1$, as its right inverse: $(\delta_0, e) (c, f) = (1, 1)$.

**Lemma.** In every split fork, $e$ is the coequalizer of $\delta_0$ and $\delta_1$.

**Proof.** For any arrow $f: b \rightarrow d$ with $f \delta_0 = f \delta_1$, take $f' = fs: c \rightarrow d$. Then, using the Eqs. (3) defining a split fork,

\[ f' e = f se = f \delta_0 t = f' \]

so $f$ factors through $e$. On the other hand, $f = ke$ for some $k: c \rightarrow d$ implies $fs = kes = k$, so $k$ is necessarily $f = fs$, and $f'$ is unique.
By a split coequalizer of $\partial_0$ and $\partial_1$ we shall mean the arrow $e$ of such a split fork on $\partial_0$ and $\partial_1$. It is possible to characterize those parallel pairs $\partial_0, \partial_1$ for which any (and hence every) coequalizer is split (Exercise 2).

Since a split fork is defined by equations involving only composites and identities, it remains a split fork under the application of any functor. Hence,

**Corollary. In every split fork, $e$ is an absolute coequalizer of $\partial_0$ and $\partial_1$.**

Here is an example of a fork in $\text{Cat}$, for $C$ any category:

$$
\begin{array}{ccc}
C^2 & \xrightarrow{\partial_0} & C \\
\xrightarrow{\partial_1} & & \xrightarrow{e} 1
\end{array}
$$

$C^2$ is the category whose objects are the arrows of $C$; $\partial_0$ and $\partial_1$ are the functors assigning to each arrow its domain and its codomain, respectively, while $e$ is the functor which sends every object of $C$ to the unique object of 1. If $C$ has a terminal object $a_0$, this fork is a split by the functor $s$ which sends the unique object of 1 to $a_0$, and the functor $t$ which sends each $c \in C$ to the unique arrow $c \rightarrow a_0$.

Here is an example of a fork in $\text{Grp}$. Let $N \triangleleft G$ be any normal subgroup of $G$ and form the semidirect product $G \ltimes N$, which has elements the pairs $\langle x, n \rangle$ for $x \in G, n \in N$ with the (evidently associative) multiplication $\langle x, n \rangle \langle y, m \rangle = \langle xy, (y^{-1}n)y \rangle m$. Then

$$
\begin{array}{ccc}
G \ltimes N & \xrightarrow{\partial_0} & G \\
\xrightarrow{\partial_1} & & \xrightarrow{e} G/N
\end{array}
$$

is a fork, where $p$ is the usual projection to the quotient group $G/N$, while $\partial_0 \langle x, n \rangle = x$, $\partial_1 \langle x, n \rangle = xn$. Moreover, in this fork $p$ is clearly the coequalizer of $\partial_0$ and $\partial_1$. This fork is not in general split, but if we apply the standard forgetful functor $U : \text{Grp} \rightarrow \text{Set}$, the resulting fork in $\text{Set}$ is split. Take $s$ to be a function sending each coset (element of $G/N$) to a representative element in $G$, while $t x = \langle x, x^{-1}(s p x) \rangle$. This example incidentally, gives one way in which any quotient group can be regarded as a coequalizer in the category of groups.

**Exercises**

1. In $\text{Ring}$ give a similar construction to show that every quotient $R/A$ of a ring $R$ by an ideal $A$ can be represented as a coequalizer, and show that the resulting fork is split after the application of the forgetful functors to sets.

2. A parallel pair $\partial_0, \partial_1 : a \rightarrow b$ is said to be contractible (Beck) if there is an arrow $t : b \rightarrow a$ with $\partial_0 t = 1$ and $\partial_1 t \partial_0 = \partial_1 t \partial_1$.
   (a) In any split fork (1), prove $\partial_0, \partial_1$ contractible;
   (b) If a contractible pair has a coequalizer, prove that this coequalizer is split.

**Beck’s Theorem**

A basic construction in familiar categories of algebras is the formation of coequalizers – in $\text{Grp}$, via factor groups, in $\text{R-Mod}$ via quotient modules, and the like. Beck’s theorem will characterize the category of $T$-algebras for any monad $T$ as a category with an adjunction in which the “forgetful” functor creates suitable coequalizers. We recall (§V.1) that a functor $G : A \rightarrow X$ creates coequalizers for a parallel pair $f, g : a \rightarrow b$ in $A$ when to each coequalizer $u : Gb \rightarrow z$ of $Gf, Gg$ in $X$ there is a unique object $c$ and a unique arrow $e : b \rightarrow c$ with $Gc = z$ and $Ge = u$ and when moreover this unique arrow $e$ is a coequalizer of $f$ and $g$.

**Theorem 1. (Beck’s theorem characterizing algebras.)** Let

$$
\langle F, G, \eta, \varepsilon \rangle : X \rightarrow A
$$

be an adjunction, $\langle T, \eta, \mu \rangle$ the monad which it defines in $X$, $X^T$ the category of $T$-algebras for this monad, and

$$
\langle F^T, G^T, \eta^T, \varepsilon^T \rangle : X \rightarrow X^T
$$

the corresponding adjunction. Then the following conditions are equivalent:

(i) The (unique) comparison functor $K : X \rightarrow X^T$ is an isomorphism;

(ii) The functor $G : A \rightarrow X$ creates coequalizers for those parallel pairs $f, g$ in $A$ for which $Gf, Gg$ has an absolute coequalizer in $X$;

(iii) The functor $G : A \rightarrow X$ creates coequalizers for those parallel pairs $f, g$ in $A$ for which $Gf, Gg$ has a split coequalizer in $X$.

**Proof.** We first show that (i) implies (ii). Consider two maps

$$
\langle x, y \rangle \xrightarrow{\partial_0} \langle y, k \rangle
$$

of $T$-algebras for which the corresponding arrows in $X$ have an absolute coequalizer

$$
\begin{array}{ccc}
x & \xrightarrow{d_0} & y \\
\xrightarrow{d_1} & & \xrightarrow{e} z
\end{array}
$$

To create a coequalizer for this parallel pair we must first find a unique $T$-algebra structure $m : Tz \rightarrow z$ on $z$ such that $e$ becomes a map of $T$-algebras, and then prove that this $e$ is, in fact, a coequalizer of $d_0, d_1$ in the category $X^T$ of $T$-algebras. But on the left side of the diagram

$$
\begin{array}{ccc}
Tx & \xrightarrow{Td_0} & Ty \\
\xrightarrow{Td_1} & & \xrightarrow{Tm} Tz
\end{array}
$$

both the upper square (with $d_0$) and the lower square (with $d_1$) commute, because $d_0$ and $d_1$ are maps of algebras; it follows that $ek$ has equal
composites with $Td_0$ and $Td_1$. But $e$ is an absolute coequalizer, so $Te$ is still a coequalizer: Therefore there is a unique vertical map $m$, as shown, which makes the right square commute.

We now wish to show that this $m$ is a structure map for $z$. The associative law for $m$ (outer square below) may be compared with the associative law for the structure map $k$ (inner square below) by the diagram

![Diagram](https://example.com/diagram)

$$T^2z \xrightarrow{Tm} Tz \xrightarrow{\mu_z} z \xrightarrow{\mu_e} Tz \xrightarrow{Tm} Tz$$

The left hand trapezoid commutes since $\mu$ is natural, and the other three trapezoids commute by the definition of $m$ above in terms of $k$ and $e$. Therefore

$$m = Tm \cdot T^2e = m \cdot \mu_e \cdot T^2e.$$ But $e$ is an absolute coequalizer, so $T^2e$ is a coequalizer and thus is epi; cancelling $T^2e$ gives the associative law for $m$. The same style of argument will prove that $m$ satisfies the unit law $m \cdot \eta_x = 1 : z \rightarrow z$.

We have found the desired unique $T$-algebra structure map $m$ on $z$, with $e$ a map of $T$-algebras by the construction of $m$. To show that $e$ is a coequalizer in $X^T$, consider any other map $f : \langle y, k \rangle \rightarrow \langle w, n \rangle$ of $T$-algebras with $f d_0 = f d_1$. Then $f : y \rightarrow w$ is an arrow in $X$ with $f d_0 = f d_1$, while $e$ is an (absolute) coequalizer of $d_0, d_1 : x \rightarrow y$. Therefore there is a map $f' : z \rightarrow w$ with $f = f' e$. An argument just like that for the diagram (3) shows that $f'$ is in fact a map of $T$-algebras. Since it is unique with $f = f' e$, this completes the proof that $e$ is a coequalizer in $X^T$, and hence that (i) implies (ii).

Next, every split coequalizer is an absolute coequalizer, hence condition (ii) of the theorem requires more creativity of $G$ than does condition (iii). Therefore (ii) implies (iii).

It remains to prove that (iii) implies (i). As a preliminary, consider a $T$-algebra $\langle x, h \rangle$; the conditions that $h : Tx \rightarrow x$ be a structure map of an algebra are exactly the conditions that

$$T^2x \xrightarrow{\mu_x} Tx \xrightarrow{h} x$$

be a fork in $X$ split by $T^2x \xrightarrow{\eta_x} Tx \xrightarrow{h} x$. Indeed, the fork condition $h \cdot \mu_x = h \cdot Th$ for (4) is just the associative law for $h$, the composite $h \cdot \eta_x = 1$ because of the unit law for $\langle x, h \rangle$, while the equations

$$\mu_x \cdot \eta_{Tx} = 1, \quad Th \cdot \eta_{Tx} = \eta_x \cdot h$$

hold by the unit law for the monad $T$ and the naturality of $\eta$.

For each object $a \in A$, the adjunction $\langle F, G, e, \eta \rangle : X \rightarrow A$ provides a fork

$$FGFGa \xrightarrow{\mu_a} FGA \xrightarrow{e} a$$

in $A$ which we call the “canonical presentation” of $a$. It does correspond to a familiar presentation if $A = Grp$; then $\eta$ is just the projection on the group $a$ of the free group generated by all the elements of $a$. If the functor $G$ is applied to the fork (5) we get a split fork in $X$; indeed, that special case of the split fork (4) when $\langle x, h \rangle$ is the $T$-algebra $\langle Ga, Ge_x \rangle$ used in the comparison theorem.

Now consider any other adjunction $\langle F', G', \eta', e' \rangle : X \rightarrow A'$ which defines the same monad in $X$. By a comparison (of $F'$ to $F$) we mean a functor $M : A \rightarrow A'$ with $MF' = F$ and $GM = G'$; as already noted, such a comparison is a morphism of adjunctions and hence satisfies $Me' = e M$.

**Lemma.** If $G$ satisfies hypothesis (iii) of the theorem on the creation of coequalizers, then there is a unique comparison $M : A \rightarrow A'$.

Since $G^T$ is now known to satisfy this hypothesis, this lemma will incidentally provide a new proof of the comparison theorem (§3).

**Proof.** If $M$ exists, then $F GM = MF'G$ and $Me = e M$, so $M$ must carry the canonical presentation of $a'$ to the canonical presentation of $M a'$. In other words, the object $M a'$ must fit in a fork

$$FGFGG a' \xrightarrow{FGe} FG a' \xrightarrow{e} a$$

in $A$, and moreover $k$ must be $M e = e M$. Map this fork to the category $X$ by the functor $G$. The result is the fork

$$FGFGG a' \xrightarrow{G \eta'_a} FG a' \xrightarrow{G e} a$$

in $X$ which is split — since $T = GF$, it is a case of the fork (4) above, for $x = G a'$. But the hypothesis (iii) insures that $G$ creates coequalizers in this case. Therefore there is exactly one possible choice for $k$ and $M a'$ above; (moreover, once $M a'$ is chosen, $e_M a'$ has the property required of $k$, so must be $k$.) This shows that the comparison $M$ is unique if it exists.

Now choose $k$ and $M a'$ in this way. To show $M a$ functor consider any $f : a' \rightarrow b'$ in $A'$. In the diagram

$$FGFGa' \xrightarrow{FGe} FG a' \xrightarrow{k} M a'$$

$$FGFG G f \xrightarrow{G \eta'_a} FG G f \xrightarrow{G e} M f$$

$$FGFGG b' \xrightarrow{G \eta'_b} FG b' \xrightarrow{G e} M b'$$

Beck’s Theorem

$$h = \eta_x$$
both left-hand squares commute, so $k_s: FGf$ must factor through the
first coequalizer $k$ by a unique arrow $Ma\to Mb$ as shown. Taking this
arrow to be $Mf$ clearly makes $M$ a functor $A\to A$, just as required for
the lemma.

By this lemma we construct both the original comparison functor
$K: A\to X^T$ and a comparison functor $M: X^T\to A$. The composite
$MK: A\to A$ is then a comparison (of the adjunction $F\to$ to itself),
hence must be the identity, again by the lemma. Similarly, $KM: X^T\to X^T$
is a comparison of $F^T$ to $F^T$, hence must be the identity. Now $MK=1$
and $KM=1$ prove $K$ an isomorphism, as required for (i).

The construction of $M$ in this theorem may be further analyzed,
using for parallel pairs the following notion of “reflection” of colimits:

**Definition.** A functor $G: A\to X$ reflects colimits of $T: J\to A$ when
every cone $\lambda: T\to a$ from $T$ to $a \in A$ for which $G\lambda: GT\to Ga$ is a
collimiting cone in $X$ is already a colimiting cone in $A$.

In particular, $G$ reflects coequalizers when every fork in $A$ which be-
comes a coequalizer in $X$ is already a coequalizer in $A$. Similarly, $G$
reflects isomorphisms when, for all arrows $t$ of $A$, $Gt$ an isomorphism
implies $t$ an isomorphism.

Beck’s theorem has an acronym PTT for “precise tripleability
theorem”. There are many other versions: A “weak” version, easier to
prove, where there are hypotheses on the coequalizers of more pairs
(Exercises 2, 3), an “equivalence” version, which gives conditions that
the comparison functor $K: A\to X^T$ be not an isomorphism but an equivalence
of categories (Exercises 2, 6), a “constructive” version which analyses
the hypotheses (certain hypotheses suffice to give a left adjoint for $K$;
others make this adjunction an equivalence: Exercises 2, 5), a “crude”
version (CTT or VTT) with strong hypotheses which apply well to the
composite of several “forgetful” functors (Exercises 9–11).

**Exercises**

(Throughout, “coequalizers” means “coequalizers of parallel pairs”)

1. If $G$ creates coequalizers, prove that $G$ reflects coequalizers.
2. Weak Tripleability Theorem (Beck’s thesis). Given the adjunction (1) and
the corresponding comparison functor $K$, give a direct proof of the following:
(a) If $A$ has all coequalizers, then $K$ has a left adjoint $L$.
(b) If, in addition, $G$ preserves all coequalizers, then the unit of this adjunction
is an isomorphism $I \cong K L$.
(c) If, in addition, $G$ reflects all coequalizers, then the counit of this adjunction
is an isomorphism $L K \cong I$.

3. (Alternative hypothesis for Exercise 2). If $A$ has all coequalizers, $G$
reflects all coequalizers, and $G$ preserves colimits, then $G$ preserves
all coequalizers.

4. (a) Show that the canonical presentation of a $T$-algebra $\langle x, h \rangle$ is
$$
\langle T^2 x, \mu_{T x} \rangle \xrightarrow{\mu_{x}} \langle T x, \mu_x \rangle \xrightarrow{h} \langle x, h \rangle.
$$

(b) Show that the comparison functor $M: X^T \to A$ in Beck’s theorem appears
as a coequalizer diagram
$$
FGFx \xrightarrow{\epsilon_x} Fx \to M(x, h).
$$

5. Given the data (1), (2), and the comparison functor $K$, let $P$ be the set of all those
parallel pairs $f, g: a \to b$ in $A$ such that $Gf, Gg$ has a split coequalizer. Using
Exercise 4(b), prove
(a) If $A$ has coequalizers of all pairs in $P$, $K$ has a left adjoint $M$.
(b) If, in addition, $G$ preserves all coequalizers of pairs in $P$, then the unit
$\eta: I \to KM$ of this adjunction is an isomorphism.
(c) If, in addition to (a), $G$ reflects coequalizers for all pairs in $P$, then the counit
$MK \to I$ of this adjunction is an isomorphism.

6. Use the results of Exercise 5 and Theorem IV.4.1 to prove the following version of
Beck’s theorem, characterizing the category of $T$-algebras up to equivalence:
Given the data (1) and (2), the following assertions are equivalent:
(i) The comparison functor $K: A \to X^T$ is an equivalence of categories.
(ii) If $f, g$ is any parallel pair in $A$ for which $Gf, Gg$ has an absolute coequalizer,
then $A$ has a coequalizer for $f, g$, and $G$ preserves and reflects coequalizers
for these pairs.
(iii) The same, with “absolute coequalizer” replaced by “split coequalizer”.

The next exercises use certain definitions of properties CTT, VTT, PTT
for a functor $G: A \to X$. Let $C_G$ (respectively $S_G$) be the set of all those parallel
pairs $\langle f, g \rangle$ in $A$ such that $\langle Gf, Gg \rangle$ has a coequalizer in $X$ (respectively,
a split coequalizer). Then $G$ has CTT when $G$ has a left adjoint, preserves
and reflects all coequalizers which exist, and when $A$ has coequalizers of all pairs
in $C_G$. Next, $G$ has VTT when $G$ has a left adjoint, reflects coequalizers of all
pairs in $S_G$, and when $A$ has split coequalizers of all pairs in $S_G$. Finally, $G$ is
PTT when $G$ has a left adjoint, preserves and reflects coequalizers for all pairs
in $S_G$, and when $A$ has coequalizers of all pairs in $S_G$. Clearly, CTT and VTT
imply PTT.

7. CTT (Crude Tripleability Theorem; Barr-Beck). If $G$ is CTT, prove that the comparison
functor $K$ is an equivalence of categories.

8. VTT. If $G$ is VTT, prove that the comparison functor is an equivalence
of categories.

9. Given functors $G_1: A \to X$, $G_2: X \to Y$, $G_3: Y \to Z$ with $G_1$ CTT, $G_2$ PTT,
and $G_3$ VTT, prove that the composite functor $G_1 G_2 G_3$ is PTT.

10. Prove that the composite of two VTT functors is VTT.

11. Prove that the composite of two CTT functors is CTT.
8. Algebras are $T$-algebras

For semi-groups, monoids, and rings, we already know (§4) that the comparison functor is an isomorphism. This result holds more generally for any variety:

**Theorem 1.** Let $\Omega$ be a set of operators, $E$ a set of identities (on the operators derived from $\Omega$), $G$ the forgetful functor from the category $\langle \Omega, E \rangle$-Alg of all small $\langle \Omega, E \rangle$-algebras to $\text{Set}$, and $T$ the resulting monad in $\text{Set}$. Then the comparison functor $K: \langle \Omega, E \rangle$-$\text{Alg} \to T \text{Set}$ is an isomorphism.

The proof will use Beck's theorem. Consider any parallel pair $f, g: A \Rightarrow B$ of morphisms of $\langle \Omega, E \rangle$-algebras for the underlying functions have an absolute coequalizer $e$:

$$GA \xrightarrow{g} GB \xrightarrow{e \omega_g} X.$$ (1)

To "create coequalizers" we must show that the set-map $e$ lifts to a unique morphism $B \to \Omega$ of algebras, and then that this map is a coequalizer of the algebra maps $f, g$. So consider any $n$-ary operator $\omega \in \Omega$ with its given actions $\omega_A$ and $\omega_B$ on the sets $A$ and $B$ (as usual, we confuse the algebra $A$ with its underlying set $|A|$). In the diagram below (ignore the right hand square)

$$\begin{array}{c|c|c|c|c|c|c|c|c|c}
A & f & B & C & \omega_A & e \omega_B & \omega_X & \omega_C & e' & A \xrightarrow{\omega} B \xrightarrow{e\omega_B} X \\
B & f & C & \omega_B & \omega_X & \omega_C & e' & \omega' & \omega & B \xrightarrow{\omega_B} C \\
\end{array}$$ (2)

the two left hand squares (with $f$ and $g$, respectively) commute because $f$ and $g$ are morphisms of $\Omega$-algebras. The function $e$ is an absolute coequalizer in $\text{Set}$ and therefore its $n$-th power $e^n$ is still a coequalizer (of $f^n$ and $g^n$). But

$$e\omega_B = ef \omega_A = e \omega_A \omega_B = e \omega_B \omega_B,$$

so $e \omega_B$ must factor uniquely through this coequalizer as $e \omega_B = \omega_X e'$. This defines the operation $\omega_X$ on $X$ so that the square (2) on $e$ commutes; that is, that $e$ is a morphism of $\Omega$-algebras. The same diagram applies to all the derived operators $\lambda$ and defines $\lambda_X$ uniquely; it follows that any identity $\lambda_B = \mu_B$ valid in $B$ is also valid in $X$, so $X$ is a $\langle \Omega, E \rangle$-algebra.

It remains to show $e$ a coequalizer for algebra. So consider any morphism $h: B \to C$ of algebras with $hf = hg$. Then $h = h \omega$ in $\text{Set}$ (apply the forgetful functor $G$), so $h$ factors as $h = h' e$ for a unique function $h'$. We must show that the right hand square in (2) above commutes for every operator $\omega$. But $h$ is a morphism of algebras, so

$$h' \omega_X e' = h' e \omega_X = h \omega_B = \omega_C h' = \omega_C h^n e^n$$

and $e'$ a coequalizer means $\omega e'$ epi, hence gives $h' \omega_X = \omega_C h^n$, as required.

## Exercises

1. Prove Theorem 1, using split coequalizers rather than absolute coequalizers, noting that each $\omega_A$ must be defined in terms of a splitting $(\iota, \iota)$ of the fork (1) as

$$\omega_A(x_1, \ldots, x_n) = e \omega_B(s(x_1, \ldots, x_m), x_n), \quad x_n \in X.$$ (For $n = 2$, observe that this is like the usual definition of the product of cosets of a normal subgroup.)

2. If $K$ is a commutative ring, show that Beck's theorem applies to the forgetful functor $K$-$\text{Alg} \to \text{K-Mod}$.

9. Compact Hausdorff Spaces

**Theorem 1.** The standard forgetful functor

$$G: \text{Cmpt Haus} \to \text{Set},$$

which assigns to each (small) compact Hausdorff space its underlying set, is monadic.

**Proof.** We already know that $G$ has a left adjoint $F$; indeed, we may take each $FX$ to be the Stone-Čech compactification $(V6.2)$ of the set $X$ with the discrete topology.

For the remainder of the proof (given in a form due to R. Paré [1971]) it is convenient to regard a topological space as a pair $(X, (-)X)$ consisting of a set $X$ and a closure operation $S \mapsto \overline{S}$ defined for all subsets $S, T \subseteq X$ with the standard properties

$$\emptyset = \emptyset, \quad S \subseteq \overline{S}, \quad S \cup T = \overline{S} \cup \overline{T} \subseteq \overline{S} \cup \overline{T},$$

with $\emptyset$ the empty subset. A continuous map $f: (X, (-)X) \to (Y, (-)Y)$ is then a function $f: X \to Y$ such that $f \overline{S} \subseteq \overline{fS}$ for all $S \subseteq X$. Also a function $f: X \to Y$ is closed if $fS \subseteq \overline{fS}$ for all $S \subseteq X$. We recall the well-known

**Lemma.** If $X$ is a compact space and $Y$ a Hausdorff space, then every continuous $f: (X, (-)X) \to (Y, (-)Y)$ is closed.

We must verify that the forgetful functor $G$,

$$(X, (-)X) \mapsto X,$$

creates coequalizers for suitable pairs. So let $f, g: (X, (-)X) \Rightarrow (Y, (-)Y)$ be a pair of continuous maps such that there is a set $W$ and an absolute coequalizer $e$,

$$X \xrightarrow{e} Y \xrightarrow{\omega} W,$$

in $\text{Set}$. Let $P$ denote the covariant power set functor $\text{Set} \to \text{Set}$; thus for each subset $S \subseteq Y, (Pe)S \subseteq W$ is the usual direct image of $S$ under $e$. 

Compact Hausdorff Spaces

Exercises

1. Prove Theorem 1, using split coequalizers rather than absolute coequalizers, noting that each $\omega_A$ must be defined in terms of a splitting $(\iota, \iota)$ of the fork (1) as

$$\omega_A(x_1, \ldots, x_n) = e \omega_B(s(x_1, \ldots, x_m), x_n), \quad x_n \in X.$$ (For $n = 2$, observe that this is like the usual definition of the product of cosets of a normal subgroup.)

2. If $K$ is a commutative ring, show that Beck's theorem applies to the forgetful functor $K$-$\text{Alg} \to \text{K-Mod}$.
Since \( e \) is an absolute coequalizer, \( Pe \) is still a coequalizer, in the diagram (of sets)

\[
\begin{array}{c}
P \rightarrow Y \xrightarrow{p} PW \\
\downarrow{\scriptstyle \iota_{PW}} \quad \downarrow{\scriptstyle \iota_{PW}} \\
PX \xrightarrow{\iota_X} PY \xrightarrow{p} PW
\end{array}
\]

Since \( f \) and \( g \) are both continuous maps, both squares on the left (the square with \( f \)), and that with \( g \)) are commutative. It follows that

\[
Pe \cdot (-) \gamma \cdot Pf = Pe \cdot (-) \gamma \cdot Pg.
\]

But \( Pe \) is a coequalizer, so \( Pe \cdot (-) \gamma \) factors through \( Pe \). This gives a unique function \( (-) \omega \) the dotted arrow in \((1)\) which makes the right hand square in the diagram commute. This function may thus be described as follows: Given a subset \( T \subset W \), choose any subset \( S \subset Y \) with \((Pe)S = T \); then \( \bar{T} = (Pe)S \), independent of the choice of \( S \). In particular, if \( e^{-1}T \subset Y \) is the usual inverse image of \( T \), then \( \bar{T} = Pe(e^{-1}T) \). It is now routine to verify that this is a closure operation on \( W \), hence that \( W \) is a topological space.

By the commutativity of the diagram, \( e \) is then continuous and closed. Since \( Y \) is compact and \( e: Y \rightarrow W \) is surjective, \( W \) is also compact. Since \( Y \) is Hausdorff, each point in \( Y \) is a closed set; since \( e \) is a closed map and is surjective, the points of \( W \) are closed. To show \( W \) is Hausdorff, consider two points \( w_1, w_2 \in W \). They are closed in \( W \), so \( e^{-1}w_1 \) and \( e^{-1}w_2 \) are disjoint closed sets in \( Y \). By a familiar property of the compact Hausdorff space \( Y \), disjoint closed sets can be separated by disjoint open sets (every compact Hausdorff space is normal), so there are disjoint open sets \( U_1, U_2 \subset Y \) with \( e^{-1}w_1 \subset U_1 \) and \( e^{-1}w_2 \subset U_2 \). Their complements \( U_1^c \) and \( U_2^c \) in \( Y \) are then closed sets with \( U_1^c \cup U_2^c = Y \). Since \( e \) is a closed map, \( e(U_1^c) \cup e(U_2^c) = W \) with

\[
(Pe)(U_1^c) \cup (Pe)(U_2^c) = W, \quad \omega_w \notin (Pe)(U_1^c) \cup (Pe)(U_2^c).
\]

So take complements again, this time in \( W \): \([Pe(U_1^c)]^c \) and \([Pe(U_2^c)]^c \) are disjoint open neighborhoods of \( w_1 \) and \( w_2 \), respectively, in \( W \). Therefore \( W \) is a Hausdorff space.

We have produced from the absolute coequalizer \( e \) in \( \mathbf{Set} \) a unique topology on \( Y \) such that \( e \) is continuous; moreover, this topology is compact Hausdorff. It remains to show that the continuous map \( e: (Y, (-) \omega \rightarrow (W, (-) \omega) \) is a coequalizer in \( \text{Cmprt. Haus} \). So consider any compact Hausdorff \((Z, (-) \omega) \) and a continuous map \( h: Y \rightarrow Z \), such that both composites in

\[
X \xrightarrow{f} Y \xrightarrow{h} Z
\]
VII. Monoids

This chapter will explore the general notion of a monoid in a category. As we have already seen in the introduction, an ordinary monoid in $\text{Set}$ is defined by the usual diagrams relative to the cartesian product $\times$ in $\text{Set}$, while a ring is a monoid in $\text{Ab}$, relative to the tensor product $\otimes$ there. Thus we shall begin with categories $B$ equipped with a suitable bifunctor such as $\times$ or $\otimes$, more generally denoted by $\boxtimes$. These categories will themselves be called “monoidal” categories because the bifunctor $\boxtimes : B \times B \to B$ is required to be associative. Usually it is associative only “up to” an isomorphism; for example, for the tensor product of vector spaces there is an isomorphism $U \otimes (V \otimes W) \cong (U \otimes V) \otimes W$. Ordinarily we simply “identify” these two iterated product spaces by this isomorphism. Closer analysis shows that more care is requisite in this identification — one must use the right isomorphism, and one must verify that the resulting identification of multiple products can be made in a “coherent” way.

Once the coherence question for monoidal categories is settled, we proceed to define monoids in such categories, the actions of monoids on objects of the category, and the construction of free monoids. Next, we introduce the simplicial category $\Delta$, which turns out to be the basic monoidal category because it contains a “universal” monoid and because of its role in simplicial resolutions and simplicial topology. Finally, compactly generated spaces are used to illustrate closed monoidal categories.

1. Monoidal Categories

A category is monoidal when it comes equipped with a “product” like the direct product $\times$, the direct sum $\oplus$, or the tensor product $\otimes$. We write this product as $\boxtimes$ (many authors write $\otimes$) to cover all cases impartially. We consider first categories equipped with a multiplication $\boxtimes$ which is strictly associative and has a strict two-sided identity object $e$. In detail, a strict monoidal category $(B, \boxtimes, e)$ is a category $B$ with a bifunctor $\boxtimes : B \times B \to B$ which is associative,

$$\boxtimes((\boxtimes x 1) = \boxtimes(1 \times \boxtimes),$$

and with an object $e$ which is a left and right unit for $\boxtimes$.

$$\boxtimes(e \times 1) = \text{id}_B = \boxtimes(1 \times e).$$

In writing the associative law (1), we have identified $(B \times B) \times B$ with $B \times (B \times B)$; in writing the unit law (2), we mean $e \times 1$ to be the functor $c \mapsto (e, c) : B \to B \times B$. The bifunctor $\boxtimes$ assigns to each pair of objects $a, b \in B$ an object $a \boxtimes b$ of $B$ and to each pair of arrows $f : a \to a', g : b \to b'$ an arrow $f \boxtimes g : a \boxtimes b \to a' \boxtimes b'$. Thus if $\mathbf{0}$ a bifunctor means that the interchange law

$$1_{\mathbf{0}} e \mathbf{1}_e = 1_{\mathbf{0}} e \mathbf{1}_e,$$

holds whenever the composites $f' f$ and $g' g$ are defined. The associative law (1) states that the binary operation $\boxtimes$ is associative both for objects and for arrows; similarly, the unit law (2) means that $e \boxtimes c = c \boxtimes e$ for objects $c$ and that $1_{\mathbf{0}} f = f = 1_{\mathbf{0}} e$ for arrows $f$.

Any monoid $M$ (in the usual sense, in $\text{Set}$), regarded as a discrete category, is a strict monoidal one with $\boxtimes$ the multiplication of elements of $M$. If $X$ is any category, the category End$(X)$ of all endofunctors $S : X \to X$ and arrows $\lambda$ between all natural transformations $\mu : S \to T$ is strict monoidal, with $\boxtimes$ the composition of functors.

A (relaxed) monoidal category is a category $B$ with a bifunctor $\boxtimes$, its multiplication, which is associative “up to” a natural isomorphism $\alpha$, and which has an object $e$ which is a left unit for $\boxtimes$ up to a natural isomorphism $\lambda$ and a right unit up to $\mu$. Moreover, “all” diagrams involving $\alpha, \lambda,$ and $\mu$ must commute.

Formally, a monoidal category $B = (\langle B, \boxtimes, e, \alpha, \lambda, \mu \rangle)$ is a category $B$, a bifunctor $\boxtimes : B \times B \to B$, an object $e \in B$, and three natural isomorphisms $\alpha, \lambda, \mu$. Explicitly,

$$\alpha_{a,b,c} : a \boxtimes (b \boxtimes c) \cong (a \boxtimes b) \boxtimes c$$

is natural for all $a, b, c \in B$, and the pentagonal diagram

$$\begin{align*}
\text{a} \boxtimes (b \boxtimes (c \boxtimes d)) & \xrightarrow{\alpha} (a \boxtimes b) \boxtimes (c \boxtimes d) \xrightarrow{\alpha} ((a \boxtimes b) \boxtimes c) \boxtimes d \\
\text{a} \boxtimes (b \boxtimes (c \boxtimes d)) & \xrightarrow{\lambda} (a \boxtimes (b \boxtimes c)) \boxtimes d
\end{align*}$$

(5)

commutes for all $a, b, c, d \in B$. Again, $\lambda$ and $\mu$ are natural

$$\lambda : a \boxtimes e \cong a, \quad \mu : a \boxtimes e \cong a$$

(6)
Monoidal Categories

There are many other examples. A discussion like that for Ab shows for each commutative ring \( K \) that \( \langle K \text{-Mod}, \otimes K, K \rangle \) is monoidal. The same holds for graded \( K \)-modules and for differential graded \( K \)-modules (= chain complexes of \( K \)-modules) under the customary definition of the tensor product for such objects (Mac Lane [1963]). Similarly, the category of all \( K \)-algebras (or, all differential graded \( K \)-algebras) is monoidal, under the familiar tensor product of algebras. For any ring \( R \), the category of all \( R - R \) bimodules is monoidal under \( \otimes_R \).

A (strict) morphism of monoidal categories.

\[ T : (B, \Box, e, \alpha, \lambda, \rho) \to (B', \Box', e', \alpha', \lambda', \rho'), \]

is a functor \( T : B \to B' \) such that, for all \( a, b, c, f, \) and \( g \)

\[ T(a \square b) = T(a) \square T(b), \quad T(f \square g) = T(f) \square T(g), \quad Te = e', \]

\[ T_{a,b,c} = a_{T(a),T(b),T(c)}, \quad T\lambda = \lambda', \quad T\rho = \rho'. \]

With these morphisms as arrows, we can form \( \text{MonCat} \), the category of all small monoidal categories. This category has (the obvious) finite products; in particular \( I \) with the evident (strict) monoidal structure is terminal in \( \text{MonCat} \). There is also a full subcategory consisting of all \text{strict} monoidal categories; naturally, the definition of morphisms \( T \) for these can omit the conditions (11) on \( \alpha, \lambda, \) and \( \rho \).

Many useful morphisms between monoidal categories are, however, not strict in the sense of (10) and (11). For example, the forgetful functor \( U : \langle K \text{-Mod}, \otimes K, \to \langle \text{Ab}, \otimes, \to \rangle \) is not strict; indeed, for \( K \)-modules \( A \) and \( B \), we have not an equality \( U(A \otimes B) = U(A) \otimes U(B) \) nor even an isomorphism, but just a natural morphism \( U(A) \otimes U(B) \to U(A \otimes B) \), expressing the fact that \( A \otimes B \) is a quotient of \( A \otimes B \). A similar situation arises for the forgetful functor \( \langle \text{Ab}, \otimes, \to \rangle \to \langle \text{Set}, \times, \to \rangle \). We shall not formulate here the properties of these "relaxed" morphisms between monoidal categories.

One might be tempted to avoid all this fuss with \( \alpha, \lambda, \) and \( \rho \) by simply identifying all isomorphic objects in \( B \). This will not do, by the following argument due to Isbell. Let \( \text{Set}_0 \) be the skeleton of the category of sets; it has a product \( X \times Y \) with projections \( p_1 \) and \( p_2 \) as usual. If \( D \) is a (the) denumerable set, then \( D = D \times D \), and both projections of this product are epimorphisms \( p_1, p_2 : D \to D \). Now suppose that the isomorphism \( \alpha : X \times (Y \times Z) \to (X \times Y) \times Z \) defined as usual to commute with the three projections, were always the identity; it is then the identity for \( X = Y = Z = D \) since \( \alpha \) is natural, \( f \times (g \times h) = (f \times g) \times h \) for any three \( f, g, h : D \to D \). But on functions is defined in terms of the projections \( p_1 \) and \( p_2 \) above, so

\[ p_1 = p_1(f \times (g \times h)) = p_1((f \times g) \times h) = (f \times g) \]

and \( p_2 \) is epimorphic, so \( f = f \times g \). The corresponding argument with \( p_2 \) gives \( f \times g = g \), hence \( f = g \) for any \( f, g : D \to D \), an absurdity. A similar argument applies to the skeleton of \( \langle \text{Ab}, \otimes, \to \rangle \).
Exercises

1. Prove that (5) and (7) imply (9). Hint: Take the pentagon (5) with \( a = b = e \) and fill in the inside, adding \( g \) in two places, the basic identity (7) twice, and suitable naturalities to get \((\lambda \sqcap 1)\lambda = \lambda \lambda : e \circ (e \sqcap (d \sqcap d)) \rightarrow e \sqcap d\), and hence \((\lambda \sqcap 1)\lambda = \lambda\).

2. Construct the product in \textit{Moncat} of two monoidal categories.

3. For \( B \) monoidal, show that \( B^{op} \) has the (evident) monoidal structure.

4. For \( B \) monoidal and \( C \) any category, show that the functor category \( BC \) is monoidal, with multiplication \( S \sqcap T \) defined by \( (S \sqcap T) c = S \sqcap T c \) and \( e : C \rightarrow B \) the constant functor \( e \). Show that the adjunction \( BC \rightarrow \mathbb{B} \) is isomorphic to \( \mathbb{B}^{op} \) is an isomorphism of monoidal categories.

5. Prove: A strict monoidal category with one object is a set (the set of arrows) with two binary operations \( \sqcap, \lambda \) which satisfy the interchange law and have a common (left and right) unit \( id \).

6. Show by examples that the axioms (5) and (7) are independent.

2. Coherence

A coherence theorem asserts: “Every diagram commutes”; more modestly, that every diagram of a certain class commutes. The class of diagrams at issue now are the diagrams in a monoidal category which, like the pentagon (1.5), are built up from instances of \( \alpha, \beta, \gamma \) and \( \Phi \) by multiplications \( \sqcap \). However, two apparently or formally different vertices of such a diagram might become equal in a particular monoidal category, in such a way as to spoil the commutativity. Hence we prove only that every “formal” diagram commutes, where a formal diagram is one in which the vertices are iterated formal \( \sqcap \)-products of “variables”. We call these formal products “binary words”: they are exactly like the well-formed formulas and terms used in logical syntax in proof theory.

The precise definition is by recursion. A binary word of length 0 is the symbol \( e_0 \) (the empty word); a binary word of length 1 is the symbol (\( \square \)) (the variable or the place holder); if \( v \) and \( w \) are binary words of lengths \( m \) and \( n \), respectively, then the symbol \( v \sqcap w = (v \sqcap w) \) is a binary word of length \( m + n \). For example, \((1 \sqcap 2) \sqcap (1 \sqcap 2) \) is a binary word of length 3 - an iterated 4-fold product, with chosen arrangement of parentheses, and a specified argument set equal to \( e_0 \). For any two binary words \( v \) and \( w \) of the same length, introduce one arrow \( v \rightarrow w \). These words with these arrows form a category \( W \) (a preorder with every arrow invertible). It is a monoidal category under multiplication \( v \rightarrow w \rightarrow \rightarrow w \), with unit \( e_0 \), and with \( \alpha, \beta, \gamma \) and \( \Phi \) the appropriate (and necessarily unique) arrows.

By its very construction (unique arrows \( v \rightarrow w \)) every diagram in \( W \) will commute. Morphisms from \( W \) to \( B \) then give the desired diagrams which commute in other monoidal categories \( B \). These morphisms are given by the following theorem, which states in effect that \( W \) is the free monoidal category on one generator (\( \bullet \)).

**Theorem 1.** For any monoidal category \( B \) and any object \( b \in B \), there is a unique morphism \( \mathbb{W} \rightarrow B \) of monoidal categories with \( \mathbb{W} \rightarrow B \).

**Proof.** We write the desired morphism as \( w \rightarrow \mathbb{W} \) to suggest that it means “Substitute \( b \) in all the blanks of the word \( w \)”. On objects \( w \) we must set

\[
(e_0)_{b} = e, \quad (\square)_{b} = b, \quad (v \sqcap w)_{b} = v_{b} \sqcap w_{b};
\]

by induction, these formulas uniquely determine all \( w \).

For words of fixed length \( n \) we now construct a certain “basic” graph \( G_{n} = G_{n,b} \). Its vertices are all words \( w \) of length \( n \) which do not involve \( e_0 \), while its edges \( v \rightarrow w \) are to be identical with certain arrows \( v \rightarrow w \) in \( B \). Call them the “basic” arrows. Here each instance

\[
\alpha : u_{b} \sqcap (v \sqcap w_{b}) \rightarrow (u \sqcap v_{b}) \sqcap w_{b}
\]

of associativity and each instance of \( \alpha^{-1} \) is basic, as are all arrows \( \beta \sqcap 1 \) or \( 1 \sqcap \beta \) with \( 1 : v \rightarrow u_{b} \) an identity and \( \beta \) already recognized as basic. Intuitively, each basic arrow is an arrow such as \( (1 \sqcap 2) \sqcap (1 \sqcap 2) \) - one instance of \( \alpha \), boxed with identities. Observe then that each basic arrow is either “directed” (it involves \( \alpha \)) or “antidirected” (with \( \alpha^{-1} \)). In the graph \( G_{n} \), the paths from \( u \) to \( w \) are thus the composable sequences of basic arrows from \( u_{b} \) to \( w_{b} \); by composition each path yields an arrow \( u \rightarrow w \) in \( B \). The crux of our proof will be to show that any two paths from \( u \) to \( w \) yield by composition the same arrow \( u \rightarrow w \) in \( B \) - i.e., that the graph \( G_{n} \) is commutative in \( B \).

First, take \( w^{n} \) to be the unique word of length \( n \) which has all pairs of parentheses starting in front. There is a directed path in \( G_{n} \) from any \( w \) to \( w^{n} \); indeed, we may choose such a path in a canonical way, successively moving outermost parentheses to the front by instances of \( \alpha \). For any two words \( v \) and \( w \) of length \( n \) the two canonical paths combine to give a path \( v \rightarrow w^{n} \rightarrow w \); this observation is really just the known proof of the “general associative law” for a product \( ab \), given the usual associativity law \( a(bc) = (ab)c \).

Define the rank \( q \) of a word \( w \) by recursion, setting \( q_{e_0} = 0, q_{(\square)} = 0 \), and

\[
q(v \sqcap w) = q(v) + q(w) + 1;
\]

observe that \( q_{w} = 0 \) means that all pairs of parentheses in \( w \) start at the front.

Now we show that \( G_{n} \) commutes. Along any path from \( u \) to \( w \), join each vertex to the “bottom” vertex \( w^{n} \) by the canonical directed path.
A glance at the diagram
\[
\begin{array}{cccc}
\text{v} & \text{v}_1 & \text{v}_2 & \text{v}_3 \\
\text{w}^a & \text{w}^b & \text{w}^c & \text{w}^d \\
\text{w} & \text{w} & \text{w} & \text{w} \\
\end{array}
\]

indicates that it will suffice to show that any two directed paths (all \(\alpha\)'s, no \(\alpha^{-1}\)) from a \(v_1\) to \(w^m\) are equal. This will be proved by induction on the rank of \(v_1 = v\). Suppose it true for all \(v\) of smaller rank, and consider two different directed paths starting at \(v\) with (directed) basic arrows \(\beta\) and \(\gamma\), as in the figure
\[
\begin{array}{ccc}
\text{v} & \text{v'} & \text{v''} \\
\text{w}^a & \text{w}^b & \text{w}^c \\
\end{array}
\]

Both \(\beta\) and \(\gamma\) decrease the rank. Hence it will suffice to show that one can "rejoin" their codomains \(v'\) and \(v''\) by directed paths to some common vertex \(z\) in such a way that the diamond from \(v\) to \(z\) is commutative. This is done by a case subdivision. If \(\beta = \gamma\), take \(z = v' = v''\). If \(\beta \neq \gamma\), write \(v = u \square w\) and observe that \(\beta\) has one of the following three forms:
\[
\begin{align*}
\beta &= \beta' \square 1_u; \quad \beta\text{ acts "inside" the first factor } u, \\
\beta &= 1_u \square \beta''; \quad \beta\text{ acts inside the second factor,} \\
\beta &= \alpha_{u,t,s}; \quad \text{where } v = u \square w = u \square (s \square t).
\end{align*}
\]

For \(\gamma\) there are three corresponding cases.

Now compare the cases for \(\beta\) and \(\gamma\). If both act inside the same factor \(u\), we can use induction on the length \(n\). If \(\beta\) acts inside \(u\) and \(\gamma\) inside \(w\), use the diamond
\[
\begin{array}{ccc}
\text{f} & \text{u} & \text{w} \\
\text{u} & \text{u} & \text{u} \\
\text{w} & \text{w} & \text{w} \\
\end{array}
\]

which commutes because \(\square\) is a bifunctor. There remains the case when one of \(\beta\) or \(\gamma\), say \(\beta\), is \(\beta = \alpha = \alpha_{u,t,s}\), as in the third case above. Since

\[
\begin{align*}
\gamma \neq \beta, \gamma \text{ must act inside } u \text{ or inside } w. \text{ If } \gamma \text{ acts inside } u, \text{ we use a diamond from } u \square (s \square t) \text{ to } (u \square s) \square t, \text{ which commutes because } z \text{ is natural. If } \gamma \\
\text{is inside } w = s \square t \text{ and actually inside } s \text{ or inside } t, \text{ naturality of } z \text{ gives a similar diamond. There remains only the case where } y \text{ is inside } s \square t \text{ but not inside } s \text{ or } t. \text{ Then } \gamma \text{ must be an instance of } \gamma, t \text{ must be a product } t = p \square q, \text{ and our diamond must then start with}
\end{align*}
\]

\[
\begin{align*}
\gamma = (u \square s) \square (p \square q) & \to (u \square s) \square p \square q, \\
\end{align*}
\]

This we can complete to a "diamond" by taking that diamond to be the pentagon of (5). This shows that the graph \(G_n\) is commutative in \(B\); it completes the coherence proof as far as associativity alone is concerned.

It is trivial to "fold in" to this proof the applications of \(\lambda\) and \(\eta\). Formally, consider the graph \(G_n\) with vertices all words of length \(n\), including words involving \(e_0\), and with edges all basic arrows constructed, just as above, by boxing instances of \(a, \lambda, \eta\) and \(g\) (and their inverses) with identities. This graph \(G_n\) is infinite, but contains the previous (finite) graph \(G_n\) built from \(x\) alone. It remains to show \(G_n\) commutative in \(B\). For each word \(w\), there is still at least one path \(w \to w^m\). But the composite arrow obtained from any such path is equal to that for a different path which first removes all \(e\)'s, then applies \(z\). Indeed, if some \(e\) is removed by \(\lambda : e \square b \to b\) after some application of \(a\), then that \(e\) can be removed before or after by naturality of \(a\), or by (7), or by (9). Moreover by (8) it does not matter in \(e \square e\) whether \(e\) is removed by \(\lambda\) or by \(\eta\).

Finally, this reduced path has composite equal to that for a canonical path in which all the \(e\)'s are removed in some specified order (say, starting with the left-most occurrence of \(e\)). This process reduces \(G_n\) to \(G_s\) and proves that \(G_s\) is commutative in \(B\), since \(G_s\) is.

We can now define the morphism \(W \to B\) required in the theorem. The category \(W\) was constructed with exactly one arrow \(v \to w\) between words \(v\) and \(w\) of the same length \(n\); the morphism will send this arrow to the composite arrow for any path \(v_n \to w_n\) in \(G_n\), since we now know the composite to be unique (independent of the choice of the path). In virtue of this same uniqueness, this construction does define a functor \(W \to B\). Moreover this functor is a morphism of monoidal categories because

\[
 f \square g = f \cdot s \cdot 1 \cdot g = (f \square 1) \cdot (1 \square g)
\]

for any arrows \(f\) and \(g\).

The coherence result can be formulated in terms of graphs whose edges are the natural transformations \(\alpha, \lambda, \eta\) and \(g\). To state this, note first that each word \(w\) of length \(n\) (in one variable) determines for each
monoidal category \( B \) a functor \( w_B : B^n = B \times \cdots \times B \to B \) of \( n \) variables, obtained by replacing each blank \((-)\) in the word \( w \) by the identity functor of \( B \). The explicit definition of this functor, like \((1)\), is by recursion; \( (e_0) : 1 \to B \) is the constant functor \( e \in B \) and \((-I) \) is the identity functor \( B \to B \), while if \( w_B \) and \( w_B' \) are already determined for words \( w \) and \( w' \) of the respective lengths \( n \) and \( n' \), then \((w \square w')_B : B^{n+n'} = B^n \times B^{n'} \to B \times B \to B \). \hspace{1cm} (2)

With this formulation, the coherence result is as follows:

**Corollary.** Let \( B \) be a monoidal category. There is a function which assigns to each pair of words \( v, w \) of the same length \( n \) a (unique) natural isomorphism

\[ \text{can}_B(v, w) : v_B \to w_B : B^n \to B, \]

called the canonical map from \( v_B \) to \( w_B \), in such a way that the identity arrow \( e \to e \) is canonical (between functors of 0 variables), the identity transformation \( 1_B : 1_B = 1_B \) is canonical, \( a, a^{-1}, \lambda, \lambda^{-1}, g, q^0 \) are canonical, and the composite as well as the \( \square \)-product of two canonical maps is canonical.

This sort of formulation, as will appear from the proof, applies also to the case considered in the theorem itself: For each \( b \in B \) there is a function which assigns to each pair of words \( v, w \) of the same length a canonical arrow \( \text{can}_B(v, w) : v_B \to w_B \), with properties like those stated for \( \text{can}_B \).

**Proof.** From the given monoidal category \( B \) we construct a category \( I(B) \) with objects all pairs \( \langle n, T \rangle \), \( T \) any functor \( T : B^n \to B \), and with arrows \( f : \langle n, T \rangle \to \langle n, T' \rangle \) all natural transformations \( f : T \to T' \). In this category we define a multiplication by \( \langle m, S \rangle \square \langle n, T \rangle = \langle m+n, S \square T \rangle \), where \( \square \) is the composite

\[ S \square T : B^m \times B^n \to B^{m+n} \to B \times B = B \to B, \]

we take the unit \( e \) to be the functor \( 1 \to B \) constant at \( e \) and define \( \lambda : e \square T \to T \) for each \( T \) and then for each \( a \in B^n \) as the arrow \( \lambda_T : a \square T \to T a \) of \( B \). This \( \lambda \) is natural in \( T \). Similar pointwise definitions give \( \alpha \) and \( \gamma \); it is routine to verify that \( I(B) \) is a (relaxed) monoidal category.

The identity functor \( I : B \to B \) is an object of \( I(B) \). Hence the theorem above stating that \( W \) is free monoidal on \((-)\) gives (for \( b = I \) a unique morphism \( W \to I(B) \) of monoidal categories with \((-) \to I \). In particular, this morphism sends each word \( w \) to the functor \( w_B \) described in \( (2) \) above, while the unique arrow \( v \to w \), for \( v \) and \( w \) of the same length, is sent to a natural transformation \( v_B \to w_B \) which we call \( \text{can}_B(v, w) \), as in \( (3) \).

Since the functor is a morphism, it must preserve \( \alpha, \lambda, \) and \( g \). Thus, using

\[ \text{can}_B(e_0, e_0) = 1_e : e \to e, \quad \text{can}_B((-), (-)) = \text{id}_B : B \to B, \]

\[ \text{can}_B((-), (-), (-)) = \alpha : B \square (B \square B) \to (B \square B) \square B, \]

\[ \text{can}_B(e_0 \square (-), (-)) = \lambda, \quad \text{can}_B((-), e_0, (-)) = q, \]

\[ \text{can}_B(v \square w, w \square w) = \text{can}_B(v, w) \square \text{can}_B(v, w). \]

This corollary states that every diagram of the following sort is commutative:

**Vertices.** Words \( w \) of length \( n \) representing functors \( w_B : B^n \to B \).

**Edges.** Natural transformations \( \lambda, \lambda^{-1}, \alpha, \gamma \), and their \( \square \)-products. Moreover, the functors in question are \( e, I, (-) \square (-) \) and their composites, and each edge is a natural transformation between the functors represented by the vertices at its ends.

**Exercises**

1. Draw a diagram showing all canonical maps between binary words of length 5.
   (It can be regarded as a polyhedral subdivision of the surface of the sphere into 19 regions — 6 pentagons (instances of \( a \)) and 3 squares (which commute by naturality)).

2. [Stasheff 1963] Show that the diagram giving all canonical maps between words of length \( n + 3 \) can be regarded as a polyhedral subdivision of the surface of the \( n \)-sphere.

3. Construct the free monoidal category on any set \( X \), and prove that it is a suitable universal property. (Hint: Its objects are words, with any \( x \) of \( X \) a word of length 1, and there is a surjection \( W_x \to M_x \) from the set \( W_x \) of words of the free monoid on \( X \). There is a (unique) arrow \( v \to w \) if and only if \( v \) and \( w \) are words with the same image in \( M_x \).)

3. **Monoids**

Following the ideas suggested in the introduction, we can now define the notion of a monoid in an arbitrary monoidal category \( \langle B, \square, e \rangle \).

A monoid \( c \) in \( B \) is an object \( c \in B \) together with two arrows \( \mu : c \square c \to c, \eta : e \to c \) such that the diagrams

\[
\begin{align*}
\text{c}_\square(c \square c) & \xrightarrow{\varepsilon} (c \square c) \square c \xrightarrow{\varepsilon \cdot 1_{c \square c}} c \square c \\
\end{align*}
\]

\[
\begin{align*}
e \square c & \xrightarrow{\alpha} c \square (e \square c) \xrightarrow{1_c \cdot \varepsilon} c.
\end{align*}
\]

\[ (1) \]
are commutative. A morphism \( f : \langle c, \mu, \eta \rangle \rightarrow \langle c', \mu', \eta' \rangle \) of monoids is an arrow \( f : c \rightarrow c' \) such that
\[
f \mu = \mu' (f \square f) : c \square c \rightarrow c', \quad f \eta = \eta' : e \rightarrow c'.
\]
With these arrows, the monoids in \( B \) constitute a category \( \text{Mon}_B \), and \( \langle c, \mu, \eta \rangle \rightarrow c \) defines a forgetful functor \( U : \text{Mon}_B \rightarrow B \).

This definition includes a variety of cases; some already noted in our introduction:

<table>
<thead>
<tr>
<th>Monoidal Category</th>
<th>Monoids Therein</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \langle \text{Set}, \times, 1 \rangle )</td>
<td>(ordinary) Monoids</td>
</tr>
<tr>
<td>( \langle \text{Top}, \times, * \rangle )</td>
<td>Topological monoids</td>
</tr>
<tr>
<td>( \langle \mathbb{C}^*, \ast, \text{Id} \rangle )</td>
<td>Monoids (cf. Chapter VI!)</td>
</tr>
<tr>
<td>( \langle \text{Ab, } \otimes, \mathbb{Z} \rangle )</td>
<td>Rings</td>
</tr>
<tr>
<td>( \langle \text{K-Mod, } \otimes_K, K \rangle )</td>
<td>( K )-algebras</td>
</tr>
<tr>
<td>( \langle \text{Graded modules, } \ldots \rangle )</td>
<td>Graded algebras</td>
</tr>
<tr>
<td>( \langle \text{DG-K-Mod, } \otimes_K, K \rangle )</td>
<td>Differential graded ( K )-algebras</td>
</tr>
<tr>
<td>( \langle \text{B}^P, \square^p, e \rangle )</td>
<td>Comonoids in ( B )</td>
</tr>
<tr>
<td>( \langle \text{K-Mod}^{pP}, \otimes_K^p, K \rangle )</td>
<td>( K )-coalgebras</td>
</tr>
<tr>
<td>( \langle \text{Cat, } \times, 1 \rangle )</td>
<td>Strict Monoidal categories</td>
</tr>
<tr>
<td>( \langle \text{O-Grph, } \star, 0 \rightarrow 0 \rangle )</td>
<td>Categories (cf. (II. 7. 3)).</td>
</tr>
</tbody>
</table>

There is a "general associative law" which states that in a monoid \( \langle c, \mu, \eta \rangle \) any two \( n \)-fold products are equal. Specifically, if \( w \) is any binary word and \( w_1, w_2 \in B \) the corresponding object of \( B \), as defined in Theorem 2.1, the \( n \)-fold product \( \mu \) is an arrow \( \mu : w_1 \rightarrow w_2 \) given by the following recursion: If \( w = e \), \( \mu = \eta \); if \( w = (\ast) \), \( \mu \) is the identity; if \( w = (-) \), \( \mu \) is \( \mu \), and in general if \( w = u \square v \), \( \mu \) is the evident composite
\[
(u \square v) c = u \square c, \quad v \square c \rightarrow c.
\]
Proposition 1 (General Associative Law). For \( \langle c, \mu, \eta \rangle \) a monoid in \( B \), the iterated products \( \mu \) and \( \mu_w \) for any two words \( v \) and \( w \) of the same length \( n \) satisfy
\[
\mu_w : \text{can}_n(v, w) = \mu : v \rightarrow c, \quad (4)
\]
where \( \text{can}_n(v, w) : v \rightarrow w \) is the canonical arrow of Theorem 2.1.

Proof. The axioms (1) and (2) for a monoid are exactly those cases of (4) where the canonical arrow in question is \( \alpha, \lambda, \) or \( \phi \). From these cases, (4) may be verified by induction, since all canonical arrows are composites of \( \alpha \)'s, \( \lambda \)'s, and \( \phi \)'s.

For example, one may define the \( n \)-th \( 
\square \)-power of every \( b \in B \) to be
\[
\mu^n = (b \square b) \square \cdots \square b
\]
with "all parentheses in front"; thus \( b^0 = e, b^1 = b, b^n+1 = b^n \square b \). For the monoid \( \langle \mu, \eta \rangle \) the \( (n+1) \)-fold product \( \mu^{n+1} : c \rightarrow c \) is then defined by recursion as
\[
\mu^0 = \eta, \quad \mu^1 = \text{id}_c, \quad \mu^{n+1} = \mu \mu^n = \mu (\mu^n \square 1).
\]
Then (4) includes the more familiar equation ("general associative law")
\[
\mu^n (\mu^{k_1} \square \cdots \square \mu^{k_n}) = \mu^{k_1 + \cdots + k_n}
\]
valid for all natural numbers \( n \) and \( k_1, \ldots, k_n \).

Theorem 2 (Construction of free monoids). If the monoidal category \( B \) has denumerable coproducts, and if for each \( a \in B \) the functors \( \square - \) and \( - \square a : B \rightarrow B \) preserve these coproducts, then the forgetful functor \( U : \text{Mon}_B \rightarrow B \) has a left adjoint.

Note: In many cases \( B = \text{Set}, B = \text{Ab}, \ldots \) the functors \( \square - \) and \( - \square a \) themselves have right adjoints, hence automatically preserve coproducts.

Proof. The distributive law \( \theta : \bigcup (a \square b) \approx a \square \bigcup b \) holds for each denumerable coproduct \( \bigcup b \). Indeed, the definition of the coproduct injections \( j_i : b_i \rightarrow \bigcup b \) shows that there is a unique arrow \( \theta \) which makes the diagram
\[
\begin{array}{c}
\bigcup b \\
\downarrow \theta \\
\bigcup b
\end{array}
\]
commute, and "preserves coproducts" means exactly that \( \theta \) is an isomorphism.

For given \( a, b \), take \( b_i = a^i \) to be the \( n \)-th power defined as in (5) and define a multiplication \( \mu \) on \( \bigcup b \) by juxtaposition \( a \square a \approx a^{n+1} \).

Formally, \( \mu \) is the unique arrow defined by the commutative diagram
\[
\bigcup b \overset{\mu}{\rightarrow} \bigcup b \leftarrow \bigcup b \overset{\lambda}{\rightarrow} \bigcup b
\]
where the vertical map \( \text{can} \) is the canonical map (iterated juxtaposition) given by the coherence theorem for \( B \). If \( \phi \) is that unique map on the coproduct \( \bigcup b \) which makes the square with the coproduct injections \( j_i \) and \( i_k \) commute for all the natural numbers \( m \) and \( n \), the map \( \theta = \theta \) is the composite of two canonical isomorphisms \( \theta \) above (because \( \square a \) is distributive over \( \bigcup b \) and \( \bigcup b \)), and the multiplication \( \mu \) is \( \mu = \phi (\theta \circ \theta) \).
A large but routine diagram (exercise!) shows this \( \mu \) to be associative, in the sense (1). A corresponding unit \( \eta_a : e \to \bigwedge a \) is defined to be the injection \( q : e = a \to \bigwedge a \) of the coproduct. All told, \( \langle \bigwedge a, \alpha, \eta_a \rangle \) is a functor in \( B \). The injection \( q_a = i_a : a \to \bigwedge a \) of the coproduct is an arrow

\[
q_a : a \to U\langle \bigwedge a, \alpha, \eta_a \rangle
\]
to the forgetful functor \( U : \text{Mon}_B \to B \).

This arrow is universal from \( a \) to \( U \). For let \( \langle c, \mu, \eta \rangle \) be any monoid in \( B \) and \( f : a \to b = U(c, \mu, \eta) \) an arrow in \( B \). Then we define an arrow \( f' : \bigwedge a \to \bigwedge b \) as the composite on the bottom of the commutative diagram

\[
\begin{array}{ccc}
\bigwedge a & \xrightarrow{f} & \bigwedge b \\
\downarrow{\mu_a} & & \downarrow{\mu_b} \\
\bigwedge a & \to & \bigwedge b
\end{array}
\]

constructed as follows. First, take \( w \) to be the word of length \( n \) with all parenthens in front, so that \( w_0 = b^n \), by our definition of \( b^n \), then \( \mu_w : c^n \to c \) is the \( n \)-fold product defined in the general associative law (6), \( i_n \) and \( j_n \) are coproduct injections, and the dotted arrows on the bottom are constructed, by universality of the coproducts, so as to make the indicated squares commute (for all \( n \)). A routine large diagram will prove that \( f' \) is a morphism of monoids; by construction \( f' \circ q_a = f \), so \( q_a \) is indeed universal and therefore \( \bigwedge a \) is a free monoid on \( a \), as asserted in the theorem.

The point of this quite formal proof is that it contains many separate instances of the same sort of formality. If \( B = \langle \text{Set}, \times, 1, \ldots \rangle \), this is the standard construction (Corollary II.7.2 of the free monoid on the set \( a \); in this case \( a^n \) is the set of words of length \( n \) spelled in letters of \( a \), and the free monoid is the disjoint union \( \bigwedge a \), with product given by composition. If \( B = \langle \text{K-Mod}, \otimes, K, \ldots \rangle \), this is the standard construction (e.g., Mac Lane [1963 b], p. 179) of the tensor algebra \( \otimes_A A^n \) on the \( K \)-module \( A \). The same construction also gives “differential graded” tensor algebras, free topological monoids, etc.

Exercises

1. Prove if \( B \) has finite products, so does \( \text{Mon}_B \).
2. (Coherence for monoids) Interpret the proposition about the canonical maps \( \mu_\alpha \) for a monoid \( \langle c, \mu, \eta \rangle \) as the following coherence theorem. Consider a graph with vertices the binary words \( w \) and with arrows \( w \to w \) those arrows \( v \to w \) which are 1, \( \mu \), \( \eta \), instances \( a(w', w', w') \) of \( a \), instances of \( \lambda \) and of \( \rho \), and all \( 1 \)-products of such arrows. Prove that any two paths \( w \to (-) \) in this graph have equal compositions, but show that this would not hold when the ending is not \((-) \) but the word \((-) \)-length 2.

3. (a) (Substitution of words in a word) Each word \( w \) of length \( n \) determines a functor \( u_w : W^N \to \text{Set} \). If \( v_1, \ldots, v_n \) are words, show that the word \( u_w(v_1, \ldots, v_n) \) has length the sum of the lengths of the \( v_i \), and that it corresponds (intuitively) to substituting \( v_1, \ldots, v_n \) in \( w \). (b) If \( w = u_w(v_1, \ldots, v_n) \), show that the canonical maps \( \mu_\alpha \) of Proposition 1 have the property that the composite

\[
w' = u_w(v_1, \ldots, v_n) \circ \mu_\alpha(v_1, \ldots, v_n) \to u_w(a, \ldots, a) \to c
\]

is equal to \( \mu_\alpha : w' \to c \). Show that this result includes Proposition 1.

4. Actions

Again, we work in a fixed monoidal category \( B \). A left action of a monoid \( \langle c, \mu, \eta \rangle \) on an object \( a \in B \) is an arrow \( v : c \cdot a \to a \) of \( B \) such that the diagram

\[
\begin{array}{ccc}
\text{c} & \xrightarrow{v} & \text{a} \\
\text{c} & \xrightarrow{\mu} & \text{c} \circ \text{c} \\
\end{array}
\]

commutes. For example, \( c \) acts on itself by the map \( \mu_0 : \square c \to c \); this is the “left regular representation” of \( c \). A morphism \( f : v \to v' \) of left actions of \( c \) is an arrow \( f : a \to a' \) in \( B \) such that \( v'(\square f) = v f : c \cdot a \to a' \). With these morphisms as arrows, the left actions \( v \) for a fixed monoid \( c \) form a category \( \text{Act} \). These definitions clearly include familiar cases: an action of an ordinary monoid on a set, a left \( R \)-module regarded as an action of the ring \( R \) on an abelian group, and similarly with rings replaced by any \( k \)-algebra, or \( DG \)-algebras.

There is a forgetful functor \( \text{Act} \to B \), defined by \( \langle v : c \cdot a \to a \rangle \to a \); it has a left adjoint which sends each \( b \in B \) to \( c \cdot b \), with action defined by the composite

\[
c \cdot (c \cdot b) \to (c \cdot c) \cdot b \to c \cdot (c \cdot b)
\]

Right actions \( a \cdot : \square c \to c \) of \( c \) are defined similarly, and commuting left- and right actions of \( c \) on \( a \) may be defined to parallel the usual bimodules (left and right \( R \)-modules).

Exercises

1. (Dubuc [1970], Prop II.1.1) Let \( \langle T, \eta, \mu \rangle \) be a monad in a category \( X \). Show that the monad \( T \) has an action on an endofunctor \( S : X \to X \) if and only if \( S \) can be lifted to the category \( T \)-algebras as \( S = T \cdot S \). Show that these actions correspond one-to-one to the liftings \( S : X \to X \).
2. Let a small strict monoidal category \( B \) (as a monoid in \( \langle \text{Cat}, \times, \ldots \rangle \) act on a category \( C \). Define then the action of a monoid in \( B \) on an object in \( C \), and use this to extend the result of Exercise 1 to the case of functors \( S : A \to X \) from any category \( A \).
3. Describe the actions of a $K$-coalgebra.
4. If $B$ has coproducts preserved by all functors $a \square -$, show that $\mathcal{Lact}$ has coproducts preserved by the forgetful functor to $B$.
5. If the base category $B$ has finite products, so does the category $\mathcal{Lact}$, in such a way that the projections $a \times d \to a$, $d$ of the product (in $B$) become morphisms of actions (in $\mathcal{Lact}$).
6. (Generalization of the tensor product of a right module by a left module.) If $B$ has coequalizers, $c$ is a monoid, $a : b \square c \to b$ is a right action, and $v : c \square a \to a$ a left action, construct a "tensor product" $b \square a \in B$ as the coequalizer of two maps $b \square (c \square a) \to b \square a$ given by the actions, and prove $\square_1$ a functor $\square : \mathcal{Lact} \times \mathcal{Lact} \to B$.
7. (Coherence result for an action.) Given a left action $v : c \square a \to a$ of a monoid $c$, describe the properties of canonical maps $w_n : w_n \cdot a$, where $w$ is any word of length $\geq 1$ with "last argument" $(-)$ (define what this means), while $w_n$ results from substituting $a$ for the last argument and $c$ for all the other arguments in $w$.

5. The Simplicial Category

We now describe a particular strict monoidal category $\mathcal{A}$ which plays a central role in topology and also provides a "universal" monoid.

This category $\mathcal{A}$ has as objects all finite ordinal numbers $n = \{0, 1, \ldots, n-1\}$ and as arrows $f : n \to n'$ all (weakly) monotone functions; that is, all functions $f$ such that $0 \leq i \leq n$ implies $f(i) \leq f(j)$. In this category, the ordinal number $0$ is initial, while the number $1$ is terminal. **Ordinal addition** is a bifunctor $+ : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$, defined on ordinals $n, m$ as the usual (ordered) sum $n + m$ and on arrows $f : n \to n'$, $g : m \to m'$ as

$$(f + g)(i) = \begin{cases} f(i), & i = 0, \ldots, n-1 \\ g(i-n), & i = n, \ldots, n+m-1. \end{cases}$$

(Thus the function $f + g$ is just $f$ and $g$ placed "side by side"). Moreover, $\langle \mathcal{A}, +, 0 \rangle$ is a strict monoidal category. Since $1$ is terminal in $\mathcal{A}$, there are unique arrows $\mu : 2 \to 1, 0 : 0 \to 1$; for the same reason, these arrows form a monoid $\langle 1, \mu, \eta \rangle$ in $\mathcal{A}$. It is "universal" in the following sense.

**Proposition 1.** Given a monoid $\langle c, \mu, \eta \rangle$ in a strict monoidal category $\langle B, \square, e \rangle$, there is a unique morphism $F : \langle \mathcal{A}, +, 0 \rangle \to \langle B, \square, e \rangle$ such that $F1 = c, F\mu = \mu'$ and $F\eta = \eta'$, as in the figure

\[
\begin{array}{c}
0 \to f \to 1 \to 2 + 1 \\
\downarrow \quad \downarrow \quad \downarrow \\
\varepsilon \to \eta' \to c \square c,
\end{array}
\quad \begin{array}{c}
\langle \mathcal{A}, +, 0 \rangle \\
\varepsilon' \quad \mu'
\end{array}
\quad \begin{array}{c}
\langle B, \square, e \rangle
\end{array}
\]

The proof depends on showing that the arrows of $\mathcal{A}$ are exactly the iterated formal products (for the binary product $\mu$). In detail, write $\mu(k^{(a)} : k \to 1$. Thus $\mu^{(a)} = \eta, \mu^{(a)}$ is the identity, $\mu^{(2)} = \mu : 2 \to 1$,

$$\mu^{(3)} = \mu(\mu + 1) = \mu(1 + \mu) : 3 \to 1,$$

and so on. Since $1$ is terminal in $\mathcal{A}$,

$$\mu^{(0)}(\mu^{(k)}) + \ldots + \mu^{(k)} = \mu^{(k)},$$

(2)

(This is the "general associative law"). On the other hand, if $f : m \to n$ is any arrow of $\mathcal{A}$, let $m_i$ be the (ordinal) number of elements in the subset $f^{-1}i$ of $m$; then

$$f = \mu^{(m_0)} + \mu^{(m_1)} + \ldots + \mu^{(m_{n-1})},$$

$$\sum_{i=0}^{n-1} m_i = m$$

(note that some of the $m_i$ may be zero). This shows that any $f$ is a sum of iterated products constructed from $\mu$ and $\eta$.

Now consider the functor $F$ required in the Proposition. Since $F(1) = c$ and $F$ is to be a morphism of monoidal categories, $F$ must have $F\eta = \eta'$; this determines the object function of $F$. Next, $F\mu = \mu'$ and $F\eta = \eta'$ imply that $F\mu^{(k)} = \mu^{(k)}$; the representation (3) of any arrow $f$ of $\mathcal{A}$ then determines the arrow function $Ff$ of $F$. Thus $F$ is unique. It remains only to show that the object and arrow functions so defined give a functor. But in $\mathcal{A}$, composites are given by (2), which corresponds exactly to the general associative law valid in $B$. q.e.d.

This universal property gives a complete characterization of $\mathcal{A}$: Its objects form the free monoid generated (under $+$) by $1$; its arrows are generated by additions and compositions from $\mu : 2 \to 1$ and $\eta : 0 \to 1$, using the associative law for $\mu$ and the left and right unit laws for $\eta$ and $\mu$.

There is another description of the arrows of $\mathcal{A}$, which starts by observing that a monotone function $f : n \to n'$ can be factored as $g \cdot h$ where $h : n \to n''$ is surjective and monotone, $g : n'' \to n'$ is monotone and injective. Moreover, this injective function $g$ will be determined just by giving the image of $g$, which is a subset of $n''$ ordinals in the set $n'$. In particular, there are exactly $n + 1$ injective monotone functions $n \to n + 1$; namely, for $i = 0, \ldots, n$, the injective monotone function $\delta_i : n \to n + 1$ whose image omits $i$, thus

$$\delta_i : n \to n + 1, \quad \delta_i(0, \ldots, n-1) = \{0, \ldots, i, \ldots, n\}.$$

(4)

where $\delta_0$ on the right indicates that $i$ is to be omitted. We display all these arrows (omitting the superscripts $n$) as

$$0 \to \delta_0 \to 2 \to \ldots \to n \to n + 1.$$

(5)
The Simplicial Category

On the other hand, a monotone \( h : n \to n' \) which is surjective is determined by the subset \( \{ j \mid h(j) = h(j+1), 0 \leq j \leq n-2 \} \) of those \( n - n' \) arguments \( j \) at which \( h \) does not increase. In particular, there are \( n \) such arrows \( n+1 \to n' \); for \( i = 0, \ldots, n-1 \) they are

\[
\sigma^i_n : n+1 \to n, \quad \sigma^i_n(i) = \sigma^i_{n+1}(i+1).
\]

We display them (without superscripts) as

\[
0 \quad 1 \quad 2 \quad 3 \quad \ldots \quad \sigma_0, \ldots, \sigma_{n-1} : n+1 \to n.
\]

These arrows may also be expressed in terms of \( \mu \) and \( \eta \). Indeed \( \delta_0 : 0 \to 1 \) is \( \eta \), \( \sigma_0 : 2 \to 1 \) is \( \mu \), and the definitions show that

\[
\delta^i_n = 1 + \eta + 1_{n-i}, \quad \sigma^i_n = 1 + \mu + 1_{n-i-1}, \quad i = 0, \ldots, n,
\]

(8)

(9)

**Lemma.** In \( \Delta \), any arrow \( f : n \to n' \) has a unique representation

\[
f = \delta_{i_1} \cdot \delta_{i_2} \cdots \delta_{i_k} \cdot \sigma_{j_1} \cdot \cdots \cdot \sigma_{j_l},
\]

(10)

where the ordinal numbers \( h \) and \( k \) satisfy \( n-k=h=n' \), while the strings of subscripts \( i \) and \( j \) satisfy

\[
n' > i_1 > \cdots > i_k \geq 0, \quad 0 \leq j_1 < \cdots < j_l < n-1.
\]

**Proof.** By induction on \( i \in n \), any monotone \( f \) is determined by its image, a subset of \( n' \), and by the set of those \( j \in n \) at which it does not increase \( \{ f(j) = f(j+1) \} \). Putting \( i_1, i_2, \ldots, i_k \), in reverse order, for those elements of \( n' \) not in the image and \( j_1, j_2, \ldots, j_l \) in order, for the elements \( j \) of \( n \) where \( f \) does not increase, it follows that the functions on both sides of (10) are equal.

In particular, the composite of any two \( \delta \)'s or \( \sigma \)'s may be put into the canonical form (10). This yields the following list of three kinds of identities on these binary composites

\[
\delta \delta = \delta_{j+1} \delta \quad i \leq j
\]

(11)

\[
\sigma \sigma = \sigma_{j+1} \quad i \leq j
\]

(12)

\[
\sigma \delta = \sigma_{j+1} \quad i < j
\]

(13)

These identities may be verified directly. For example, (11) asserts that \( \delta_{i+1} \cdot \delta_{i} = \delta_{i+1} \cdot \delta_i \) for any \( j \leq n \); one checks that each side of this equation is a monotone injection, and that both sides have the same image.

Proposition 2. The category \( \Delta \), with objects all finite ordinals, is generated by the arrows \( \delta^i_n : n \to n+1 \) and \( \sigma^i_n : n+1 \to n \) subject to the relations (11), (12), and (13).

**Proof.** These relations suffice to put any composite of \( \delta \)'s and \( \sigma \)'s into the unique form (10) of the Lemma.

The category \( \Delta \) has a direct geometric interpretation by affine simplices, which give a functor

\[
\Delta : \Delta \to \text{Top}
\]

(14)

representing \( \Delta \) as a subcategory of \( \text{Top} \). On objects \( n \) of \( \Delta \), take \( \Delta_n \) to be the empty topological space, and \( \Delta_{n+1} \) to be the "standard" \( n \)-dimensional affine simplex – the subspace of Euclidean \( \mathbb{R}^{n+1} \) consisting of the following points

\[
\Delta_{n+1} = \{ p = (t_0, \ldots, t_n) | t_0 \geq 0, \ldots, t_n \geq 0, \sum t_i = 1 \}.
\]

Here the non-negative real numbers \( t_0, \ldots, t_n \) are the barycentric coordinates of the point \( p \in \Delta_{n+1} \). On arrows \( f : n+1 \to m+1 \), \( \Delta_f : \Delta_{n+1} \to \Delta_{m+1} \) is the (affine) map defined by

\[
\Delta_f(t_0, \ldots, t_n) = (s_0, \ldots, s_m), \quad s_j = \sum_{i=j} \gamma_i t_i.
\]

Note carefully that (in this notation) \( J_{n+1} \) has dimension \( n \) and \( n+1 \) vertices, while \( \Delta_f \) is the (unique) affine map which sends the vertex \( i \) of \( \Delta_{n+1} \) to the vertex \( f(i) \) of \( \Delta_{m+1} \); for example, \( \delta_i : \Delta_{n+1} \to \Delta_{n+2} \) is that affine map which sends the \( n \)-simplex \( \Delta_{n+1} \) to that \( n+1 \)-dimensional face of \( \Delta_{n+2} \) which is opposite vertex number \( i \). Geometrically, the "boundary" of a tetrahedron \( \Delta_4 \) consists of the four triangular faces which are the images of \( \Delta_3 \) under \( \delta_0, \delta_1, \delta_2, \delta_3 \). Using standard properties of affine geometry (Mac Lane-Birkhoff [1967], Chap. 12) one may verify (exercise) that \( \Delta \) as defined is indeed a functor \( \Delta \to \text{Top} \).

Note that this functor \( \Delta \) sends the ordinal number \( n+1 \) to the \( n \)-dimensional simplex: \( \Delta \) is a subcategory of \( \text{Top} \), but the geometric dimension is one less than the arithmetic one used in \( \Delta \).

By \( \Delta^+ \) we denote the full subcategory of \( \Delta \) with objects all the positive ordinals \( \{ 1, 2, 3, \ldots \} \) (omit only 0). Topologists use this category, call it \( \Delta^+ \), and rewrite its objects (using the geometric dimension) as \( \{ 0, 1, 2, \ldots \} \). Here we stick to our \( \Delta \), which contains the real 0, an object which is necessary if all face and degeneracy operations are to be expressed, as in (3), in terms of binary product \( \mu \) and unit \( \eta \).

Contravariant functors on the category \( \Delta^+ \to \text{Set} \) are traditionally known as "simplicial sets". Thus, a simplicial object \( S \) in a category \( X \) is defined to be a functor \( S : (\Delta^+)^{op} \to X \), and a morphism \( S \to S' \) of simplicial objects is a natural
transformation \( \theta : S \to S' \). If we write this functor \( S \) as

\[ n+1 \to S_n, \quad \delta_i \mapsto d_i, \quad \sigma_j \mapsto s_j, \]

so that \( S_n \) is in geometric dimension \( n \), then a simplicial object in \( X \) may be described in the traditional (and more complicated) way as a list of \( S_0, S_1, \ldots, S_n, \ldots \) of objects of \( X \) (the object of \( n \)-simplices) with arrows ("face operators") \( d_i : S_n \to S_{n-1} \) for \( i = 0, \ldots, n \) and \( n > 0 \), and arrows ("degeneracies") \( s_j : S_n \to S_{n+1} \) for \( i = 0, \ldots, n, n \geq 0 \) which satisfy the identities dual to (11), (12), and (13).

\[
\begin{align*}
d_i d_{i+1} & = d_i, & i & \leq j & \quad (14^\theta) \\
s_j s_i & = s_j s_i, & i & \leq j & \quad (12^\theta) \\
d_i s_i & = s_{i-1} d_i, & i & < j & = 1, & i = j, j + 1, & \quad (13^\theta)
\end{align*}
\]

An augmented simplicial object in \( X \) is a functor \( S : A^{op} \to X \). A simplicial object \( S \) may be augmented (i.e., extended to a functor \( S' \)) by finding one object \( S_1 \in X \) and one arrow \( \varepsilon : S_0 \to S_1 \) of \( X \) with \( \varepsilon d_0 = \varepsilon d_i : S_i \to S_1 \), thus \( S(\delta_i) = \varepsilon \). Such an arrow \( \varepsilon \) is (traditionally) an augmentation of \( S \).

A simplicial object \( S \) in an abelian category \( A \) (e.g., \( A = \text{Ab} \)) gives homology, via a suitable "boundary" operation. Specifically construct from \( S \) the arrows

\[
S_0 \leftarrow \varepsilon \leftarrow S_1 \leftarrow \cdots
\]

(15)

where the boundary homomorphism \( \partial : S_{n+1} \to S_n \) is the arrow defined as the alternating sum \( \partial = d_0 - d_1 + \cdots + (-1)^{n+1} d_{n+1} \). The relations (11^\theta) on the faces \( d_i \) imply that \( \partial \partial = 0 \). (This means that the diagram (15) is a chain complex in \( A \).) Since \( \partial \partial = 0 \),

\[
\text{Im} \{ \partial : S_{n+1} \to S_n \} \leq \text{Ker} \{ \partial : S_n \to S_{n-1} \}
\]

and we can take the quotient object (see Chap. VIII) to be the \( n \)-th homology of \( S \):

\[
H_n(S) = \text{Ker} \{ \partial : S_{n} \to S_{n-1} \} / \text{Im} \{ \partial : S_{n+1} \to S_{n} \}.
\]

Each augmentation of the functor \( S \) yields an augmentation of this chain complex; that is, an object \( S_{-1} \) of \( A \) and an arrow \( \varepsilon : S_0 \to S_{-1} \) with \( \varepsilon d_0 = 0 \), hence an arrow \( H_0(S) \to S_{-1} \).

The singular homology of a topological space is a classical example. Consider the composite functor

\[
A^{op} \times \text{Top} \xrightarrow{\text{hom}(\varepsilon \ldots)} \text{Set} \xrightarrow{\mathcal{Z}} \text{Ab}
\]

where \( \mathcal{Z} : \to \text{Top} \) is the functor described above, while \( \mathcal{Z} \) assigns to each set the free abelian group generated by the elements of that set. This composite determines for each topological space \( X \) an augmented simplicial object \( S = S(X) \) in \( \text{Ab} \). Each arrow \( h \in \text{hom}(d_{n+1}, X) \) is a singular \( n \)-simplex in \( X \), so \( S_{n+1} \) is the free abelian group generated by all such simplices (all finite linear combinations with integral coefficients of singular \( n \)-simplices). The associated chain complex is the singular chain complex of the space \( X \), with its homology the singular homology (see e.g. Mac Lane [1963] Chap. II).

We may summarize the protean aspects of \( A \) thus:

(a) \( A \) is the category of finite ordinal numbers, hence a full subcategory of the category of all (linearly) ordered sets.

(b) \( A \) is a full subcategory of \( \text{Cat} \), if we interpret each ordinal \( n \) as a category (finite preorder); the objects of \( A \) are the categories \( 0, 1, 2, 3, \ldots \).

(c) \( A \) is the strict monoidal category containing the universal monoid, its arrows are all "iterated multiplications" \( \mu^{(0)} + \cdots + \mu^{(m-1)} \).

(d) \( A \) is a subcategory of \( \text{Top} \), consisting of the standard ordered simplices (one for each dimension), with order preserving affine mappings.

The simplicial objects defined via \( A \) provide a means of treating many questions in algebraic topology, especially those dealing with homology, CW-complexes, Eilenberg-Mac Lane spaces, and cohomology operations. This line of development is presented in May [1967], Lamotke [1968], and Gabriel-Zisman [1967], the last presentation making full use of categorical techniques.

**Exercises**

1. In \( A \), show that an arrow \( f : n \to n' \) is monic (or epi) if and only if the function \( f \) is injective (respectively, surjective).

2. (a) Show that the subcategory \( A_{\text{mon}} \subset A \) of all monics in \( A \) is generated by the arrows \( \delta_i \), subject to the relations (11).

(b) Show that every arrow in \( A_{\text{mon}} \) is uniquely an iterated sum of \( n : 0 \to 1 \) and \( i : 1 \to 1 \).

3. (a) Show that the subcategory \( A_{\text{epi}} \subset A \) of all epics in \( A \) is generated by the arrows \( \sigma_j \), subject to the relations (12).

(b) A semigroup \( \langle c, \mu \rangle \) in a strict monoidal category \( \langle C, \otimes, e \rangle \) is an object \( c \) with an arrow \( \mu : c \otimes c \to c \) which is associative, that is, \( \mu(\mu \otimes 1) = \mu(1 \otimes \mu) \).

Show that \( 2 \to 1 \) is a universal semigroup in \( A_{\text{epi}} \).

4. Show that the category of simplicial objects in \( \text{Set} \) is small-complete.

6. Monads and Homology

Monads and their duals, the comonads, play via \( A \) a central role in homological algebra, as we may now briefly indicate. Let \( L = \langle L, e, \delta \rangle \) be a comonad in a category \( A \); in other words \( L : A \to A \) is an endo-


functors. cohomology various functors; derived construct a of algebra, homological sense A

This amounts to saying that \( \langle L, e, \delta \rangle \) is a comonoid in the strict monoidal category \( A^4 \) of endofunctors of \( A \), where the functor \( \Box \) (multiplication) is composition.

Now \( \Delta \) contains the universal monoid \( \langle 1, 0, 1 + 1 \rangle \), so \( A^e \) contains the universal comonoid \( \langle 1, 1, 0, 1 + 1 \rangle \). Thus, by the dual of Theorem 5.1, any comonoid in a strict monoidal category \( \langle B, \Box, e \rangle \) determines a unique morphism \( A^e \to B \) of monoidal categories, carrying the universal comonoid to the given one. This morphism \( A^e \to B \) is an augmented simplicial object in \( B \) (and \( A^* \), \( A \), \( A \) is a simplicial object).

In particular, each comonad \( \langle L, e, \delta \rangle \) in \( A \), as a comonoid in the functor category \( A^4 \), determines an augmented simplicial object (functor) \( A^e \to A^4 \), with

\[
\langle L, 0, 1 \rangle \xrightarrow{\delta} \langle L, 1 \rangle \xrightarrow{e} \langle L, L \rangle.
\]

Thus \( n \mapsto L^n = L \to \cdots \to L \), \( e \) is the augmentation, \( \delta = \delta_0 : L \to L^2 \) is the degeneracy arrow, and the faces and degeneracies in higher dimensions are given by the duals of the equations (8) and (9) of \( \S \) 5 (which express \( \delta \) and \( \sigma \) in terms of \( \mu \) and \( \eta \)):

\[
\delta_i^0 = L^i : L^i \to L^2, \quad i = 0, \ldots, n,
\]

(2)

\[
\delta_i^0 = L^i \delta : L^i \to L^{i+1}, \quad i = 0, \ldots, n - 1.
\]

(3)

The whole simplicial object has the form

\[
Smp L = \{ L \leftarrow d_0 \xrightarrow{d_1} L^2 \leftarrow L^3 \cdots \xrightarrow{d_i} L^i \cdots \xrightarrow{d_n} L^n \}.
\]

Now suppose that \( A \) is an Ab-category (e.g., an abelian category, or that we have applied to \( Smp L \) a functor to some Ab-category). The simplicial identities on the face operations \( d_i \) then show that the alternating sums

\[
\delta \delta = 0,
\]

so are the boundary morphisms of a chain complex called \( L^e \),

\[
L^e : L \xrightarrow{e} L^2 \xrightarrow{\delta} \cdots \xrightarrow{\delta} L^n \xrightarrow{\delta} \cdots
\]

with an augmentation \( e : L \to a \). This complex is a standard “resolution” of an abelian category, and so may be used to construct derived functors; in particular, various cohomology functors.
The degeneracy operators \( s_i : \mathbb{Z}(\Pi)^{n} \to \mathbb{Z}(\Pi)^{n+1} \), as determined by \( \delta \) according to (3), are the \( \Pi \)-module maps
\[
s_i(x_1 \cdots x_n) = x_1 \cdots x_{i-1} 1 x_{i+1} \cdots x_n \quad 0 \leq i \leq n - 1.
\]
Since \( \Pi - \text{Mod} \) is already an abelian category, this (augmented) simplicial object determines an augmented chain complex in \( \Pi - \text{Mod} \) of the form
\[
\mathbb{Z} \leftarrow \mathbb{Z}(\Pi) \leftarrow \mathbb{Z}(\Pi)^{2} \leftarrow \cdots \leftarrow \mathbb{Z}(\Pi)^{n} \leftarrow \cdots
\]
This is a “free resolution” of the trivial \( \Pi \)-module \( \mathbb{Z} \); it is, in fact, the standard resolution used to define the homology and cohomology of the group \( \Pi \). (Mac Lane [1963], Theorem IV.5.1)

The homology of \( \Pi \) is obtained from the resolution as follows. Take a \( \Pi \)-module \( A \) and the corresponding functor \( \hom_{\Pi}(-, A) : (\Pi - \text{Mod})^{op} \to \text{Ab} \), where \( \hom_{\Pi}(-, -) \) denotes the abelian group of \( \Pi \)-module morphisms. Apply this functor to the chain complex above (dropping the augmentation \( \mathbb{Z}(\Pi) \to \mathbb{Z} \)) to get a “cochain complex
\[
\hom_{\Pi}(\mathbb{Z}(\Pi), A) \to \hom_{\Pi}(\mathbb{Z}(\Pi)^{2}, A) \to \cdots
\]
with coboundary \( \delta = \hom_{\Pi}(\delta, A) \). The cohomology groups of this complex are exactly the cohomology groups \( H^{n}(\Pi, A) \) of the group \( \Pi \) with coefficients in \( A \). The formulas for \( \delta \) above give \( \delta \) explicitly. Thus, for example, \( H^{0}(\Pi, A) = \{ a | a \in A \text{ and } xa = a \text{ for all } x \} \); \( H_{1}(\Pi, A) \) is the group of “crossed homomorphisms” \( \Pi \to A \) modulo the principal crossed homomorphisms, and \( H^{2}(\Pi, A) \) is the group of all group extensions of the additive group \( A \) by the multiplicative group \( \Pi \), with operations (conjugation) given by the \( \Pi \)-module structure of \( A \) (Mac Lane [1963], IV.2, IV.3).

The higher cohomology groups of groups appear in obstruction problems (Mac Lane [1963], IV.8), in the theory of the \( K(\Pi, 1) \) spaces in topology (Mac Lane [1963], IV.11), and class field theory (Cassels–Fröhlich [1967]).

The homology of \( \Pi \) with coefficients in a right-\( \Pi \)-module \( C \) is found in a similar way: To the standard resolutions apply not the functor \( \hom_{\Pi}(-, A) \) but the (covariant, additive) functor \( C \otimes_{\Pi} - : \Pi - \text{Mod} \to \text{Ab} \). The homology of the resulting chain complex in \( \text{Ab} \) is the homology \( H_{n}(\Pi, C) \) of \( \Pi \) with coefficients in \( C \). For example (Mac Lane [1963], Prop. X.5.2)
\[
H_{0}(\Pi, Z) = Z, \quad H_{1}(\Pi, Z) = \Pi/\{ [\Pi, \Pi] \};
\]
the latter is the factor commutator group of \( \Pi \).

### 7. Closed Categories

The ideas broached in this chapter have extensive further developments which we shall indicate briefly. First, a monoidal category \( B \) is said to be symmetric when it is equipped with isomorphisms
\[
\gamma_{a,b} : a \bigcirc b \cong b \bigcirc a,
\]
natural in \( a, b \in B \), such that the diagrams
\[
\begin{align*}
\gamma_{a,b} \circ \gamma_{b,c} & = 1, \quad \alpha_{a} = \lambda_{a} \circ \gamma_{a,c} : b \bigcirc b \cong b, \\
(a \bigcirc (b \bigcirc c)) & \to (a \bigcirc b) \bigcirc c \to c \bigcirc (a \bigcirc b) \\
(a \bigcirc (c \bigcirc b)) & \to (a \bigcirc c) \bigcirc b \to (c \bigcirc a) \bigcirc b
\end{align*}
\]
all commute. This selection of conditions suffices (Mac Lane [1963b]) to prove that “all” such diagrams commute, much as in the coherence theorem of §2 above. Monoidal categories \( (B, \bigcirc, e, \cdot) \), where \( \bigcirc \) is the categorical product or coproduct, are automatically symmetric when \( \gamma : a \times b \cong b \times a \) is taken to be the (canonical) isomorphism which commutes with the projections.

A closed category \( V \) is a symmetric, monoidal category in which each functor \( - \bigcirc b : V \to V \) has a specified right adjoint \( \bigcirc_{b} : V \to V \). For example, \( \langle \text{Ab}, \bigcirc, \cdot \rangle \) is closed; the adjoint is given for abelian groups \( A \) and \( B \) as \( A^{\#} = \text{hom}(B, A) \), the abelian group of all morphisms \( B \to A \). Similarly, \( \langle \text{K-Mod}, \otimes, \cdot \rangle \) is closed for any commutative ring \( K \). The cartesian closed categories, such as \( \text{Set} \) and \( \text{Cat} \), are also closed categories in this sense. In all these cases, the functor \( \bigcirc_{b} : V \to V \) is a sort of “internal hom functor”.

An \( Ab \)-category (and in particular, an abelian category) has already been described (§1.8) as a category with “hom-sets” in \( \text{Ab} \). Similarly, one can describe “categories” with “hom-sets” in any monoidal category \( B \): A set \( R \) of “objects” \( r, s, t \); to each pair of objects \( r, s \) an object \( R(r,s) \in B \); to each ordered triple an arrow (composition!)
\[
R(s,t) \circ R(r,s) \to R(r,t)
\]
in \( B \); to each object \( r \), an arrow \( e \to R(r,r) \) in \( B \) (unit!). These data are subject to the usual associativity and unit axioms on composition. The result is called a \( B \)-category, a \( B \)-based category, or a category relative to \( B \) — and often, replacing the letter \( B \) by \( V \), a \( V \)-category. But observe that this structure \( R \) is not yet a category in the ordinary sense; it has only hom-objects \( R(r,s) \) and not hom-sets. These can be obtained only applying to the hom-objects \( R(r,s) \) a suitable functor \( U : B \to \text{Set} \), say \( U = B(e, -) \), to get hom-sets \( UR(r,s) \). When there are such hom-sets,
one says that the ordinary category \( UR \) has been “enriched” by the objects \( R(r, s) \in B \).

Practically all the basic theory of categories applies to enriched categories, provided that the basic category \( B \) is not just monoidal, but closed. This development (for a presentation, see Dubuc [1970] and references there) may provide a powerful method of treating at one time the cases of ordinary categories, additive categories based on closed categories of chain complexes (for relative homological algebra), and categories based on a suitable cartesian closed variant of \( \text{Top} \).

8. Compactly Generated Spaces

A convenient category of topological spaces should be cartesian closed. The familiar adjunction which makes \( \text{Set} \) cartesian closed,

\[
\text{Set}(X \times Y, Z) \cong \text{Set}(X, \text{Set}(Y, Z)), \quad Z^Y = \text{Set}(Y, Z),
\]

which sends each \( f : X \times Y \to Z \) to \( f^* : X \to Z^Y \), with \( (f^* x) y = f(x, y) \), may be considered also for topological spaces \( X, Y, \) and \( Z \). We obtain a topological space \( \text{Cop}(Y, Z) \) by imposing on the set \( \text{Top}(Y, Z) \) of all continuous maps \( Y \to Z \) the compact open topology: A subbase for the open sets consists of the sets \( N(C, U) \) where \( C \) is any compact subset of \( Y \), \( U \) any open subset of \( Z \), and \( N(C, U) \) consists of all those continuous \( h : Y \to Z \) for which \( hC \subseteq U \). A standard argument (which we will not need) shows that the basic adjunction \( f \mapsto f^* \) of (1) restricts to give an adjunction

\[
\text{Top}(X \times Y, Z) \cong \text{Top}(X, \text{Cop}(Y, Z)),
\]

provided \( Y \) is locally compact Hausdorff.

There have been many attempts to repair this situation for more general spaces \( Y \) by using a variety of other topologies on the function space or other topologies on the product space. The best device is to so restrict the category of topological spaces that the (categorical) product \( X \times Y \) (with its intrinsic topology as a product) does always have a right adjoint (which will be a function space with a uniquely determined topology).

A topological space \( X \) is compactly generated when each subset \( A \subseteq X \) which intersects every compact subset \( C \) of \( X \) in a closed set is itself closed. By \( \text{CGHaus} \) we denote the category with objects all compactly generated Hausdorff spaces (= Kelley spaces), with arrows all continuous functions \( X \to X' \).

**Proposition 1.** \( \text{CGHaus} \) is a full coreflective subcategory of \( \text{Haus} \).

It is a full subcategory by definition. To each Hausdorff space \( Y \) we construct a compactly generated space \( K \) with the same points as \( Y \) (the “Kelleyfication” of \( Y \)) by requiring that \( A \subseteq Y \) be closed in \( K \) if and only if \( A \cap C \) is closed in \( Y \) for all compact sets \( C \subseteq Y \). Thus all closed sets of \( Y \) are closed in \( K \) if \( K \) is Hausdorff, and the identity function \( \varepsilon : K \to Y \) is continuous. Any continuous map \( f : X \to Y \) from a compactly generated Hausdorff space \( X \) factors as \( f = \varepsilon f' \).

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

where \( f' : X \to K \) is the same function (as \( f \)) and is continuous because \( X \) is compactly generated. This shows that \( \varepsilon \) is universal from \( K \) to \( Y \), so is the counit of an adjunction which makes \( \text{CGHaus} \) coreflective in \( \text{Haus} \), as desired.

The description of \( K \) means also (see Fig. (3)) that a function \( g : Y \to Z \) to a topological space \( Z \) is continuous, on \( K \) as \( g \varepsilon : K \to Z \), if and only if the original \( g \) is continuous on all compact subsets of \( Y \). Observe also that metrizable spaces and locally compact Hausdorff spaces are compactly generated.

**Proposition 2.** \( \text{CGHaus} \) is (small) complete and cocomplete.

**Proof.** The category \( \text{Haus} \) is complete (Proposition V.9.2) and a right adjoint such as \( K \) preserves limits. Hence \( \text{CGHaus} \) is complete. In particular, the product (written \( \square \)) of two spaces \( X \) and \( Y \) in \( \text{CGHaus} \) is obtained from their “ordinary” product \( X \times Y \) in \( \text{Haus} \) as

\[
X \square Y = K(X \times Y).
\]

In other words, the \( \square \)-product of Kelley spaces is the product of the underlying sets, with the Kelleyfication of the usual product topology.

Cocompleteness follows readily. Since any coproduct in \( \text{Haus} \) (= disjoint union) of compactly generated spaces is also compactly generated, it will suffice to construct the coequalizer of a parallel pair \( f, g : Y \to X \) in \( \text{CGHaus} \). Take the coequalizer \( p : X \rightharpoonup Q \) in \( \text{Haus} \) (Prop. V.9.2) and form \( KQ \):

\[
\begin{array}{ccc}
Y & \xrightarrow{f} & K \leftarrow X \\
\varepsilon \downarrow & & \downarrow p \\
Q & &
\end{array}
\]

Since \( \varepsilon : KQ \rightharpoonup Q \) is universal, there is a unique continuous \( p' : X \to KQ \) with \( p' \varepsilon = p \) and \( p' f = p' g \). Since \( p' \) is also a map in \( \text{Haus} \), and \( p \) is the
coequalizer of \( f \) and \( g \) there, there is a continuous \( t : Q \to KQ \) with \( p' = tp \). Then \( p = e'p = etp \), so \( e' = t \) and \( e' = et \). But \( e \) is monic (in Haus), so \( e' = 1 \), and \( e \) is an isomorphism: The coequalizer in Haus lies in CGHaus.

For example, if \( A \) is a subset of a compactly generated Hausdorff space \( X \), then we get an identification space \( X/A \) as a coequalizer in CGHaus (collapse all of \( A \) to a point in CGHaus). It is the largest Hausdorff quotient of the space \( X/A \) (collapsed in Top); its topology is automatically compactly generated.

**Theorem 3.** CGHaus is a cartesian closed category.

For two compactly generated Hausdorff spaces \( X \) and \( Y \) define
\[
X^Y = K(\text{Cop}(Y, X)),
\]
the function space with the Kelleyification of the compact-open topology. Define \( \epsilon : X^Y \to X \) by evaluation; \( \langle f, y \rangle \mapsto fy \). We claim that \( \epsilon \) is continuous; it suffices to prove that \( \epsilon : X^Y \to X \) is continuous on compact sets. Since any compact subset of the product space is contained in the product of its projections, it suffices to show that \( \epsilon \) is continuous on any set of the form \( D \times C \), where \( D \) is compact in \( \text{Cop}(Y, X) \) and \( C \) is compact in \( Y \). Consider \( \langle f, y \rangle \in D \times C \), and let \( U \) be an open set of \( X \) containing \( fy \). Since \( f : Y \to X \) is continuous, there exists a neighborhood \( M \) of \( y \) in \( C \) whose closure satisfies \( f M \subseteq U \). But \( N(M, U) \) is a set of the subbase for \( \text{Cop}(Y, X) \) and \( [N(M, U) \cap D] \times M \) is open in \( D \times C \), contains \( \langle f, y \rangle \), and is mapped by \( \epsilon \) into \( U \). This proves \( \epsilon \) continuous.

It remains to show \( \epsilon \) universal from \( - \boxtimes Y \to X \). So consider any map \( h : Z \boxtimes Y \to X \) in CGHaus. Then we construct \( k : Z \to \text{Set}(Y, X) \) as \( k = h^t \); that is, so that \( (kz)y = h(z, y) \) for all \( z \in Z \) and \( y \in Y \). A direct proof shows that \( kZ : Y \to X \) is continuous; thus \( kZ \in X^Y \). Next, we prove that \( z \mapsto kZ \) is continuous \( Z \to X^Y \). Since \( Z \) is compactly generated, it is enough to show \( Z \to \text{Cop}(Y, X) \) continuous. So let \( N(C, U) \) be one of the open sets for the subbase of the compact-open topology, and suppose that \( kZ \in N(C, U) \); thus \( h(C \times z) \subseteq U \). Since \( U \) is open, \( C \) compact, and \( h \) continuous, there is a neighborhood \( V \) of \( z \) such that \( h(C \times V) \subseteq U \). This implies that \( kV \subseteq N(C, U) \). Therefore \( k \) is continuous.

We now have the commutative diagram
\[
\begin{array}{ccc}
X^Y & \xrightarrow{\epsilon} & X \\
\downarrow & & \downarrow \\
Z \boxtimes Y & \xrightarrow{\epsilon} & Z
\end{array}
\]
by the adjunction in \text{Set}, there is at most one \( k \) with \( e(k \boxtimes 1) = h \), and we have just shown this \( k \) continuous. Therefore \( e \) is universal, and defines the desired adjunction
\[
\text{CGHaus}(Z \boxtimes Y, X) \cong \text{CGHaus}(Z, X^Y).
\]  

Since \( \square \) designates the product in CGHaus, this category is cartesian closed.

This adjunction (6) is a bijection of sets. One also wishes the corresponding homeomorphism
\[
X^2, \square \cong (X)^2
\]
of function spaces. This follows from the adjunction (6) for categorical reasons (Ex. IV.6.3).

This summarizes the basic properties of the category CGHaus. More extensive work (Steenrod [1967] and elsewhere) indicates that it is the convenient category for topological studies; Dubuc and Porta [1971] show that it is appropriate for topological algebra (extensions of the Gelfand duality). All told, this suggests that in Top we have been studying the wrong mathematical objects.

The right ones are the spaces in CGHaus.

9. Loops and Suspensions

For homotopy theory, we consider the category CGHaus of pointed compactly generated Hausdorff spaces – with objects the spaces \( X \in \text{CGHaus} \) with a selected base point, \( *_X \), and with arrows the continuous maps preserving the base point. Let \( X^{*Y} \) be the subspace of \( X^Y \) consisting of all base-point preserving maps. Since it is a closed subspace, it is compactly generated. It has a natural base-point (the continuous function sending all of \( X \) to \( *_X \). In the standard adjunction \( f \mapsto f^* \),
\[
\text{CGHaus}(Z \boxtimes Y, X) \cong \text{CGHaus}(Z, X^Y)
\]
consider on the right the indicated subset: Those \( f' : Z \to X^{*Y} \) which preserve base-point. Thus \( (f'z)_* = *_X \) and \( (f^*z)_* = *_X \); that is, for all
Loops and Suspensions

$z \in Z$ and $y \in Y$

$$f(z,*) = * = f(*,y).$$

These are exactly the continuous functions $f$ which collapse the "wedge" $Z \vee Y = (Z \sqcup *) \vee (Y \sqcup *)$ to a point. The corresponding identification space is called the smash product

$$(Z \sqcup Y)/[(Z \sqcup *) \vee (Y \sqcup *)] = Z \wedge Y$$

(or sometimes $Z \neq Y$). This gives an adjunction

$$\text{CGHaus}_s(Z \wedge Y, X) \cong \text{CGHaus}_s(Z, X \wedge Y).$$

(1)

The circle $S^1$ may be obtained from the closed unit interval $I = \{0 \leq t \leq 1\}$ as the identification space $S^1 = I/\{0,1\}$; we regard it as a pointed space with base point $0 = +1$. The functors $\Sigma$ (reduced suspension) and $\Omega$ (loop space) on $\text{CGHaus}_s$ to $\text{CGHaus}_s$ are defined as

$$\Sigma X = X \wedge S^1, \quad \Omega X = X^{\vee S^1};$$

by the bijection above $\Sigma: \text{CGHaus}_s \rightarrow \text{CGHaus}_s$, $\Omega$ as right adjoint. The points of $\Omega X$ are the loops in $X$ at the base point; that is, the continuous maps $f: I \rightarrow X$ with $f(0) = f(1) = x$. On the other hand, $\Sigma X$ is the cylinder $X \times I$ with top $X \times \{1\}$, bottom $X \times \{0\}$, and generator $x \times I$ all collapsed to a single point, the (new) base point; equivalently it is the double cone over $X (X \times I$ with top and bottom collapsed) and with the generator over $\bullet$ collapsed, as in the figure

$$X = \bigwedge X, \quad \Sigma X = \bigwedge X.$$

For example, $\Sigma S^1 = S^1 \wedge S^1$ is the two sphere $S^2$, $\Sigma^2 S^1$ the $(n+1)$-sphere. The unit $X \rightarrow \Omega\Sigma X$ of the adjunction sends $x \in X$ to the function $\langle x, -\rangle: I \rightarrow \Sigma X$; it has a vivid geometric picture: it sends each point $x \in X$ to that generator of the cone which passes through $x$; this generator is a loop from north pole to south pole = north pole, hence a point of $\Omega\Sigma X$.

By iteration, $\Sigma$ is the left adjoint of $\Omega^n: \text{CGHaus}_s \rightarrow \text{CGHaus}_s$; this adjunction has a unit $X \rightarrow \Omega^n\Sigma X$ which can be written as a composite

$$X \rightarrow \Omega\Sigma X \rightarrow \Omega\Omega\Sigma X \rightarrow \cdots$$

and $\Omega^n$, as a right adjoint, preserves products: $\Omega^n(X \square Y) \cong \Omega^nX \square \Omega^nY$.

These and similar facts can be obtained either by direct topological arguments, or by application of the properties of adjunctions.

Exercises

1. Construct a left adjoint for $\text{Set}_s(S, -): \text{Set}_s \rightarrow \text{Set}_s$.
2. Show that the smash product in $\text{CGHaus}_s$ is commutative and associative up to natural isomorphisms which make $\text{CGHaus}_s$ a symmetric monoidal category with unit the two-point space.
3. In $\text{Top}_s$, show that $- \times X$ does not have a right adjoint (because it does not preserve coproducts).
4. The Path space functor $P: \text{CGHaus}_s \rightarrow \text{CGHaus}_s$ has $P X = X^{\vee S^1}$, where 0 is taken as the base point of the interval $I$. For each path $f \in PX, f(1)$ defines a natural transformation $\pi: P \rightarrow \text{Id}$. Show that $\pi$ can be obtained as the pullback of a diagram $P \rightarrow \text{Id} \rightarrow \ast$. (Classically, $\Omega X$ is the "fibre" of $\pi_2: PX \rightarrow X$.)
5. Describe the counit of the $\Sigma \Omega$ adjunction.

Notes

Monoidal categories were first explicitly formulated by Bénabou [1963, 1964], who called them "catégories avec multiplication" and by Mac Lane [1963b], who called them "categories with multiplication"; the renaming is due to Eilenberg. Coherence theorems were initiated by Stasheff in a 1963 treatment of higher homotopies, by Mac Lane [1963b], and by Epstein [1966], who needed them for a general definition of Steenrod operations. Coherence theorems are undergoing active development; Lambek [1968] found a fascinating connection with the cut-elimination theorems of Gentzen-style proof theory; following his lead, Kelly-Mac Lane [1970] proved a coherence theorem for closed categories. The simplicial category, long implicit in the boundary formulas of algebraic topology, became explicit in the study of Eilenberg-Mac Lane spaces and of the Eilenberg-Zilber theorem about 1950, and played a role in the development of homological algebra (see the notes to Chap. VI). Our discussion of monads and homology is only a slight introduction to the recent proliferation of conceptual schemes for the organization of homological algebra.

Compactly generated spaces first appeared in John Kelley’s 1955 book on General Topology; their convenience for topology was emphasized by Steenrod [1967], Gabriel-Zisman [1967], and others. There are alternative closed categories convenient for topology, notably the quasi-topological spaces due to Spanier.

The suspension $\Sigma$ of a topological space is a tool long used in homotopy theory. The Cartan-Serre attack (about 1950) on the difficult problem of computing the homotopy groups of spheres made essential use of loop spaces and suspension. These constructions originally seemed thoroughly geometric. Thus the natural map $X \rightarrow \Omega\Sigma X$ came from a topological insight, but now appears in conceptual terms, as the unit of an adjunction.
VIII. Abelian Categories

This chapter will formulate the special properties which hold in categories such as \( \mathbf{Ab}, \mathbf{R-Mod}, \mathbf{Mod-R}, \) and \( \mathbf{R-Mod-S}: \) They are all \( \mathbf{Ab}\)-categories (the hom-sets are abelian groups and composition is bilinear), all finite limits and colimits exist, and these limits - especially kernel and cokernel - are well behaved. This leads to a set of axioms describing an "abelian" category. The axioms suffice to prove all the facts about commuting diagrams and connecting morphisms which are proved in \( \mathbf{Ab} \) by methods of chasing elements. We carry the subject exactly to this point, leaving the subsequent development of homological algebra to more specialized treatments.

1. Kernels and Cokernels

Recall (§ I.5) that a null object \( z \) in a category is an object which is both initial and terminal. If \( C \) has a null object, then to any \( a, b \in C \) the unique arrows \( a \to z \) and \( z \to b \) have a composite \( 0 = 0_z : a \to b \) called the zero arrow from \( a \) to \( b \). It follows that any composite with one factor a zero is itself a zero arrow. The null object is unique up to isomorphism, and the notion of zero arrow is independent of the choice of the null.

Let \( C \) have a null object. A kernel of an arrow \( f : a \to b \) is defined to be an equalizer of the arrows \( f,0 : a \to b \). Put more directly, \( k : s \to a \) is a kernel of \( f : a \to b \) when \( f k = 0 \), and every \( h \) with \( fh = 0 \) factors uniquely through \( k \) (as \( h = kh') \)

\[
\begin{array}{ccc}
S & \xrightarrow{k} & a \\
\uparrow k & & \downarrow f \\
C & \xrightarrow{h} & b
\end{array}
\]

Thus any category with all equalizers (or, more generally, with all pullbacks or with all finite limits) and with a zero has kernels for all arrows, and the kernel \( k : s \to a \) of \( f \) is unique, up to an isomorphism of \( s \). Like all equalizers, a kernel \( k \) is necessarily monic (\( kg' = kh' \) implies \( g' = h' \), by the unique factorization requirement in the definition). Hence it is convenient to think of the kernel \( k : s \to a \) as a subobject of \( a \) as an equivalence class of monics \( s \to a \).

For example, in \( \mathbf{Grp} \) the group \( 1 \) with just one element (the identity element) is a null object, and for any two groups the zero morphism \( G \to H \) is the unique morphism which sends all of \( G \) to the identity element in \( H \). The kernel of an arbitrary morphism \( f : G \to H \) of groups is the insertion \( N \to G \) of the usual kernel \( N \), (with \( N = \{ x \in G \mid f(x) = 1 \} \)). Note that \( N \) is a normal subgroup of \( G \), so in \( \mathbf{Grp} \) every kernel is monic but there are monics which are not kernels.

In the category \( \mathbf{Set}_* \) of pointed sets (§ I.7), the one-point set is a null object and the zero map \( P \to Q \) is the function taking all of \( P \) to the base point \( *_Q \) in \( Q \). For any morphism \( f : P \to Q \) of pointed sets, the kernel \( S \to P \) is the insertion of the subset \( S \) of those \( x \in P \) with \( fx = *_Q \), where the base point of \( S \) is identical with the base point of \( P \). Much the same description gives kernels in \( \mathbf{Top}_* \). In \( \mathbf{Grp} \), an epimorphism is determined (up to isomorphism) by its kernel, but this is by no means the case in \( \mathbf{Set}_* \) or in \( \mathbf{Top}_* \).

In any \( \mathbf{Ab}\)-category \( A \), all equalizers are kernels. Indeed, in such a category each hom-set \( A(b,c) \) is an abelian group. Hence, given a parallel pair \( f,g : b \to c \), a third arrow \( h : a \to b \) satisfies \( fh = gh \) if and only if \( (f - g)h = 0 \). Therefore the universal such \( h \) can be described either as the equalizer of \( f \) and \( g \) or as the kernel of \( f - g \). This is the reason one usually deals with kernels and not with equalizers in \( \mathbf{R-Mod}, \mathbf{Ab} \), etc.

The dual notion of cokernel has already been described, in § III.3.

Now suppose that the category \( C \) has a null object \( z \), and kernels and cokernels for all arrows. For each object \( c \in C \), the set \( P_c \) of all arrows \( f \) with codomain \( c \) has a preorder \( \leq \), with \( g \leq f \) defined to mean that \( g \) factors through \( f \) (i.e., that \( g = fg' \) for some arrow \( g' \)). This reflexive and transitive relation \( \leq \) defines a usual equivalence relation \( \equiv \), with \( f \equiv g \) meaning that \( f \leq g \) and \( g \leq f \). The equivalence classes of arrows \( f \in P_c \) under this relation form a partially ordered set, which contains the partially ordered set of subobjects of \( c \) (without the inclusion relation already defined for subobjects in § V.7).

Dually, the set \( Q^c \) of all arrows \( u \) with domain \( c \) is preordered, with \( u \leq v \) when \( v \) factors through \( u \) (i.e., \( v = vu' \) for some \( u' \)).

Now choose a kernel for each arrow \( u \) from \( c \) and a cokernel for each arrow \( f \) to \( c \). Then the definitions of kernel and cokernel state that

\[
f \leq \ker u \Leftrightarrow u f = 0 \Leftrightarrow \coker f \leq u.
\]

These logical equivalences state exactly that the functions

\[
\ker : Q^c \to P_c, \quad \coker : P \to Q^c
\]
Kernels and Cokernels

define a Galois connection from the preorder $Q'$ to the preorder $P'$, as defined in § IV.5. As for any such connection, the triangular identities read

$$\ker(\coker(\ker u)) = \ker u, \quad \coker(\ker(\coker f)) = \coker f.$$  

and $q$ is a kernel if and only if $g = \ker(\coker g)$; these facts are also readily provable directly from the definitions.

If $C$ has a null object, kernels, and cokernels, then any arrow $f$ of $C$ has a canonical factorization

$$f = mq, \quad m = \ker(\coker f).$$  

(2)

**Lemma 1.** If also $f = m'tq'$, where $m'$ is a kernel, then in the commutative square

$$\begin{array}{ccc}
q & \xrightarrow{f} & m' \\
\downarrow & & \downarrow \\
q' & \xrightarrow{m'tq'} & m
\end{array}$$

there is a (unique) diagonal arrow $t$ with $m = m't$ and $q' = tq$. Moreover, if $C$ has equalizers and every monic in $C$ is a kernel, then $q$ is epi.

**Proof.** By assumption, $m' = \ker p'$ where $p' = \coker m'$; take also

$$p = \coker m = \coker f.$$  

Then $p'm' = 0$, so $p'f = p'm'q' = 0$, and $p'$ factors through $p$ as $p' = wp$ for some $w$. Then $p'm = wpm = 0$, so $m$ factors through $m' = \ker p'$ as $m = m't$ for a unique monic $t$. Moreover, $m'q' = m'tq$ and $m'$ is monic, so $q' = tq$. This gives the desired diagram (3).

Next, to prove that $q$ is epi, consider some parallel pair of arrows

$$r, s$$

with $rq = sq$. Then $q$ factors through the equalizer $e$ of $r$ and $s$. If $f = mq = meq'$, then $m' = me$ is monic, hence by

**2. Additive Categories**

An $Ab$-category $A$, as defined in § I.8, is a category in which each hom-set $A(b, c)$ is an additive abelian group (not necessarily small) and composition of arrows is bilinear relative to this addition. Thus each abelian group $A(b, c)$ has a zero element $0 : b \rightarrow c$, called the zero arrow (even though $A$ may not have a null object in the previous sense). Again, a composite with a zero arrow is necessarily zero, since composition is distributive over addition.

**Proposition 1.** The following properties of an object $z$ in an Ab-category $A$ are equivalent: (i) $z$ is initial; (ii) $z$ is terminal; (iii) $1_z = 0 : z \rightarrow z$; (iv) the abelian group $A(z, z)$ is the zero group. In particular, any initial (or any terminal) object in $A$ is a null object.

**Proof.** If $z$ is initial, there is a unique map $z \rightarrow z$, hence $1_z = 0$ and $A(z, z) = 0$. If $1_z = 0$, then $f : b \rightarrow z$ has $f = 1_zf = 0f = 0 : b \rightarrow c$, so there is a unique arrow, namely 0, from $b$ to $z$, and $z$ is terminal. The rest follows by duality.

When there is a null (= initial and terminal) object $z$ in the Ab-category $A$, the unique maps $b \rightarrow z$ and $z \rightarrow c$ are the zero elements of $A(b, z)$ and $A(z, c)$ respectively. Hence the composite $b \rightarrow z \rightarrow c$, which is the zero morphism $0 : b \rightarrow c$, as defined in § 1, is also the zero element of the abelian group $A(b, c)$.

Next we consider products and coproducts in the Ab-category $A$.

**Definition.** A biproduct diagram for the objects $a, b \in A$ is a diagram

$$a \xrightarrow{p_1} c \xleftarrow{p_2} b$$

with arrows $p_1, p_2, i_1, i_2$ which satisfy the identities

$$p_1i_1 = 1_a, \quad p_2i_2 = 1_b, \quad i_1p_1 + i_2p_2 = 1_c.$$  

(2)

**Theorem 2.** Two objects $a$ and $b$ in an Ab-category $A$ have a product in $A$ if and only if they have a biproduct in $A$. Specifically, given a biproduct diagram (1), the object $c$ with the projections $p_1$ and $p_2$ is a product of $a$ and $b$, while, dually, $c$ with $i_1$ and $i_2$ is a coproduct. In particular, two objects $a$ and $b$ have a product in $A$ if and only if they have a coproduct in $A.
Proof. First assume we have the biproduct diagram (1) with the condition (2). Then
\[ p_1 i_2 = p_1 (i_1 p_1 + i_2 p_2) i_2 = 1 \cdot p_1 i_2 + p_1 i_2 \cdot 1 = p_1 i_2 + p_1 i_2; \]
subtracting, \( p_1 i_2 = 0 \); symmetrically \( p_2 i_1 = 0 \). (These are familiar equations for the usual biproduct of modules.) Now consider any diagram \( a \xrightarrow{f} b \). The sum \( h = i_1 f_1 + i_2 f_2 : b \to c \) then has \( p_i h = f_i \); conversely, if \( h' : d \to c \) has \( p_i h' = f_i \) for \( i = 1, 2 \), then
\[ h' = (i_1 p_1 + i_2 p_2) h' = i_1 p_1 h' + i_2 p_2 h' = i_1 f_1 + i_2 f_2, \]
so \( h' = h \). This states that there is a unique \( h : d \to c \) with \( p_i h = f_i \) for \( i = 1, 2 \), so the diagram \( a \xrightarrow{f} b \) is indeed a product. The assignment \( h \mapsto (f_1, f_2) \) is an isomorphism
\[ A(d, c) \cong A(d, a) \oplus A(d, b) \]
of abelian groups, where \( \oplus \) on the right is the direct sum of abelian groups.

Conversely, given a product diagram \( a \xrightarrow{f} a \times b \xrightarrow{p_1, p_2} b \), the definition of this product provides a unique arrow \( i_1 : a \to a \times b \) with components \( p_1 i_1 = 1_a \), \( p_2 i_1 = 0 \) and a unique \( i_2 : b \to a \times b \) with \( p_1 i_2 = 0 \), \( p_2 i_2 = 1_b \). Then
\[ p_1 (i_1 p_1 + i_2 p_2) = p_1 + 0 p_2 = p_1, \quad p_2 (i_1 p_1 + i_2 p_2) = p_2, \]
so \( i_1 p_1 + i_2 p_2 : a \times b \to a \times b \) is the unique arrow with components \( p_1 \) and \( p_2 \), hence is the identity \( 1_{a \times b} \). Thus the given product diagram does indeed yield a biproduct, with (1) and (2).

In special categories, such as \( \text{Ab} \) and \( \text{R-Mod} \), the biproduct is often called a direct sum. Note also that the description of the biproduct diagram is "internal," since it involves only the objects \( a, b, c \), and the arrows between them, while the standard categorical description of the product (or the coproduct) is "external," since it refers to construction of arrows in the whole category.

Given objects \( a, b \in A \), the biproduct diagram (1), if it exists, is determined uniquely up to an isomorphism of the object \( c \). If all such biproducts exist, then a choice of \( c = a \oplus b \) for each pair \( (a, b) \) determines a bifunctor \( \oplus : A \times A \to A \), with \( f_1 \oplus f_2 \) defined for arrows \( f_1 : a \to a' \) and \( f_2 : b \to b' \) either by the equations
\[ p_j (f_1 \oplus f_2) = f_j p_j, \quad j = 1, 2, \]
(i.e., defined as for a product \( \times = \oplus \)) or by the equations
\[ (f_1 \oplus f_2) i_k = i_k f_k, \quad k = 1, 2, \]
that is, as for a coproduct \( \oplus = \bigoplus \), with \( i_1, i_2 \), the injections of the second coproduct. Indeed, the first pair (3) of equations determines \( f_1 \oplus f_2 \), uniquely as the arrow with components \( f_1, f_2 \); then by the defining equations for the second biproduct
\[ (f_1 \oplus f_2) i_k = (i_1 p_1 + i_2 p_2) (f_1 \oplus f_2) i_k = i_k f_k, \]
as in the second pair of equations, and dually.

The conclusion may also be formulated thus: The identification of the product functor \( a \times b \), with mapping function defined by (3), with the coproduct functor \( a \bigoplus b \), mapping function defined by (4), is a natural isomorphism.

Iteration, for given \( a_1, \ldots, a_n \in A \), yields a biproduct \( \bigoplus_j a_j \) characterized (up to isomorphism in \( A \)) by the diagram
\[ a_j \xrightarrow{h_j} a_j \xrightarrow{\pi_j} a_k, \quad j, k = 1, \ldots, n \]
and the equations
\[ i_1 p_1 + \cdots + i_n p_n = 1, \quad p_k i_j = \delta_{k,j} = 0 \quad k+j \quad 1=k+j. \]

Moreover, for given \( c_1, \ldots, c_m \in A \) there is an isomorphism
\[ A(\bigoplus_k c_k, \bigoplus_j a_j) \cong \bigoplus_k A(c_k, a_j) \]
of abelian groups, where \( \Sigma \) denotes the iterated biproduct of abelian groups. This implies that each arrow \( f : \bigoplus_j c_j \to \bigoplus_a a_j \) is determined by the \( n \times m \) matrix of its components \( f_{ik} = p_k f_i : a_i \to c_k \). Composition of arrows is then given by the usual matrix product of the matrices of components. In other words, the equations (5) contain the familiar calculus of matrices (cf. §III.5).

An additive category is by definition an \( A-b \)-category which has a zero object \( 0 \) and a biproduct for each pair of its objects.

Proposition 3. For parallel arrows \( f, f' : a \to b \) in an additive category \( A \),
\[ f + f' = \delta^A(f \oplus f') \delta_A : a \to b, \]
where \( \delta_A : a \to a \times a \) is the diagonal map, \( \delta^A : b \oplus b = b \bigoplus b \to \text{the codiagonal} \).

Here the diagonal is defined by \( p_i \delta = 1_a = p_0 \delta_1 \) and the codiagonal by \( \delta^A i_1 = 1_b = \delta \delta i_2 \). The proof is a direct calculation:
\[ \delta^A (f \oplus f') \delta_A = \delta^A (f \oplus f') (i_1 p_1 + i_2 p_2) \delta_A = \delta^A (f \oplus f') i_1 + \delta^A (f \oplus f') i_2 = \delta^A i_1 + \delta^A i_2 f' = f + f'. \]
Additive Categories

This proposition suggests that the additive structure of $A$ can be derived from the biproduct (cf. Exercise 4).

If $A$ and $B$ are $Ab$-categories, an additive functor $T: A \to B$ is a functor from $A$ to $B$ with

$$T(f + f') = Tf + Tf'$$

for any parallel pair of arrows $f, f': b \to c$ in $A$. It follows that $T0 = 0$.

Since the additive structure of $A$ can be described in terms of the biproduct structure of $A$, this condition (7) can also be reformulated as follows:

**Proposition 4.** If $A$ and $B$ are $Ab$-categories, while $A$ has all binary biproducts, then a functor $T: A \to B$ is additive if and only if $T$ carries each binary biproduct diagram in $A$ to a biproduct diagram in $B$.

**Proof.** Each of the equations $p_i i_1 = 1$, $p_1 i_2 = 1$, and $i_x p_1 + i_2 p_2 = 1$ describing a biproduct in terms of its insertions $i_j$ and projections $p_j$ is preserved by an additive functor; therefore each additive functor preserves biproducts.

Conversely, suppose that $T$ preserves all binary biproducts. Then a parallel pair of arrows $f_1, f_2: a \to d$ has $T(f_1 \oplus f_2) = Tf_1 \oplus Tf_2$ and therefore $T(f_1 + f_2) = Tf_1 + Tf_2$ by the formula (6) for sum in terms of direct sum and the equations $T(\delta_a) = \delta_{Ta}$, $T(\delta_d) = \delta_{Td}$, which follows at once from the definition of the diagonal $\delta$ and the codiagonal $\delta$ in terms of product and coproduct.

Our proposition can also be modified: $T$ is additive if and only if $T$ carries each binary product diagram in $A$ to a product diagram in $B$, or, if and only if it carries each binary coproduct in $A$ to a coproduct in $B$.

Many familiar functors for $Ab$-categories $A$ are additive. For example, if $A$ has small hom-sets each hom-functor

$$A(a, -): A \to Ab, \quad A(-, a): A^{op} \to Ab$$

is additive. If $A$ and $B$ are $Ab$-categories, so is $A \times B$, and the projections $A \times B \to A$, $A \times B \to B$ of this product are additive functors. The tensor product of abelian groups is a functor $Ab \times Ab \to Ab$, additive in each of its arguments, and so is the torsion product.

**Exercises**

1. In any additive category $A$, show that the canonical map

$$\kappa: a_1 \oplus \cdots \oplus a_n \to a_1 \times \cdots \times a_n$$

(defined in $\S$ III.5) is an isomorphism (This is essentially a reformulation of Theorem 2).

2. Define the corresponding canonical map $\kappa$ of an infinite coproduct to the corresponding infinite product, and show by an example that it need not be an isomorphism in every additive category.

3. Abelian Categories

**Definition.** An abelian category $A$ is an $Ab$-category satisfying the following conditions

(i) $A$ has a null object,

(ii) $A$ has binary biproducts,

(iii) Every arrow in $A$ has a kernel and a cokernel,

(iv) Every monic arrow is a kernel, and every epi a cokernel.

The first two conditions insure that $A$ is an additive category, as described in $\S$ 2. Instead of requiring a null object in (i), we could by Proposition 2.1 require a terminal object or an initial object. Instead of requiring all biproducts $a \oplus b$, we could require all products $a \times b$ or all binary coproducts.

With (i) and (ii), the existence of kernels in condition (iii) implies that $A$ has all finite limits. Indeed, the equalizer of $f, g: a \to b$ may be constructed as the kernel of $f - g$. (i) and (ii) give finite products, and finite products and equalizers give all finite limits. Dually, the existence of cokernels implies the existence of all finite colimits.
Abelian Categories

Condition (iv) is powerful. It implies, for example, that any arrow $f$ which is both monic and epic is an isomorphism. For $f: a \to b$ monic means $f = \ker g$ for some $g$, hence $gf = 0 = 0f$. But $f$ is epic, so cancels to give $g = 0: b \to c$, and the kernel of $g = 0$ is equivalent to the identity of $b$, hence is an isomorphism.

The categories $\mathsf{R-Mod}$, $\mathsf{Mod-R}$, $\mathsf{Ab}$ (and many others) are all abelian, with the usual kernels and cokernels. If $A$ is abelian, so is any functor category $A^J$, for arbitrary $J$. Specifically, if $S, T: J \to A$ are any two functors, the set $\text{Nat}(S, T) = A^J(S, T)$ of all natural transformations $\alpha: S \Rightarrow T$ is an abelian group, with addition defined termwise as $(\alpha + \beta)_j = \alpha_j + \beta_j: S_j \to T_j$ for each $j \in J$. The functor $N: J \to A$ everywhere equal to the null object of $A$ is the null functor in $A^J$; the biproduct $S \oplus T$ of two functors is defined termwise, as $(S \oplus T)_j = S_a \oplus T_a$, and the kernel $K$ of a natural transformation $\alpha: S \to T$ is defined termwise, so that for each $j$, $K_j = \ker \alpha_j$ is the kernel of $\alpha_j$. All the axioms follow, to make $A^J$ abelian.

**Proposition 1.** In an abelian category $A$, every arrow $f$ has a factorization $f = me$, with $m$ monic and $e$ epic: moreover,

$$m = \ker(\text{coker } f), \quad e = \text{coker}(\ker f).$$

(1)

Given any other factorization $f' = m'e'$ with $m'$ monic and $e'$ epic and a commutative square

$$
\begin{array}{ccc}
& f & \\
\alpha & \downarrow & \beta \\
& e' & \\
\end{array}
$$

(2)

as shown at the left above, there is a unique $k$ with $e'g = ke$, $m'k = hm$.

(i.e., with the squares at the right commutative).

**Proof.** To construct such a factorization of $f$, take $m = \ker(\text{coker } f)$. Since $\ker(\text{coker } f) \cdot f = 0$, $f$ factors as $f = me$ for a unique $m$, and by Lemma 1 of §1, $e$ is epic. Now $m$ is monic, so for any composable $t$, $ft = 0$ if and only if $et = 0$. This implies that $\ker f = ker e$. But $e$ is epic, so the arrow $e = \ker(\text{coker } f) = \text{coker}(\ker f)$. We have proved (1).

Now regard $f$ and $f'$ as objects in the arrow category $A^J$; a morphism $(g, h): f \to f'$ is then just a commutative square as in (2) above. Consider the factorizations $f = me$ and $f' = m'e'$, and set $u = \ker f = \ker e$. Then $0 = hfu = m'e'gu$, so $e'gu = 0$, and $e'$ must factor through $e = \ker(u)$ as $e'g = ke$ for a unique $k$. Then also $m'ke = m'e'g = hme'$, so $m'k = hm$, and both squares commute in the rectangle of (2).

This completes the proof. The second part shows that any morphism $(g, h): f \to f'$ must carry a factorization of $f$ to a factorization of $f'$, so that the factorization is functorial. In particular, for the identity morphism $(1, 1): f \to f$, this proves that any two monic-epi factorizations $f = me$ and $f = m'e'$ are isomorphic ($k$ an isomorphism above).

From this factorization, we define (the usual) image and cokernel of $f = me: a \to b$ as

$$m = \text{im } f, \quad e = \text{coim } f,$$

(3)

uniquely up to isomorphism. Thus the image $m$ of $f$ is a subobject of its codomain $b$, its cokernel a quotient object of its domain. More generally, if $f = m_t e_t$ with $m_t$ monic, $f$ an isomorphism and $e$ epic, then $m_t = \text{im } f$, $e_t = \text{coim } f$ and $t$ is (the usual) isomorphism of cokernels to image. This is the situation which arises in familiar concrete categories like $\mathsf{Ab}$. If $f: B \to C$ is a morphism in $\mathsf{Ab}$ with kernel a subgroup $K$ of $B$, image a subgroup $S$ of $C$, then $f$ factors as a three-fold composite

$$
B \longrightarrow B/K \longrightarrow S \longrightarrow C,
$$

with $e_1$ the projection on the standard quotient group, $m_1$ the inclusion, and $u$ the evident isomorphism of the cokernels of $u$ to the image $S$. This three-fold factorization arises because each quotient object $B/K$ has a canonical representation (by cosets).

Exact sequences work as usual in any abelian category.

**Definition.** A composite pair of arrows,

$$
\begin{array}{ccc}
\vdots & f & \rightarrow \\
0 & a & \rightarrow \\
\oplus & b & \rightarrow \\
\rightarrow & c & \rightarrow \\
\rightarrow & 0 & \rightarrow
\end{array}
$$

(4)

is exact at $b$ when $im f \equiv \ker g$ (equivalence as subobjects of $b$) - or, equivalently, when $\text{coker } f \equiv \text{coim } g$. Observe that $im f \leq \ker g$ if and only if $gf = 0$, while $im f \geq \ker g$ if and only if every $k$ with $kg = 0$ factors as $k = mk'$, where $m$ is the first factor in the monic-epi factorization $f = me$. This bipartite definition of exactness is just the usual condition (say in $\mathsf{Ab}$): $(f, g)$ exact means that the composite $gf$ is zero and that every element killed by $g$ is in the image of $f$.

**Definition.** The diagram (with 0 the null object)

$$
\begin{array}{ccc}
0 & \longrightarrow & a \\
\longrightarrow & b & \longrightarrow \\
\longrightarrow & c & \longrightarrow \\
\longrightarrow & 0 & \longrightarrow
\end{array}
$$

(5)

is a short exact sequence when it is exact at $a$, at $b$, and at $c$.

Since $0 \to a$ is the zero arrow, exactness at $a$ means just that $f$ is monic; dually, exactness at $c$ means $g$ epic. All told, (5) short exact thus is equivalent to

$$f = \ker g, \quad g = \text{coker } f.$$  

(6)
Similarly, the statement that \( h = \ker f \) becomes the statement that the sequence
\[
\begin{array}{ccc}
a & \longrightarrow & b \\
\downarrow & & \downarrow \\
c & \longrightarrow & 0
\end{array}
\]  
(7)
is exact at \( b \) and at \( c \). Classically, such a sequence (7) was called a short right exact sequence. Similarly, \( k = \ker f \) is expressed by a short left exact sequence.

The monic-epi factorization \( f = me \) of any arrow \( f \) determines two short exact sequences which appear (with the bordering zeros omitted) as the top and side of the following commutative diagram:
\[
\begin{array}{ccc}
\ker f & \longrightarrow & \text{coim } f \\
\downarrow & & \downarrow \\
T & \longrightarrow & \text{im } f \\
\downarrow & & \downarrow \\
\text{coker } f & \longrightarrow & \\
\end{array}
\]  
(8)

A functor \( T : A \rightarrow B \) between abelian categories \( A \) and \( B \) is, by definition, exact when it preserves all finite limits and all finite colimits. In particular, an exact functor preserves kernels and cokernels, which means that
\[
\ker(Tf) = T(\ker f), \quad \text{coker}(Tf) = T(\text{coker } f); \quad (9)
\]
it also preserves images, coimages, and carries exact sequences to exact sequences. By the familiar construction of limits from products and equalizers and dual constructions, \( T : A \rightarrow B \) is exact if and only if \( f \) is additive and preserves kernels and cokernels.

A functor \( T \) is left exact when it preserves all finite limits. In other words, \( T \) is left exact if and only if \( \ker(Tf) = T(\ker f) \) for all \( f \); the last condition is equivalent to the requirement that \( T \) preserves short left exact sequences.

Abelian categories have a more economical description, not involving a given abelian group structure on each hom-set. Explicitly, let \( A \) be any category which satisfies the axioms (i), (ii), (iii), and (iv) just as above, except that (ii) is replaced by
(ii') \( A \) has binary products and binary coproducts.

Then the formula (2.6) can be used to introduce an addition in each hom-set \( A(a, b) \), and with this addition \( A \) is an abelian category. The somewhat fussy proof, Freyd [1964], Schubert [1970], will be omitted here because it seems of little use for the applications, where the categories usually come equipped with the needed addition in each \( A(a, b) \).

**Exercises**

1. For \( A, B \) abelian categories, show that an additive functor \( T : A \rightarrow B \) is exact if and only if it carries all short exact sequences in \( A \) to short exact sequences in \( B \).
2. Prove: \( A \) and \( B \) abelian implies that the product category \( A \times B \) is abelian.
3. Show that the category of all free abelian groups is not abelian.
4. Show that the category of all finite abelian groups (with arrows all morphisms of such) is abelian.
5. If \( R \) is a left noetherian ring, show that the category of all finitely generated left \( R \)-modules (with arrows all morphisms of such modules) is abelian.
6. For subobjects \( u \leq v \) of an object \( u \) in an abelian category, define a "quotient" object \( u/v \) (to agree with the usual notion in \( Ab \)). If \( gf = 0 \), prove that \( \ker g/\text{im } f \) is isomorphic to the dual object \( \text{coim } g/\ker f \).

**4. Diagram Lemmas**

In an abelian category \( A \), a chain complex is a sequence
\[
\cdots \rightarrow c_{n+1} \xrightarrow{\partial_{n+1}} c_n \xrightarrow{\partial_n} c_{n-1} \rightarrow \cdots \]  
(1)
of composable arrows, with \( \partial_n \partial_{n+1} = 0 \) for all \( n \). The sequence need not be exact at \( c_n \); the deviation from exactness is measured by the \( n \)-th homology object (for the quotient, cf. Exercise 3.6)
\[
H_n = \ker(\partial_n : c_n \rightarrow c_{n-1})/\text{im}(\partial_{n+1} : c_{n+1} \rightarrow c_n). \quad (2)
\]
Initially in algebraic topology one used chain complexes only in \( Ab \) or in \( K-\text{Mod} \) (especially for \( K \) the integers modulo a prime), but more general considerations of sheaf theory and homological algebra use complexes in many other abelian categories. The definition (2) of homology applies in any abelian category; the development of its properties depends on certain manipulations of exact sequences, normally proved in \( Ab \) by chasing elements around diagrams. We will now show how the basic diagram lemmas hold in any (fixed) abelian category \( A \).

A morphism \( \langle m, e \rangle \rightarrow \langle m', e' \rangle \) of short exact sequences (in \( A \)) is by definition a triple \( \langle f, g, h \rangle \) of arrows in \( A \) such that the diagram
\[
\begin{array}{ccc}
0 & \rightarrow & m' \rightarrow e' \rightarrow 0 \\
0 & \rightarrow & m \rightarrow e \rightarrow 0 \\
0 & \rightarrow & m' \rightarrow e' \rightarrow 0
\end{array}
\]  
commutes. The short exact sequences with these morphisms constitute a category \( \text{Ses } A \); in an evident way, it is additive. A first basic lemma is:

**Lemma 1.** (The short five lemma.) In any commutative diagram (3) with short exact rows, \( f \) and \( h \) monic imply \( g \) monic, and \( f \) and \( h \) epi imply \( g \) epi.
In Ab, take any element \( x \) in \( \ker g \); then \( g(x) = 0 \):

\[
\xymatrix{ x' \ar[r]^m & x \ar[r]^f & e(x) \\
0 \ar[r] & f(x') \ar[r] & 0 \ar[r] & 0 \ar[r] & k \}
\]

so \( k(x) = 0, e(x) = 0 \). By exactness of the first row, there must be an element \( x' \) with \( m(x') = x \). By exactness of the second row, \( f(x') = 0 \), therefore \( x' = 0 \), and so \( x = 0 \). This argument is a “diagram chase” with \( x \).

In any abelian category, the same argument can be done without elements. Take \( k = \ker g \). Then \( k = \ker e \). Since \( e \) is monic, \( e(x) = 0 \). Therefore \( k \) factors through \( m = \ker e \) as \( k = mk' \). But \( 0 = gk = gmk = m'f'k' \), and \( m' \) and \( f \) are monic, so \( k = 0 \). Since \( k = \ker g \), this proves \( g \) monic.

The proof that \( g \) is epi is dual.

In Ab, a pullback of a monic or an epi is monic or epi, respectively. This holds for pullbacks of monics in any category (Lemma V.7), and for pullbacks of epis in an abelian category, as follows.

**Proposition 2.** Given a pullback square (on the right below)

\[
\begin{array}{ccc}
A & \xrightarrow{b} & C \\
\downarrow & & \downarrow \\
B & \xrightarrow{f} & D
\end{array}
\]

in an abelian category, \( f \) epi implies \( f' \) epi. Also, the kernel \( k \) off factors as \( k = gk' \) for \( k' \) which is the kernel of \( f' \).

In particular, given a short exact sequence \( a \to b \to c \), each arrow \( g: d \to c \) to the right-hand end object \( c \) yields by pullback a short exact sequence \( a \to b \to c \). This operation (and its dual) is basic to the description of Ext \((c, a)\) (the set of “all” short exact sequences from \( a \) to \( c \)) as a bifunctor for an abelian category (Mac Lane [1963] Chap. II).

**Proof:** The pullback \( s \) (like any pullback) is constructed from products and equalizers thus: Take \( b \times d \) with projections \( p_1 \) and \( p_2 \), form the left exact sequence

\[
\begin{array}{ccc}
0 & \rightarrow & s \\
\downarrow & & \downarrow \\
0 & \rightarrow & a \times b \rightarrow b \rightarrow c
\end{array}
\]

(i.e., \( m \) is a kernel), and set \( f' = p_1, f = p_2 m \).

Here \( f p_1 - g p_2 \) is epi. For suppose \( h(f p_1 - g p_2) = 0 \) for some \( h \). Then, using the injection \( i_1 \) of the biproduct,

\[
0 = h(f p_1 - g p_2) i_1 = h f p_1 i_1 = h f,
\]

and \( h = 0 \) because \( f \) is given to be epi.

Now suppose \( u f' = 0 \) for some \( u \). Since \( f' = p_2 m \), \( u p_2 m \) is an epimorphism, so \( v p_2 m \) factors through \( f p_1 - g p_2 \) as \( u p_1 = u' f p_1 i_1 = 0 f f' \). But \( p_2 i_1 = 0 \), so

\[
0 = u p_2 i_1 = u'(f p_1 - g p_2) i_1 = u' f f' i_1 = u f.
\]

Since \( f \) is epi, \( u' = 0 \); therefore \( f' \) is epi, as desired.

Finally, consider \( k = \ker f \). The pair of arrows \( k: a \to b \) and \( 0: a \to d \) have \( f k = 0 = g 0 \), so by the definition of the pullback \( s \) there is a unique arrow \( k': a \to s \) with \( g k' = k \) and \( f k' = 0 \), since \( k \) is monic, so is \( k' \). To show the kernel \( \ker f' \) consider any arrow \( v \) with \( f v = 0 \). Then \( g v \) factors through \( k = \ker f \) as \( g v = k v' \) for some \( v' \). Then \( g v = g f v' \) and \( f v = 0 = f' k' v' \), so by the uniqueness involved in the definition of a pullback, \( v = k' v' \). Therefore \( k' = \ker f' \), as desired.

In virtue of this Proposition, diagram chases can be made in any abelian category using “members” (in \( A \)) instead of elements (in \( Ab \)). Call an arrow \( x \) with codomain \( a \) a member of \( a \), written \( x \in a \), and define \( x \equiv y \) for two members of \( a \) to mean that there are epis \( u, v \) with \( xu = yv \). This relation is manifestly reflexive and symmetric. To prove it transitive, suppose also that \( yw \equiv zw \) for epis \( w \) and \( r \) and form the pullback square displayed at the upper left in the diagram

By Proposition 2, \( u' \) and \( w' \) are epis, and hence \( x \equiv z \). Then a member of \( a \) is an equivalence class, for the relation \( \equiv \), of arrows to \( a \). Since every arrow \( x \) has a factorization \( x = me \), every member of \( a \) is represented by a subobject (a monic \( m \)) of \( a \), but we shall not need to use this fact. Each object \( a \) has zero member, the (equivalence class of the) zero arrow \( 0 \to a \). Each member \( x \in a \) has a “negative” \(-x \).

For any arrow \( f: a \to b \), each member \( x \in a \) gives \( f \in \epsilon_0 b \), and \( x \equiv y \) in \( a \) implies \( f x \equiv f y \) in \( b \), so any arrow from \( a \) to \( b \) carries members of \( a \) to members of \( b \) — just as if they were elements of sets.

**Theorem 3.** (Elementary rules for chasing diagrams.) For the members in any abelian category

(i) \( f : a \to b \) is monic if and only if, for all \( x, y \in a \), \( f x \equiv 0 \) implies \( x \equiv 0 \);
(ii) \( f : a \to b \) is monic if and only if, for all \( x, x' \in a \), \( f x \equiv f x' \) implies \( x \equiv x' \);
(iii) \( g : b \to c \) is epi if and only if each \( z \in c \) there exist \( a \in b \) with \( g y \equiv z \).
By exactness at $a_3$ and Rule (v) of the theorem, there is a $y \in e_{m_2}$ with $g_2 y \equiv x$. Then $0 \equiv f_3 x \equiv f_3 g_2 y \equiv f_2 y$, so by exactness at $b_2$ there is a $y' e_{m_1}$ with $h_1 y' \equiv f_2 y$. Since $f_1$ is epi, there is a $z e_{m_0}$ with $h_1 f_1 z \equiv f_2 y$ or $f_2 g_1 z \equiv f_2 y$. But $f_2$ is monic, so, by Rule (ii), $g_1 z \equiv y$ and $x \equiv g_2 y \equiv g_2 g_1 z \equiv 0$. Since any $x$ with $f_4 x \equiv 0$ is itself $0$, $f_4$ is monic, as required.

As another illustration, consider any morphism $\langle f, g, h \rangle$ of short exact sequences, as in (3); add the kernels and cokernels of $f, g$, and $h$ to form a diagram

$\begin{array}{ccc}
0 & \rightarrow & Ke f \rightarrow Ke g \rightarrow Ke h \\
| & \downarrow f & \downarrow j \\
0 & \rightarrow & a \rightarrow b \rightarrow c \rightarrow 0
\end{array}$

$\begin{array}{ccc}
0 & \rightarrow & a \rightarrow b \rightarrow c \rightarrow 0 \\
| & \downarrow f & \downarrow j \\
0 & \rightarrow & Ke f \rightarrow Ke g \rightarrow Ke h
\end{array}$

where $Ke f$ is the domain of $\ker f$, $Co f$ the codomain of $\coker f$, etc.

In this diagram the columns (with 0's added top and bottom) are exact sequences by construction, and both middle rows are given to be exact. By the definitions of kernel and cokernel, one may add unique arrows $m_0, e_0$ in the top row and $m_1, e_1$ in the bottom row so as to make the added squares commute. An easy diagram chase (by the method of Theorem 3) shows the first row exact at $Ke f$ and $Ke g$; dually, the last row is exact at $Co g$ and $Co h$. However, the first row is not necessarily a short exact sequence because $e_0$ need not be epi; moreover, this happens precisely when $m_1$ is not monic. An easy example of this phenomenon (in Ab; $g \equiv 0$) is

$\begin{array}{ccc}
0 & \rightarrow & 0 \rightarrow Z \rightarrow 0 \\
| & \downarrow g & \downarrow i \\
0 & \rightarrow & Z \rightarrow 0
\end{array}$

The failure of exactness can be repaired by the following striking lemma, which produces an added $\delta$ called the connecting homomorphism—it is essentially the connecting homomorphism used for relative homology (a complex modulo a subcomplex) and for the connecting maps between derived functors in homological algebra.

** Lemma 5 (Ker-coker sequence = Snake lemma.)** Given a morphism $\langle f, g, h \rangle$ of short exact sequences, as in (3), there is an arrow $\delta : Ke h \rightarrow Co f$
such that the following sequence is exact
\[ 0 \rightarrow K e f \rightarrow K e g \rightarrow K e h \rightarrow C o f \rightarrow C o g \rightarrow C o h \rightarrow 0 \] (6)

**Proof.** From the map of short exact sequences we first build a different diagram: on the left in
\[
\begin{array}{cccc}
\alpha & \beta & \gamma & c_0 \\
\| & & & \\
\alpha' & \beta' & \gamma' & c_0 \\
\downarrow & & & \\
\alpha'' & \beta'' & \gamma'' & c_0 \\
\end{array}
\]

\[ c_0 = K e h, \text{ } d \text{ is the pullback of } e \text{ and } k = \ker h, \text{ so that } u \text{ is epi with kernel } \] s as in Proposition 2; dually, \( d' \) is the pushout of \( p = \coker f \) and \( m' \), with cokernel \( s' \) as shown. Right down the middle runs an arrow \( \delta_0 = p' g' k' : d' \rightarrow d' \), with \( s' \delta_0 = h k u = 0 \) and \( \delta_0 s = u' p f = 0 \). Since \( u' = \ker s' \) and \( u = \coker s \), this means that \( \delta_0 \) factors uniquely as
\[ \delta_0 = u' \delta u : d' \rightarrow c_0 \rightarrow a_1 \rightarrow d'. \]

The middle factor is the connecting "arrow \( \delta : c_0 \rightarrow a_1 \)."

The effect of this arrow \( \delta \) on a member \( x \in c_0 \) can be described by the zig-zag staircase shown at the right of (7) above. Indeed, since \( e \) is epi there is a member \( y \in b \) with \( e y = k x \). Then \( e' \gamma y = \gamma e y = h k x \equiv 0 \), so by exactness there is a member \( z \in c_0 \) in the kernel \( m' \equiv k \), \( k x \equiv g y \). We claim that \( \delta x \equiv z \) is then the member \( z_1 = p z \in a_1 \). For, \( d \) is a pullback so there is an \( x_0 \in d \)
with \( u x_0 \equiv x \), \( k' x_0 \equiv y \). Then
\[ u' \delta(2) \equiv u' \delta x_0 = \delta_0 x_0 \equiv p' g' y \equiv u' z_1 \]
and \( u' \) is monic, \( \delta x \equiv z_1 \). This argument also proves that (the equivalence class of) the member \( z_1 \) is independent of the choices made in the construction of the zig-zag (7). This zig-zag is exactly the description usually given for the action of a connecting morphism \( \delta \) on the elements of abelian groups.

Using the zig-zag description we can now prove the exactness of the ker-coker sequence (6), say the exactness at \( K e h \). First, to show that \( \delta e_0 = 0 \), it suffices to show \( \delta e_0 w = 0 \) for any member \( w \in b_0 = K e g \). But the member \( e_0 w = x \in c_0 \) has \( k x = k e_0 w = e j w \), where \( j = \ker g \) as in (5); hence in the zig-zag (7) we may choose \( y = j w \). Then \( g y = g j w \equiv 0 \), which proves that \( \delta_0 = 0 \). On the other hand, consider any \( x \in c_0 \) with \( \delta x \equiv 0 \). This means that the \( z_1 \) constructed in the zig-zag has \( z_1 \equiv 0 \); by exactness there is a member \( z_2 \in c_0 \) with \( f z_2 \equiv z \) which means that \( g m z_2 \equiv g y \). Now form the "difference" member \( y_0 = y - m z_2 \in b \). By Rule (vi) above, this difference member has \( e y_0 \equiv e y = k x \) and \( g y_0 \equiv 0 \). But \( j : b_0 \rightarrow b \) is \( \ker g \), so there is an \( x_0 \in b_0 \) with \( j x_0 \equiv y_0 \):

\[
\begin{array}{cccc}
\alpha & \beta & \gamma & c_0 \\
\| & & & \\
\alpha' & \beta' & \gamma' & c_0 \\
\downarrow & & & \\
\alpha'' & \beta'' & \gamma'' & c_0 \\
\end{array}
\]

Then \( k e_0 x_0 = e y_0 \equiv k x \) and \( k \) is monic, \( e_0 x_0 \equiv x \). We have shown that each \( x \) with \( \delta x \equiv 0 \) has the form \( x \equiv e_0 x_0 \), so is in the image of \( e_0 \). This proves exactness; in fact, it is exactly like the usual exactness proof with honest elements.

**Exercises**

1. In the five-lemma, obtain minimal hypotheses (on \( f_1, f_2 \), and \( f_4 \) only) for \( f_3 \) to be monic.
2. In the five-lemma, prove \( f_5 \) epi using members (not co-members) [Hint: Rule (vi) of Theorem 3 is necessary in this proof].
3. Complete the proof of the exactness of the ker-coker sequence.
4. Show that the connecting morphism \( \delta \) is natural; i.e., that it is a natural transformation between two appropriate functors defined on a suitable category whose objects are morphisms (3) of short exact sequences.
5. A 3 \times 3 diagram is one of the form (bordered by zeros)

\[
\begin{array}{ccc}
\alpha & \beta & \gamma \\
\| & & \\
\alpha' & \beta' & \gamma' \\
\downarrow & & \\
\alpha'' & \beta'' & \gamma'' \\
\end{array}
\]

(a) Give a direct proof of the 3 \times 3 lemma: If a 3 \times 3 diagram is commutative and all three columns and the last two rows are short exact sequences, then so is the first row.
(b) Show that this lemma also follows from the ker-coker sequence.
(c) Prove the middle 3 \times 3 lemma: If a 3 \times 3 diagram is commutative, and all three columns and the first and third rows are short exact sequences, then so is the middle row.
6. For two arrows \( f : a \rightarrow b \) and \( g : b \rightarrow c \) establish an exact sequence
\[ 0 \rightarrow K e f \rightarrow K e g \rightarrow K e h \rightarrow C o f \rightarrow C o g \rightarrow 0. \]
7. Show explicitly that the category \( \text{Seq}(A) \) is not in general abelian.
Diagram Lemmas

Notes

Shortly after the discovery of categories, Eilenberg and Steenrod [1952] showed how the language of categories and functors could be used to give an axiomatic description of the homology and cohomology of a topological space. This, in turn, suggested the problem of describing those categories in which the values of such a homology theory could lie. After discussions with Eilenberg, this was done by Mac Lane [1948, 1950]. His notion of an “abelian bicategory” was clumsy, and the subject languished until Buchsbaum’s axiomatic study [1955] and the discovery by Grothendieck [1957] that categories of sheaves (of abelian groups) over a topological space were abelian categories but not categories of modules, and that homological algebra in these categories was needed for a complete treatment of sheaf cohomology (Godement [1958]). With this impetus, abelian categories joined the establishment.

This chapter has given only an elementary theory of abelian categories—a demonstration directly from the axioms of all the usual diagram lemmas. Our method of “chasing members” is an adaptation of the method given by Mac Lane [1963, Chap. XII]: the critical point is the snake lemma, which must construct an arrow. Earlier proofs of this lemma in abelian categories were obscure; the present version is due to M. André (private communication). These diagram lemmas can also be proved in abelian categories from the case of $R$-modules by using suitable embedding theorems (Lubkin-Heron-Freyd-Mitchell). These beautiful theorems construct for any small abelian category $A$ a faithful, exact functor $A \to \text{Ab}$ and a full and faithful exact functor $A \to \text{R-Mod}$ for a suitable ring $R$. For proofs we refer to Mitchell [1965], Freyd [1964], and Pareigis [1970].

These sources will also indicate the further elegant developments for abelian categories: A Krull-Remak-Schmidt theorem, Morita duality, the construction of “injective envelopes” in suitable abelian categories, the structure of Grothendieck categories, and the locally Noetherian categories (Gabriel [1962]).

IX. Special Limits

This chapter covers two useful types of limits (and colimits): The filtered limits, which are limits taken over preordered sets which are directed (and, more generally, over certain filtered categories), and the “ends”, which are limits obtained from certain bifunctors, and which behave like integrals.

1. Filtered Limits

A preorder $P$ is said to be directed when any two elements $p, q \in P$ have an upper bound in $P$; that is, an $r$ with $p \leq r$ and $q \leq r$ (there is no requirement that $r$ be unique). It follows that any finite set of elements of $P$ has an upper bound in $P$. A directed preorder is also called a “directed set” or a “filtered set”.

This notion (renamed) generalizes to categories. A category $J$ is filtered when $J$ is not empty and

(a) To any two objects $j, j' \in J$ there is $k \in J$ and arrows $j \to k, j' \to k$:

(b) To any two parallel arrows $u, v : i \to j$ in $J$, there is $k \in J$ and an arrow $w : j \to k$ such that $wu = wv$, as in the commutative diamond

Condition (a) states that the discrete diagram $(j, j')$ is the base of a cone with vertex $k$. Condition (b) states that $i \to j$ is the base of a cone. It follows that any finite diagram in a filtered category $J$ is the base of at least one cone with a vertex $k \in J$. 207
A filtered colimit is by definition a limit of a functor \( F : J \to C \) defined on a filtered category \( J \).

Classically, colimits were defined only over directed preorders (sometimes just over directed orders). This has proved to be a needless conceptual restriction of the notion of colimit. What does remain relevant is the interchange formulas for filtered colimits (§2) and the possibility of obtaining all colimits from finite coproducts, coequalizers, and colimits over directed preorders. Since we already know that (infinite) coproducts and coequalizers give all colimits (Theorem V.2.1) this needs only the following result.

**Theorem 1.** A category \( C \) with finite coproducts and colimits over all (small) directed preorders has all (small) coproducts.

**Proof.** We wish to construct a colimit for a functor \( F : J \to C \), where \( J \) is a set (= a discrete category). Let \( J^+ \) be the preorder with objects all finite subsets \( S \subset J \), ordered by inclusion; clearly, \( J^+ \) is filtered. Let \( F^+ \)

assign to each finite subset \( S \) the coproduct \( \amalg F_S \), taken over all \( s \in S \). If \( S \subset T \) is an arrow \( u : S \to T \) of \( J^+ \), take \( F^+ u \) to be the unique (dotted) arrow which makes the diagram

\[
\begin{array}{ccc}
F^+ S &=& F^+ T \\
\downarrow u & & \downarrow u \\
F_S & & F_T \\
\end{array}
\]

commute for every \( s \in S \), with \( i \) and \( i' \) the injections of the coproducts. This evidently makes \( F^+ \) a functor \( J^+ \to C \) which agrees on \( J \) with the given functor \( F \), if \( J \) is included in \( J^+ \) by identifying each \( j \) with the one-point subset \( \{j\} \).

Now consider any natural transformation \( \theta : F^+ \to G \) to some other functor \( G : J^+ \to C \). For each \( s \in S \) the diagram

\[
\begin{array}{ccc}
F^+ S &=& G S \\
\downarrow \theta_S & & \downarrow \theta_G \\
F_S & & G \{s\} \\
\end{array}
\]

commutes. By the definition of coproducts, this means that \( \theta \) is completely determined by the values \( \theta_S \) of \( \theta \) on \( F_S \). In particular, each cone \( v^+ : F^+ \to C \) over \( F^+ \) is completely determined by its values on \( J \), which form a cone \( v : F \to C \) over \( F \). Moreover, \( v^+ \) is a limiting cone if and only if \( v \) is. Thus we can calculate the desired coproduct \( \amalg F_j \), which is the colimit of \( F \), as the colimit of \( F^+ \), known to exist because \( J^+ \) is a directed preorder.

As a typical application, we construct colimits in \( \text{Grp} \).

**Proposition 2.** The forgetful functor \( \text{Grp} \to \text{Set} \) creates filtered colimits.

---

**Filtered Limits**

**Proof.** We are given a filtered category \( J \) and a functor \( G : J \to \text{Grp} \); it assigns to arrows \( j \to k \) group morphisms \( G_j \to G_k \); we shall write \( G \_j \) both for the group and its underlying set. We are also given a limiting cone \( \mu \) for the composite functor \( J \to \text{Grp} \to \text{Set} \); it has a set \( S \) as vertex and assigns to each \( j \in J \) a function \( \mu_j : G_j \to S \). We first show that there is a unique group structure on the set \( S \) which will make all functions \( \mu_j \) morphisms of groups. First note that to each \( s \in S \) there is at least one index \( j \) with a group element \( g_j \) for which \( \mu_j g_j = s \). Otherwise we could omit \( s \) from \( S \) to have a cone with a smaller set \( S' \) as vertex, an evident contradiction to the universality of \( S \) (there would be two functions \( S \to S' \) having the same composite with \( \mu \)).

Now we define a product of any two elements \( s, t \in S \). Write \( s = \mu_j g_j \) and \( t = \mu_k g_k \) for some \( j, k \in J \); since \( J \) is filtered, there is in \( J \) a cone \( \eta \)

over \( j, k \) with some vertex \( i \). The image of this cone under \( G \) is \( G_i \to G_j \to G_k \), so \( s \) and \( t \in S \) both come from elements of the group \( G_i \); define their product in \( S \) to be \( \mu_i r \), of their product in \( G_i \). This product is independent of the choice of \( i \), because a different choice \( i' \) is part of a cone \( G_j \to G_i \to G_{i'} \) of group morphisms. Also, the product of three factors \( r, s, t \) is associative, because we can choose \( G_j \) to contain pre-images of all three, and multiplication is known to be associative in \( G_i \). Each group \( G_i \) has a unit element, and each \( G_j \to G_k \) maps unit to unit; the common image of these units is a unit for the multiplication in \( S \). Inverses are formed similarly.

We now have found a (unique) group structure on \( S \) for which \( \mu_j : G_j \to S \) is a morphism of groups. This states that \( \mu \) is a cone from \( G \) to \( S \) in \( \text{Grp} \). It is universal there: If \( v : G \to T \) is another cone in \( \text{Grp} \), it is also a cone in \( \text{Set} \), so there is a unique set map \( f : S \to T \) with \( f \mu = v \); one checks as above that this map \( f \) must be a morphism of groups.

This argument is clearly not restricted to \( \text{Grp} \); it applies to each category \( \text{Alg} \), of algebras of a fixed type \( \tau \) (defined by operators and identities). The same remark applies to the following corollary.

**Corollary 3.** \( \text{Grp} \) has all (small) colimits.

**Proof.** First, the one-element group is an initial object in \( \text{Grp} \). Next, any two groups \( G \) and \( H \) have a coproduct \( G \amalg H \). Indeed, any pair of
homomorphisms \( G \rightarrow L, H \rightarrow L \) to a third group \( L \) factors through the subgroup of \( L \) generated by the images of \( G \) and \( H \). The cardinal number of this subgroup is bounded; this verifies the solution set condition for an application of the adjoint functor theorem to construct the coproduct \( G \star H \).

These two observations show that \( \text{Grp} \) has all finite coproducts. By Proposition 2, it has all filtered colimits. Hence, by Theorem 1, it has all small coproducts. To get all small colimits we then need only coequalizers, and the coequalizer of two homomorphisms \( u, v: G \rightarrow H \) is the projection \( H \rightarrow H/N \) on the quotient group by the least normal subgroup \( N \) containing all the elements \((ug)\,(vg)^{-1}\) for \( g \in G \).

This proof gives an explicit picture of the coproduct in \( \text{Grp} \). The coproduct \( G \star H \) of two groups is usually called the free product; its elements are finite words \( \langle g_1, h_1, g_2, h_2, \ldots, g_n, h_n \rangle \) spelled in letters \( g_i \in G \) and \( h_i \in H \); these words are multiplied by juxtaposition, while equality is given by successive cancellations (if \( h_i = 1 \) in \( H \), drop it and multiply \( g_i g_{i+1} \) in \( G \), etc.). A direct proof of associativity of the multiplication from this definition is fuzzy. By this corollary an infinite coproduct \( \coprod_i G_i \) of groups \( G_i \) is obtained by pasting together all the finite coproducts

\[
G_1 \star G_2 \star \cdots \star G_k
\]

(the inclusion maps make this a subgroup of any coproduct of more factors). Thus \( \coprod_i G_i \) is the union of all these finite coproducts.

**Exercises**

1. Use the adjoint functor theorem to prove in one step that \( \text{Grp} \) has all small colimits.
2. Prove that \( \text{Alg} \) has all small colimits; in particular, describe the initial object (when is it empty?)?

2. **Interchange of Limits**

Consider a bifunctor \( F: P \times J \rightarrow X \) to a cocomplete category \( X \). For values \( p \in P \) of the parameter \( p \), the colimits of \( F(p,-): J \rightarrow X \) define functors \( p \mapsto \text{Colim}_J F(p,j) \) of \( p \), so that the colimiting cones.

\[
\mu_{p,j}: F(p,j) \rightarrow \text{Colim}_J F(p,j)
\]

are natural in \( p \) (Theorem V.3.1). One may prove readily (§8 below) that

\[
\text{Colim}_p \text{Colim}_J F(p,j) \cong \text{Colim}_J \text{Colim}_p F(p,j)
\]

with the isomorphism given by the "canonical" map. Dually, limits commute. But limits need not commute with colimits, because the canonical map

\[
\kappa: \text{Colim}_J \text{Lim}_p F(p,j) \rightarrow \text{Lim}_p \text{Colim}_J F(p,j)
\]

need not be an isomorphism.

This canonical map exists as soon as all four limits and colimits in (3) exist, and is constructed as in the following diagram

\[
\begin{array}{c}
\text{Colim}_J F(p,j) \xleftarrow{\mu_{p,j}} \text{Lim}_p F(p,j) \xrightarrow{\mu_{j}} \text{Colim}_p F(p,j) \\
\kappa \\
\end{array}
\]

where \( v \) and \( v_{-j} \) for each \( j \) are limiting cones, and \( \mu, \mu_{p,j} \) for each \( p \) the colimiting cones. Since \( v \) is a cone in \( p \) and \( \mu \) is natural in \( p \), the composite \( \mu_{p,j} v_{-j} \) for fixed \( j \) is a cone in \( p \); by the universality of \( v \) there then exist arrows \( \alpha \) for each \( j \) making the left hand squares commute. Since \( \mu_{p,j} \) is a cone in \( j \), so is \( \alpha \); by the universality of \( \mu \), there is then a map \( \kappa \) making the right hand square commute. It is the desired canonical arrow.

This \( \kappa \) need not be an isomorphism. Consider, for example, the case when \( P = \{1, 2\} \) and \( J = \{1, 2\} \) are both discrete 2-object categories. The canonical \( \kappa \) (in evident notation)

\[
\kappa: (A_1 \times B_1) \coprod (A_2 \times B_2) \rightarrow (A_1 \coprod A_2) \times (B_1 \coprod B_2)
\]

is given by two components \( \alpha_1 \) and \( \alpha_2 \), where \( \alpha_1 \) is determined by

\[
A_1 \coprod A_2 \xleftarrow{\kappa} (A_1 \coprod A_2) \times (B_1 \coprod B_2) \rightarrow B_1 \coprod B_2
\]

In \( \text{Ab} \), \( \kappa \) is evidently an isomorphism, but in \( \text{Set} \) it is not—the domain of \( \kappa \) is a disjoint union of two sets, while the codomain of \( \kappa \) is the four-fold disjoint union

\[
(A_1 \times B_1) \coprod (A_1 \times B_2) \coprod (A_2 \times B_1) \coprod (A_2 \times B_2)
\]

We now turn to conditions which suffice to make \( \kappa \) an isomorphism.

**Theorem 1.** If the category \( P \) is finite while \( J \) is small and filtered, then for any bifunctor \( F: P \times J \rightarrow \text{Set} \) the canonical arrow

\[
\kappa: \text{Colim}_J \text{Lim}_p F(p,j) \rightarrow \text{Lim}_p \text{Colim}_J F(p,j)
\]

is an isomorphism.
This states that finite limits commute with filtered colimits in Set.

**Proof.** By the construction of colimits in terms of coproducts and coequalizers (dual of Theorem V.2.2),

\[ \text{Colim}_j F(p, j) = \sqcup_j F(p, j)/E, \]

where \( \sqcup_j \) is the disjoint union and \( E \) is the equivalence relation defined for elements \( x \in F(p, j) \) and \( x' \in F(p, j) \) in that union by \( x Ex' \) if and only if there are arrows \( u : j \to k, u' : j \to k \) with \( F(p, u)x = F(p, u')x' \). Write \((x, j)\) for the \( E\)-equivalence class of an element \( x \in F(p, j) \). Now \( J \) is filtered; condition (a) in the definition of “filtered” implies that any finite list \((x_1, j_1), ..., (x_m, j_m)\) of such elements can be written as a list \((y_1, k_1), ..., (y_n, k_n)\) with one second index \( k \). Condition (b) in the definition implies that every equality between elements of this list takes place after application of a suitable one arrow \( w : k \to k' \).

For any functor \( G : P \to \text{Set} \), \( \lim_p Gp = \text{Cone}(* , G) \), the set of cones \( \tau \) over \( G \) with vertex a point *. If \( GP = \text{Colim}_j F(p, j) \) and \( P \) is finite, each such cone consists of a finite number of elements of \( \text{Colim}_j F(p, j) \) and the conditions that \( \tau \) be a cone involve a finite number of equations between these elements. Since \( J \) is directed, the observations above now mean that each cone \( \tau \) can consist of elements \( \tau_x = (y_{ij}, k) \) for some one index \( k \), where \( y_{ij} \in F(p, k) \) already constitute a cone \( y_x : * \to F(-, k) \). This cone \( y \) is an element of \( \lim_p F(p, k) \); its equivalence class \( y, k \) is an element of \( \text{Colim}_j \lim_p \). The map \( \tau \mapsto (y, k) \in \text{Colim}_j \lim_p F(p, j) \).

which is independent of the choices made, is the desired (two-sided) inverse of the canonical arrow \( k \).

**Exercises**

1. Show that \( x \) of (3) is natural for arrows \( e : F \to F' \) in \( X^{p \times j} \).
2. (Verdier). A category \( J \) is pseudo-filtered when it satisfies condition \( (h) \) for filtered categories and the following condition \( (a') \): Any two arrows \( i \to j, i \to j' \) with the same domain can be embedded in a commutative diamond

\[
\begin{array}{ccc}
\ast & \to & j \\
\downarrow & & \downarrow \\
& & i
\end{array}
\]

Prove that a category \( J \) is filtered if and only if it is connected and pseudo-filtered. Prove that a category is pseudo-filtered if and only if its connected components are filtered.

3. In Set, show that coproducts commute with pullback.
4. Using Exercises 2 and 3, show that pseudo-filtered colimits commute with pullbacks in Set.

### Final Functors

**3. Final Functors**

Colimits may often be computed over subcategories. For example, the colimit of a functor \( F : N \to \text{Cat} \), where \( N \) is the linearly ordered set of natural numbers, is clearly the same as the colimit of the restriction of \( F \) to any infinite subset \( S \) of \( N \) (i.e., to any subcategory which contains at least one object "beyond" each object of \( N \)). In classical terminology, such a subset \( S \) was called "cofinal" in \( N \); it now seems preferable to drop the "co", as not related to dualizations. Also, we will replace the subset \( S \) first by the inclusion functor \( S \to N \) and then by an arbitrary functor.

A functor \( L : J' \to J \) is called **final** if for each \( k \in J \) the comma category \( (k \downarrow L) \) is non-empty and connected. This means that to each \( k \) there is an object \( j' \in J' \) and an arrow \( k \to Lj' \), and that any two such arrows can be joined to give finite commutative diagram of the form

\[
\begin{array}{ccc}
Lj' & \to & \ast \\
\downarrow & \simeq & \downarrow \\
& \simeq & \downarrow \\
Lj & \to & \ast
\end{array}
\]

A subcategory is called **final** when the corresponding inclusion functor is final. For example, if \( J \) is a linear order, \( J' \subset J \) and \( L \) the inclusion, then \( L \) final means simply that to each \( k \in J \) there is \( j' \in J' \) such that \( k \leq j' \).

For \( L : J' \to J \) and \( F : J \to X \) there is a canonical map

\[
\begin{array}{ccc}
\text{Colim}_j F \to \text{Colim} F
\end{array}
\]

defined when both colimits exist; if \( \mu' : FL \to \text{Colim} F \) and \( \mu' \) are the colimiting cones, \( \mu \) is the unique arrow of the arrow \( X \) with \( \mu_j = \mu_j' \) for all \( j \in J' \).

The main theorem now is:

**Theorem 1.** If \( L : J' \to J \) is final and \( F : J \to X \) is a functor such that \( X = \text{Colim} F \) exists, then \( \text{Colim} F \) exists and the canonical map (1) is an isomorphism.

**Proof.** Given a colimiting cone \( \mu : FL \to \text{Colim} F = X \), we construct arrows \( i_k : F_k \to X \) for each \( k \in J \) by choosing an arrow \( u : k \to Lj' \) and taking \( i_k \) to be the composite

\[
\begin{array}{ccc}
F_k & \overset{\mu_k}{\to} & FLj' \\
\downarrow & \simeq & \downarrow \\
& \simeq & \downarrow \\
& \mu_j & \to \ar X
\end{array}
\]

Since \( \mu \) is a cone and \( (k \downarrow L) \) is connected, the connectivity diagram above readily shows \( i_k \) independent of the choice of \( u \) and \( j' \). It follows at once that \( r : F \to X \) is a cone with vertex \( x \) and base \( F \). On the other hand, if \( \lambda : F \to y \) is another cone with this base \( F \), then \( \lambda X : FL \to y \) is a cone with base \( FL \), so by the universal property of \( \mu \) there exists a unique \( f : x \to y \).
Special Limits

with \( f \mu = \lambda L \), and hence (because \( \lambda_k = \lambda_{k'} \cdot F u \)) with \( f t = \lambda \). This shows that \( \tau \) is a limiting cone and hence that \( x = \text{Colim} F \); clearly this also makes the canonical map \( h \) an isomorphism.

The condition that \( L \) be final is necessary for the validity of this theorem (cf. Exercise 5). The dual of this result (the dual of final is "initial") is useful for limits.

**Exercises**

1. If \( j \in J \) and \( \{ j \} \) is the discrete subcategory of \( J \) with just the one object \( j \), show that the inclusion \( \{ j \} \to J \) is final if and only if \( j \) is a terminal object in \( J \). What does this say about colimits and terminal objects?
2. Prove that a composite of final functors is final.
3. If \( J \) is filtered, \( L : J \to J \) is full, and each \( (k, L) \) is non-empty \( (k \in J) \) prove that \( L \) is final.
4. For the covariant hom-functor \( J(k, -) : J \to \text{Set} \), use the Yoneda Lemma to show that \( \text{Colim} \ J(k, -) \) is the one-point set.
5. (Converse of Theorem of the text). Let \( L : J \to J \) be a functor, where \( J' \) and \( J \) have small hom-sets, such that for every \( F : J \to X \) with \( X \) complete the canonical map \( \text{Colim} F L \to \text{Colim} F \) is an isomorphism. Prove that \( L \) must be final (Hint: Use \( F = J(k, -) \), \( X = \text{Set} \), and Exercise 4).

4. Diagonal Naturality

We next consider an extension of the concept of naturality. Given categories \( C, B \) and functors \( S, T : C^{op} \times C \to B \), a **dinatural transformation** \( \alpha : S \Rightarrow T \) is a function \( \alpha \) which assigns to each object \( c \in C \) an arrow \( \alpha_c : S(c, c) \to T(c, c) \) of \( B \), called the **component** of \( \alpha \) at \( c \), in such a way that for every arrow \( f : c \to c' \) of \( C \) the following hexagonal diagram

\[
\begin{array}{ccc}
S(c, c) & \xrightarrow{f \cdot 1} & T(c, c) \\
\downarrow S(f, 1) & & \downarrow T(1, f) \\
S(c', c) & \xrightarrow{1 \cdot f} & T(c', c') \\
\downarrow S(f, 1) & & \downarrow T(1, f) \\
S(c', c') & \xrightarrow{f \cdot 1} & T(c', c') \\
\end{array}
\]

is commutative. Observe that the contravariance of \( S \) and \( T \) in the first argument is used in forming the arrows \( S(f, 1) \) and \( T(f, 1) \) in this diagram.

Every ordinary natural transformation \( \tau : S \Rightarrow T \) between the bifunctors \( S \) and \( T \), with components \( \tau_{c,c'} : S(c, c') \to T(c, c') \), will yield a dinatural transformation \( \alpha : S \Rightarrow T \) between the same bifunctors, with components just the diagonal components of \( \tau \); thus \( \alpha_c = \tau_{c,c} \). More interesting examples arise from functors which are "dummy" in one or more variables. For example, \( T : C^{op} \times C \to B \) is dummy in its first variable if it is a composite

\[
C^{op} \times C \xrightarrow{Q} C \xrightarrow{T_0} B,
\]

where \( Q \) is the projection on the second factor and \( T_0 \) is some functor (of one variable). Put differently, each functor \( T_0 : C \to B \) of one variable may be treated as a bifunctor \( C^{op} \times C \to B \), dummy in the first variable. Again, a functor dummy in both variables is in effect a constant object \( b \in B \) with \( T(c, c') = b \) for all objects \( c, c' \in C \) and \( T(f, f') = 1_b \) for all arrows \( f \) and \( f' \) in \( C \).

The following types of dinatural transformations \( S \Rightarrow T \) arise. If \( S \) is dummy in its second variable and \( T \) dummy in its first variable, a dinatural transformation \( \alpha : S \Rightarrow T \) sends a functor \( S_0 : C^{op} \to B \) to a covariant \( T_0 : C \to B \) by components \( \alpha_c : S_0 c \to T_0 c \) which make the diagrams

\[
\begin{array}{ccc}
S_0 c & \xrightarrow{s_0 f} & T_0 c \\
\downarrow s_0 f & & \downarrow T_0 f \\
S_0 c' & \xrightarrow{s_0 f} & T_0 c'
\end{array}
\]

(2)

commute for each arrow \( f : c \to c' \) of \( C \). Such a dinatural transformation might be called a natural transformation of the contravariant functor \( S_0 \) to the covariant functor \( T_0 \). (Dually, of a covariant to a contravariant one.)

If \( T = b : C^{op} \times C \to B \) is dummy in both variables, a dinatural transformation \( \alpha : S \Rightarrow b \) consists of components \( \alpha_c : S(c, c) \to b \) which make the diagram

\[
\begin{array}{ccc}
S(c', c) & \xrightarrow{s(f, 1)} & S(c', c') \\
\downarrow s(f, 1) & & \downarrow s(f, 1) \\
S(c, c) & \xrightarrow{\alpha_c} & b
\end{array}
\]

(3)

commute for every \( f : c \to c' \). (The right hand side of the hexagon (1) has collapsed to one object \( b \).) Such a transformation \( \alpha : S \Rightarrow b \) is called an **extranatural** transformation, a "supernatural" transformation or a **wedge** from \( S \) to \( b \). The same terms are applied to the dual concept \( \beta : b \Rightarrow T \), given by components \( \beta_b : b \Rightarrow T(c, c) \) such that every square

\[
\begin{array}{ccc}
& \xrightarrow{b} & T(c, c) \\
\beta_b & \downarrow & \downarrow T(1, f) \\
& & T(c', c') \xrightarrow{T(1, f)} T(c, c')
\end{array}
\]

(4)

is commutative. (The left hand side of the hexagon (1) has collapsed.)
We give an example of each type. A Euclidean vector space $E$ is a
vector space over the field $\mathbb{R}$ of real numbers equipped with an inner
product function $\langle \cdot , \cdot \rangle : E \times E \to \mathbb{R}$ which is bilinear, symmetric, and
positive definite. These spaces are the objects of a category $\text{Euclid}$, with
arrows those linear maps which preserve inner product. There are
two functors
\[
U : \text{Euclid} \to \text{Vct}_\mathbb{R}, \quad (\text{Euclid})^{op} \to \text{Vct}_\mathbb{R}
\]
to the category of real vector spaces: The (covariant) forgetful functor $U$
"forget the inner product" and the contravariant functor "take the dual
space". Now for each Euclidean vector space $E$ the assignment $e \mapsto \langle e, - \rangle$, for $e \in E$, is a linear function $\kappa_e : E \to E^*; 
$ these functions $\kappa_e$ are the
components for a transformation $\kappa$ which is dinatural from $U$ to $*$
(dual of type (2)): This is the fact, familiar in Riemannian geometry, that
each Euclidean vector space is \textit{naturally} isomorphic to its dual – and
we need the notion of dinaturality to express this fact categorically.

\textit{Evaluation.} $E_x$, for $X$ a (small) set, takes the value of each function
$h : X \to A$ at each argument $x \in X$. If the (small) set $A$ is fixed, we may
regard $E_x$ as a function
\[
E_x : \text{hom}(X, A) \times X \to A. \quad \langle h, x \rangle \mapsto h x,
\]
defined for each object $X \in \text{Set}$. For two small sets $X$ and $Y$, $\text{hom}(X, A) \times Y$
the (object function of a) functor $\text{Set}^{op} \times \text{Set} \to \text{Set}$, while for every
arrow $f : Y \to X$ the obvious property $h(f(x)) = (hf)x$ of evaluation states
that the square like (3) always commutes. Hence the functions $E_x$ are
the components of an \textit{extranatural transformation}
\[
E : \text{hom}(-, A) \times (-)^{op} \to A.
\]
Observe that $E$ is also natural (in the usual sense) in the argument $A$;
we say that $\text{hom}(X, A) \times X \to A$ by evaluation is dinatural ($=$ extrana-
tural) in $X$ and natural in $A$.

\textit{Counts.} For functors $F : X \times P \to A$ and $G : P^{op} \times A \to X$ a bijection
\[
A(F(x, p), a) \cong X(x, G(p, a)) \tag{5}
\]
natural in $x, p,$ and $a$ is an adjunction with parameter $p$ (Theorem IV.7.3);
its counit, obtained by setting $x = G(p, a)$ in (5), is a collection of
\[
\varepsilon_{(p, a)} : F(G(p, a), p) \to a \tag{6}
\]
natural in $a$ and dinatural ($=$ extranatural) in $p$. This includes the case
of evaluation above.

Here is an example of the dual type of dinaturality. In any category $C$
the identity function assigns to each object $c$ the identity arrow $1_c : c \to c ,
which may be regarded as an element $1_c \in \text{hom}(c, c)$ or as an arrow
\[
1_c : * \to \text{hom}(c, c), \text{ where } * \text{ is the one-point set. Now } \text{hom}(c, c^\prime)
\text{ is the object function of a functor } C^{op} \times C \to \text{Set}, \text{ and for each arrow } f : c \to c^\prime
\text{ the identity function 1 has the evident property } f \circ 1_c = 1_{c^\prime} \circ f, \text{ which states
in the present language that 1 is a dinatural transformation}
\[
1 : * \Rightarrow \text{hom}(-, -).
\]

All three types of dinatural transformations occur in combination
with natural transformations in the previous sense (and indeed we will
usually simply call all three types "natural transformations", dropping
the "di" except where it is needed for emphasis). Thus given categories
and functors
\[
S : C^{op} \times C \to A, \quad T : A \times D^{op} \times D \to B
\]
a natural transformation $\gamma : S \Rightarrow T$ is a function which assigns to each
triple of objects $c \in C, a \in A, \text{ and } d \in D$ an arrow
\[
\gamma(c, a, d) : S(c, a, a) \to T(a, d, d)
\]
of $B$ such that (i) for $c$ and $d$ fixed, $\gamma(c, - , d)$ is natural in $a$, in the usual
sense; (ii) for $a$ and $d$ fixed, $\gamma(- , - , d)$ is dinatural in $c$; (iii) for $c$ and $a$
fixed, $\gamma(c, a, -)$ is dinatural in $d$. In the description of these natural trans-
formations any one of the categories $A, B$, or $C$ may be replaced by a
product of several categories, and in each case naturality in a product
argument $c \in C = C \times C'$ may be replaced by naturality in each argu-
ment of the pair $c = \langle c', c'' \rangle$ when the other is fixed (see Exercise 3
below). For example, in any category the operation of composition
\[
\text{hom}(b, c) \times \text{hom}(a, b) \Rightarrow \text{hom}(a, c)
\]
is natural; i.e., natural in $a$, dinatural in $b$, and natural in $c$.

The composite of two dinatural transformations need not be di-
natural at all, but any dinatural transformation $\alpha : S \Rightarrow T$ may be composed
on either side with transformations which are natural in both arguments.
If $\sigma : S' \Rightarrow S$ and $\tau : T \Rightarrow T'$ are natural transformations, the composite
arrows
\[
S'(c, c) \Rightarrow S'(c, c) \Rightarrow T(c, c) \Rightarrow S\tau(c, c) \Rightarrow T\tau(c, c)
\]
are the components of a dinatural transformation $S' \Rightarrow T'$. Here is a
more interesting case (easily proved).

**Proposition 1.** Given functors
\[
R : C \to B, \quad S : C \times C^{op} \times C \to B, \quad T : C \to B
\]
and functions (for all $c, d \in C$)
\[
\varrho(c, d) : R(c) \Rightarrow S(c, d, d), \quad \sigma(d, c) : S(d, d, c) \Rightarrow T(c)
\]
which are natural in c and dinatural in d, the function which assigns to each \( c \in C \) the composite arrow

\[ R(c) \overset{d \mapsto c}{\longrightarrow} S(c, c) \overset{\eta(c, c)}{\longrightarrow} T(c) \]

is a natural transformation \( R \Rightarrow T \).

**Exercises**

1. Prove that the unit \( \eta_x : x \Rightarrow G(p, F(x, p)) \) of an adjunction with parameter is dinatural in \( p \), and that this property is equivalent to the naturality of the adjunction itself in \( p \) (cf. IV 7, Exercise 2). Dualize.
2. Formulate the triangular identities for an adjunction with parameter.
3. (Naturality by separation of arguments.) Given \( b \in B \), a functor

\[ S : (C \times D)^p \times C \times D \rightarrow B, \]

and a function \( \beta \) assigning to \( c \in C, d \in D \) an arrow

\[ \beta_{c,d} : S(c, d, c, d) \rightarrow b \]

of \( B \), show that \( \beta : S \Rightarrow b \) is dinatural if and only if it is dinatural in \( c \) (for each fixed \( d \)) and dinatural in \( d \) (for each fixed \( c \)). State the dual result.
4. Extend the composition rule of Proposition 1 to the case when \( S \) is a functor \( C \times C \times C \Rightarrow C \Rightarrow B. \) Do the same for any odd number of factors \( C \).
5. For \( S : C^{op} \times C \Rightarrow B \) and \( b, b' \in B \), show that dinatural transformations \( b \Rightarrow S \) and \( S \Rightarrow b' \) do not in general have a well defined composite \( b \Rightarrow b' \).
6. Extending Exercises 3 and 4, find a general rule for the composition of natural transformations in many variables.

**5. Ends**

An “end” is a special (and especially useful) type of limit, defined by universal wedges in place of universal cones.

**Definition.** An end of a functor \( S : C^{op} \times C \Rightarrow X \) is a universal dinatural transformation from a constant \( e \) to \( S \); that is, an end of \( S \) is a pair \( \langle e, \omega \rangle \), where \( e \) is an object of \( X \) and \( \omega : e \Rightarrow S \) is a wedge (a dinatural transformation) with the property that to every wedge \( \beta : x \Rightarrow S \) there is a unique arrow \( h : x \Rightarrow e \) of \( B \) with \( \beta_a = \omega_a h \) for all \( a \in C \).

Thus for each arrow \( f : b \Rightarrow c \) of \( C \) there is a diagram

\[ \begin{array}{ccc}
    x & \overset{h}{\rightarrow} & e \\
    \downarrow & & \downarrow \\
    S(b, b) & \overset{\beta(b, b)}{\rightarrow} & S(b, c) \\
    \downarrow & & \downarrow \\
    S(c, c) & \overset{\omega(c, c)}{\rightarrow} & S(c, c) \\
    \end{array} \]

such that both quadrilaterals commute (these are the dinaturality conditions); the universal property of \( \omega \) states that there is a unique \( h \) such that both triangles (at the left) commute.

The uniqueness property which applies to any universal states in this case that if \( \langle e, \omega \rangle \) and \( \langle e', \omega' \rangle \) are two ends for \( S \), there is a unique isomorphism \( u : e \Rightarrow e' \) with \( \omega : u \Rightarrow \omega \) (i.e., with \( \omega_e = u \cdot \omega_e \) for each \( e \in C \)).

We call \( \omega \) the ending wedge or the universal wedge, with components \( \omega_e \), while the object \( e \) itself, by abuse of language, is called the “end” of \( S \) and is written with integral notation as

\[ e = \int S(c, c) = \text{End of } S. \]

Note that the “variable of integration” \( e \) appears twice under the integral sign (once contravariant, once covariant) and is “bound” by the integral sign, in that the result no longer depends on \( e \) and so is unchanged if “\( e \)” is replaced by any other letter standing for an object of the category \( C \). These properties are like those of the letter \( x \) in the usual integral \( \int f(x) \, dx \) of the calculus.

Natural transformations provide an example of ends. Two functors \( U, V : C \Rightarrow X \) define a functor \( \text{hom}(U, V) : C^{op} \times C \Rightarrow Set \), and if \( Y \) is any set, a wedge \( \tau : Y \Rightarrow \text{hom}(U, V) \), with components

\[ \tau_c : Y \Rightarrow \text{hom}(U, C, V_c), \quad c \in C, \]

assigns to each \( y \in Y \) and to each \( c \in C \) an arrow \( \tau_{y,c} : U c \Rightarrow V_c \) of \( X \) such that for every arrow \( f : b \Rightarrow c \) one has the “wedge condition” \( V_f \circ \tau_{y,b} = \tau_{y,f} \circ U f \). But this condition is just the commutativity of the square

\[ \begin{array}{ccc}
    U b & \overset{f}{\longrightarrow} & V b \\
    \downarrow \quad \downarrow & & \quad \downarrow \quad \downarrow \\
    U c & \overset{f}{\longrightarrow} & V c \end{array} \]

which asserts that \( \tau_{y,b} \) for fixed \( y \) is a natural transformation \( \tau_{y,-} : U \Rightarrow V \). Thus, if we write \( \text{Nat}(U, V) \) for the set of all such natural transformations, the assignment \( y \Rightarrow \tau_{y,-} \) is the unique function \( Y \Rightarrow \text{Nat}(U, V) \) which makes the following diagram commute.

\[ \begin{array}{ccc}
    Y & \overset{\tau}{\longrightarrow} & \text{hom}(U, C, V_c) \\
    \downarrow \quad \downarrow \quad \downarrow & & \quad \downarrow \quad \downarrow \quad \downarrow \\
    \text{Nat}(U, V) & \overset{\omega}{\longrightarrow} & \text{hom}(U, C, V_c), \end{array} \]

where \( \omega \) assigns to each natural \( \lambda : U \Rightarrow V \) its component \( \lambda_c : U c \Rightarrow V_c \). This states exactly that \( \omega \) is a universal wedge. Hence

\[ \text{Nat}(U, V) = \int \text{hom}(U, C, V_c) ; \quad U, V : C \Rightarrow X. \]
Every end is manifestly a limit—specifically, a limit of a suitable diagram in $X$ made up of pieces like those pieces $S(b, b) \to S(b, c) \to S(c, c)$, one for each $f$ in $C$, which come up in the diagram (1) defining an end. This can be stated formally in terms of the following construction (to be used only in this section) of a category $C^\downarrow$ depending on $C$. The objects of $C^\downarrow$ are all symbols $c^\downarrow$ and $f^\downarrow$ for $c \in C$ and $f$ an arrow in $C$ (note especially that $c^\downarrow$ and $(1,)^\downarrow$ are different objects). The arrows of $C^\downarrow$ are the identity arrows for these objects, plus for each arrow $f : b \to c$ in $C$ two arrows

$$b^\downarrow \to f^\downarrow \to c^\downarrow$$

in $C^\downarrow$. The only meaningful compositions for these arrows in $C^\downarrow$ are compositions with one factor an identity arrow. Thus we have defined a category $C^\downarrow$, called the subdivision category of $C$.

Each functor $S : C^\op \times C \to X$ defines a functor $S^\downarrow : C^\downarrow \to X$ by the assignments indicated (from top to bottom) in the following figure for a typical $f : b \to c$ in $C$:

$$
\begin{array}{cccc}
C^\downarrow & b^\downarrow & f^\downarrow & c^\downarrow \\
S^\downarrow & \downarrow & \downarrow & \downarrow \\
X & S(b, b) & S(f, b) & S(c, c)
\end{array}
$$

Inspection of this figure shows that a cone $\tau : x \to S^\downarrow$ is exactly the same thing as a wedge $\omega : x \to \omega S$. This proves that a limit of $S^\downarrow$ is an end of $S$, in the following sense.

**Proposition 1.** For any functor $S : C^\op \times C \to X$ and the associated functor $S^\downarrow : C^\downarrow \to X$, as defined above, there is an isomorphism

$$\theta : \{ S(c, c) \cong \lim \left[ S^\downarrow : C^\downarrow \to X \right] \}.$$  

In more detail, if either the indicated end or the indicated limit exists, then both exist, and there is a unique arrow $\theta$ in $X$ such that the diagram

$$
\begin{array}{ccc}
\{ S(c, c) \} & \to & S(c, c) \\
\theta & \downarrow & \downarrow \\
\lim S^\downarrow & \to & S^\downarrow(c)
\end{array}
$$

commutes for every $c \in C$, where $\omega$ is the ending wedge and $\lambda$ the limiting cone: moreover, this arrow $\theta$ is an isomorphism.

**Corollary 2.** If $X$ is small-complete and $C$ is small, every functor $S : C^\op \times C \to X$ has an end in $X$.

The Proposition above has reduced ends to limits. The converse is easier. Every limit may be regarded as an end!

### Ends

**Proposition 3.** For each functor $T : C \to X$ let $S$ be the composite functor

$$C^\op \times C \xrightarrow{Q} C \xrightarrow{\tau} X$$

where $Q$ is the second projection $<c, c'> \mapsto c'$ of the product. Then $<e, \tau : e \mapsto T>$ is a limit for $T$ in $X$ if and only if $<e, \tau : e \mapsto S>$ is an end for $S$ in $X$.

**Proof.** The components $\tau_e$ of a cone $e \mapsto T$ make the triangle

$$
\begin{array}{ccc}
e & \to & Tb = S(b, b) \\
\tau_e & \downarrow & \downarrow \\
Tc = S(c, c) & \xrightarrow{S(1, f) = T_1} & Tc = S(c, c)
\end{array}
$$

commutes ($S(-,-)$ is "dummy" in the first variable), and this in turn states exactly that $\tau : e \mapsto S$ is a wedge. It follows that $\tau$ is universal as a cone if and only if it is universal as a wedge.

This conclusion reads: There is an isomorphism

$$\{ S(c, c) \} \cong \lim_{c} Tc,$$

valid when either the end or the limit exists, carrying the ending wedge to the limiting cone; the indicated notation thus allows us to write any limit as an integral (an end) without explicitly mentioning the dummy variable (the first variable of $S$).

A functor $H : X \to Y$ is said to preserve the end of a functor $S : C^\op \times C \to X$ when $\omega : e \mapsto S$ an end of $S$ in $X$ implies that $H \omega : H e \mapsto HS$ is an end for $H S$; in symbols

$$H \{ S(c, c) \} \cong \{ H S(c, c) \}.$$

Similarly, $H$ creates the end of $S$ when to each end $v : y \mapsto HS$ in $Y$ there is a unique wedge $\omega : e \mapsto S$ with $H \omega = v$, and this wedge $\omega$ is an end of $S$.

Since an end (of the functor $S$) is the same thing as a limit (of the corresponding $S^\downarrow$), the properties we have established for the preservation of limits carry over to the preservation of ends. For example, the hom-functors preserve (and reverse, see §6) ends:

$$\{ X(x, \{ S(c, c) \} ) \} \cong \{ X(x, S(c, c)) \},$$

$$\{ X(\{ S(c, c), x \} ) \} \cong \{ X(S(c, c), x) \}.$$
6. Coends

The definition of the coend of a functor $\mathcal{F} : C^{op} \times C \to X$ is dual to that of an end. A coend of $\mathcal{F}$ is a pair $\langle d, c \rangle$ consisting of an object $d \in X$ and a dinatural transformation $\zeta$ (a wedge), universal among dinatural transformations from $\mathcal{F}$ to a constant. The object $d$ (when it exists, unique up to isomorphism) will usually be written with an integral sign and with the bound variable $c$ as superscript; thus

$$S(c, c) \xrightarrow{\zeta} \int S(c, c) = d.$$  

The formal properties of coends are dual to those of ends.

Coends are familiar under other names. For example, the tensor product of modules over a ring $R$ is a coend. Specifically, a ring $R$ is an $R$-bimodule with one object (which we call $R$ again) and with arrows the elements $r \in R$, composition of arrows being their product in $R$. A left $R$-module $B$ is an additive functor $R \to Ab$ which sends the (one) object $R$ to the abelian group $B$ and each arrow $r$ in $R$ to the scalar multiplication $r \cdot B : b \mapsto rb$ in $B$. Similarly, a right $R$-module $A$ is an additive functor $R^{op} \to Ab$ (contravariant on $R$ to $Ab$). If $\otimes$ is the usual tensor product in $Ab$, then $R \to A \otimes B$ is a bifunctor $R^{op} \times R \to Ab$. Moreover, the coend

$$\int A \otimes B = A \otimes_R B$$

is exactly the usual tensor product over $R$. Indeed, a wedge $\zeta$ from the bifunctor $A \otimes B$ to an abelian group $M$ is precisely a (single) morphism $\varphi : A \otimes B \to M$ of abelian groups such that the diagram

$$\begin{array}{ccc}
A \otimes B & \xrightarrow{a \otimes b} & B \\
\downarrow & & \downarrow \varphi \\
A \otimes B & & M
\end{array}$$

commutes for every arrow $r$ in $R$. With the above interpretation of modules as functors, this means for elements $a \in A$ and $b \in B$ that

$$\varphi(a \otimes b) = \varphi(a \otimes r b).$$

Therefore $M$ is an end precisely when $M$ is $A \otimes B$ modulo all $a \otimes b - a \otimes r b$, and this is precisely the usual description of the tensor product $M = A \otimes B$.

The point of these observations is not the reduction of the familiar to the unfamiliar (tensor products to coends) but the extension of the familiar to cover many more cases. If $B$ is any monoidal category with multiplication $\square$, as in Chapter VII, then any two functors $T : P^{op} \to B$ and $S : P \to B$ have a "tensor product"

$$\square_p S = \int (T p) \square (S p).$$

Coends

an object of $B$. The simplicial category $\mathcal{A}$ of § VII.5 has a functor $\mathcal{A} : \mathcal{A} \to \text{Top}$ (each ordinal $n + 1$ realized by the $n$-dimensional affine simplex), while any $\mathcal{A} : \mathcal{A}^{op} \to \text{Set}$ is called a simplicial set. Now the copower $S \cdot X$ (set $S$ times space $X$) is just the disjoint union $\text{copower of copies of } X$. Hence $n \to S^n \cdot \mathcal{A}$ is a functor $\mathcal{A}^{op} \times \mathcal{A} \to \text{Top}$ and the coend

$$\int S^n \cdot \mathcal{A}$$

is the usual geometric realization (Lamotke [1968], p. 34; May [1976], p. 55) of the simplicial set $S$. The coend formula describes the geometric realization in one gulp: Take the disjoint union of affine $n$-simplices, one for each $i \in S_n$, and paste them together according to the given face and degeneracy operations (arrows of $\mathcal{A}$). There is a similar efficient description of the (Stasheff-Milgram) classifying space of a topological monoid (best situated in the category $\text{CGHaus}$ of § VII.8); see Mac Lane [1970].

Exercises

1. For $S : C^{op} \times C \to \text{Set}$, prove that the set $\text{Wedge} (\ast, S)$ of all wedges $\omega : \ast \to X$ from the one-point set $\ast$ to $S$ is an end of $S$, with ending wedge given by $\omega \circ \ast \in \text{Set}(c, c)$. Compare with the explicit description of a limit in Set as a set of cones.

2. Show directly (without using limits) that a category $X$ with all small products and with equalizers has all small ends (cf. the corresponding proof for limits in § V.2).

3. To each category $C$ there is a "twisted arrow category" $C^\ast$ with objects the arrows $f : a \to b$ of $C$ and arrows $\langle h, k \rangle : f \to f'$ the arrows $h : a \to a$ (note the twist!) and $k : b \to b'$ such that $f' = k f h$. Then $\langle f : a \to b \rangle \mapsto \langle a, b \rangle$ is a functor $C \to C^\ast \times C$. For any $S : C^{op} \times C \to \text{Set}$, prove that cones $c \to S$ correspond to wedges $c \to S$, and use this fact to give another proof of the reduction of ends to limits (Proposition 5.1).

4. Let $\text{Fin}$ (the skeletal category of finite non-empty sets) be the category with objects finite nonzero ordinals $\omega$ and arrows all functions $f : \omega \to \omega$. For each set $X$, an assignment $n \mapsto \mathcal{V} \in \text{Fin}$ defines the object function of a functor $\text{Fin}^{op} \to \text{Set}$. For each ring $R$, the assignment $n \mapsto R^n$ becomes a covariant functor $\mathcal{V} : \text{Fin} \to \text{R-Mod}$ if each function $f : n \to m$ (arrow in $\text{Fin}$) takes a list $a_0, \ldots, a_{n-1} \in R^m$ to $b_0, \ldots, b_{m-1} \in R^m$, where $b_i = \sum a_i$ the sum over all those $j \in \omega$ with $f_j = i$. Show that the free $R$-module generated by the set $X$ is the coend

$$\int X^\ast \cdot R(\ast),$$

and show that this formula is essentially the usual description of the elements of the free module as finite formal sums $\sum_{i=1}^n x_i a_i$. Define the tensor product as a coend $\int (S \star) \cdot (T \circ) c$, where $\star$ denotes the copower. Show that the tensor product is a functor $\mathcal{D}^{op} \times \text{Set} \to \text{Set}^{op} \to D$.
7. Ends with Parameters

The basic formal properties for ends are much like those for integrals in calculus. All these properties will apply equally well to limits (regarded as ends with a dummy variable).

**Proposition 1.** (End or limit of a natural transformation). Given a natural transformation \( \gamma : S \to S' \) between functors \( S, S' : C^{op} \times C \to X \) which both have ends \( \langle e, \omega \rangle \) and \( \langle e', \omega' \rangle \), respectively, there is a unique arrow \( g = \int \gamma_{c, c} : e \to e' \) in \( X \) such that the following diagram commutes for every \( c \in C \):

\[
\begin{array}{c}
\int S(c, c) \xrightarrow{\int \gamma_{c, c}} \int S'(c, c) \\
\int S'(c, c) \xrightarrow{\int \omega'} \int S'(c, c) \\
\int S'(c, c) \xrightarrow{\int \omega} \int S'(c, c) \\
\int S'(c, c) \xrightarrow{\int \gamma_{c, c}} \int S'(c, c)
\end{array}
\]

(1)

**Proof.** The composites \( \gamma_{c, c} \circ \omega \) define a wedge, so \( g \) exists and is unique by the universality of the wedge \( \omega' \).

We call the arrow \( g \) the end of the natural transformation \( \gamma \).

Composing \( \gamma \) with another \( \gamma' : S' \to S'' \) yields the rule

\[
\int (\gamma' \circ \gamma)_{c, c} = \int (\gamma'_{c, c} \circ \gamma_{c, c}) = \int (\gamma'_{c, c} \circ \gamma_{c, c}) = \int (\gamma_{c, c} \circ \gamma_{c, c}).
\]

(2)

By this composition rule, a limit (or an end) involving a parameter \( p \) (in some category \( P \)) can be shown to be a functor of that parameter in the following sense.

**Theorem 2.** (Parameter Theorem for ends and limits). Let \( T : P \times C^{op} \times C \to X \) be a functor such that \( T(p, -, -) \) for each object \( p \in P \) has an end

\[
\omega_p : \int T(p, c, c) \to T(p, -, -)
\]

in \( X \). Then there is a unique functor \( U : P \to X \) with object function \( U p = \int T(p, c, c) \) such that the components of the wedge (3) for each \( c \in C \) define a transformation \( (\omega_p)_* : U p \to T(p, c, c) \), natural in \( p \).

**Proof.** Each arrow \( s : p \to q \) of \( P \) defines a natural transformation \( \gamma = T(s, -,-) : T(p, -,-) \to T(q, -,-) \). Hence the arrow function of the desired functor \( U \) must have \( U s = \int T(s, c, c) \), defined as in (1), and the composition rule (2) shows that this definition of \( U s \) does determine a functor \( U : P \to X \).

The functor \( U \) will be written \( U = \int T(-, c, c) \); thus

\[
\int T(-, c, c) \overset{p}{\to} \int T(p, c, c), \quad \int T(-, c, c) \overset{s}{\to} \int T(s, c, c).
\]

(4)

7. Ends with Parameters

The notation suggests that this functor \( U \) is itself an end. Indeed, regard \( \int T(-, c, c) \) as an object of the functor category \( X^{P^{op}} \) and rewrite \( T : P \times C^{op} \times C \to X \) as the functor \( T^* : C^{op} \times C \to X^P \) given on arrows (or objects) \( f, f' \) of \( C \) by

\[
T^*(f, f') = T(-, f, f'): P \to X.
\]

Put differently, \( T^* \) is the image of \( T \) under the standard adjunction

\[
\text{Cat}(P \times C^{op} \times C, X) \cong \text{Cat}(C^{op} \times C, X^P).
\]

**Theorem 3.** (Parameter theorem, continued). Under the same hypotheses on \( T \), the functor \( T^* \) has the end

\[
\omega^*_p : \int T(-, c, c) \to T(-, c, c)
\]

where \( (\omega^*_p)_p = (\omega_p)_c \), for all \( p \in P \) and \( c \in C \).

**Proof.** The end \( \int T(-, c, c) \) is an object of \( X^P \), while \( T^* \) is a functor with codomain \( X^P \). By the previous theorem, the arrows \( (\omega_p)_c \) of \( X \) provide for each \( c \) an arrow of \( X^P \) (a natural transformation)

\[
\omega^*_p : \int T(-, c, c) \to T(-, c, c)
\]

its component at \( p \) is \( (\omega^*_p)_p = (\omega_p)_c \). Moreover, varying \( c \), \( \omega^* \) is a wedge

\[
\int T(-, c, c) \to T^*.
\]

It is a universal wedge, for, given any object \( F \in X^P \) and any wedge \( \beta : F \to T^* \), each component \( \beta_p \) factors uniquely through the corresponding component \( \omega^*_p \), so \( \beta \) itself factors uniquely through \( \omega^* \). This gives the end for \( T^* \), as required.

This theorem can also be formulated wholly in terms of the functor category \( X^P \), as was done in the case of limits in Theorem V.3.1.

**Exercises**

1. (Dubuc) Construct a functor category \( X^P \) and a functor \( T : C \to X^P \) which has a limit not a pointwise limit (Suggestion: Take \( C = 2 \)).

2. State and prove the parameter theorem for coends.

3. If \( X \) is small complete and \( C \) is small, use Proposition 1 to prove that \( \text{Lim} : X^{C^{op}} \to X \) is a functor (cf. Ex. V.2.3).

4. For any categories \( X \) and \( P \), show that the functor \( X^P \to X^{[P]} \) induced by inclusion (of the discrete subcategory \([P]\)) creates ends and coends (cf. Theorem V.3.2).
8. Iterated Ends and Limits

We now describe when the "double integral" can be obtained as an "iterated" integral (Fubini!).

Proposition. Let $S : P^{op} \times P \times C^{op} \times C \to X$ be a functor such that the end \( \int S(p, q, c, c) \) exists for all pairs \( \langle p, q \rangle \) of objects of $P$; by the parameter theorems, regard these ends as a bifunctor $P^{op} \times P \to X$, and regard $S$ as a bifunctor $(P \times C)^{op} \times (P \times C) \to X$. Then there is an isomorphism

\[
\theta : \int \int S(p, q, c, c) \cong \int \int S(p, q, c, c).
\]

Indeed, the "double end" on the left exists if and only if the end \( \int S(p, q, c, c) \) on the right exists, and then there is a unique arrow $\theta$ in $X$ such that the diagram

\[
\begin{array}{ccc}
\int S(p, q, c, c) & \xrightarrow{5e} & \int S(p, q, c, c) \\
\theta \downarrow & & \downarrow \theta \\
\int \int S(p, q, c, c) & \xrightarrow{5e} & \int \int S(p, q, c, c)
\end{array}
\]

commutes, where the horizontal arrows $\xi, \eta, \omega$ are the universal wedges belonging to the corresponding ends; moreover, the arrow $\theta$ is an isomorphism.

Proof. For each $\langle p, q \rangle \in P \times P$ we are given the end

\[
\omega_{p,q} : \int \int S(p, q, c, c) \to S(p, q, c, c).
\]

For any $x \in X$ each $P$-indexed family $\eta_x : \cdot \to \int S(p, q, c, c)$ of arrows of $X$ determines a $(P \times C)$-indexed family $\eta_{p,c}$ as the composites

\[
\eta_{p,c} : x \xrightarrow{5e} \int S(p, q, c, c) \xrightarrow{\omega_{p,q}} S(p, q, c, c);
\]

for $p$ fixed, $\eta_{p,-}$ is trivially a wedge in $c$. Conversely, since $\omega_{p,q}$ is universal, every $(P \times C)$-indexed family which is natural in $c$ for each $p$ is such a composite, for a unique family $q$. Now $q$ or $\xi$ is extranatural in $p$ (the latter for some $c$) if and only if the corresponding square below

\[
\begin{array}{ccc}
x & \xrightarrow{5e} & \int S(p, q, c, c) \\
\eta_x \downarrow & & \downarrow \eta_x \\
\int S(q, q, c, c) & \xrightarrow{\chi} & \int S(q, q, c, c)
\end{array}
\]

\[
\begin{array}{ccc}
x & \xrightarrow{5e} & S(p, q, c, c) \\
\eta_x \downarrow & & \downarrow \eta_x \\
S(q, q, c, c) & \xrightarrow{S_{p,q,c}} & S(p, q, c, c)
\end{array}
\]

commutes for each arrow $s : p \to q$ in $P$. Also, the first square commutes precisely when it commutes after composition with the arrows $\omega_{p,q,c}$ for all objects $c$. Form the cubical diagram with these two squares as front and back faces and with edges $1, \omega_{p,q,c}, \omega_{p,q,c},$ and $\omega_{q,c}$ (front to back). By our definitions the four side faces involving these edges commute; hence the front square commutes if and only if the back square commutes for all $c$. Therefore $q$ is a wedge (in $p$) if and only if $\xi$ is a wedge (in $p, q, c$), so that wedges from $x$ to $\int \int S(p, q, c, c)$ correspond one-one to wedges from $x$ to $S$. Since the end is a universal wedge, and since a universal is determined up to isomorphism, this gives the isomorphism $\theta$ of the proposition.

Note one essential point: This proposition reduces double to iterated integrals provided the inner integral $\int \int S(p, q, c, c)$ exists for all pairs $\langle p, q \rangle$ (not just for $p = q$). The case of limits involves no such refinement.

The familiar result on change of order of integrals follows from this one, expanding a double integral in two ways.

Corollary. Let $S : P^{op} \times P \times C^{op} \times C \to X$ be a functor such that the ends $\int S(p, q, c, c)$ and $\int S(p, q, b, c)$ exist, for all $p, q \in P$ and $b, c \in C$. By the parameter theorems regard these ends as bifunctors $(P, q, b, c)$ respectively; then there is an isomorphism

\[
\theta : \int \int S(p, q, c, c) \cong \int \int S(p, q, c, c).
\]

Indeed, the (outside) iterated end on the left exists if and only if the (outside) iterated end on the right exists, and the isomorphism $\theta$ is the unique arrow in $X$ such that the diagrams

\[
\begin{array}{ccc}
\int \int S(p, q, c, c) & \xrightarrow{5e} & \int \int S(p, q, c, c) \\
\theta \downarrow & & \downarrow \theta \\
\int \int S(p, q, c, c) & \xrightarrow{5e} & \int \int S(p, q, c, c)
\end{array}
\]

commute for all $p \in P$ and $c \in C$, where the horizontal arrows are the appropriate components of the universal wedges for the integrals involved.

These results include the corresponding facts for limits and colimits. Thus, for a functor $F : P \times C \to X$ with $P$ and $C$ small, $X$ complete

\[
\lim_p \lim_q F(p, q) \cong \lim_q \lim_p F(p, q) \cong \lim_p \lim_q F(p, q),
\]

with the corresponding formula for colimits.
X. Kan Extensions

If \( M \) is a subset of \( C \), any function \( t : M \to A \) to a non-empty set \( A \) can be extended to all of \( C \) in many ways, but there is no canonical or unique way of defining such an extension. However, if \( M \) is a subcategory of \( C \), each functor \( T : M \to A \) has in principle two canonical (or extreme) "extensions" from \( M \) to functors \( L, R : C \to A \). These extensions are characterized by the universality of appropriate natural transformations; they need not always exist, but when \( M \) is small and \( A \) is complete and cocomplete they do exist, and can be given as certain limits or as certain ends. These "Kan extensions" are fundamental concepts in category theory. With them we find again that each fundamental concept can be expressed in terms of the others. This chapter begins by expressing adjoints as limits and ends by expressing "everything" as Kan extensions.

1. Adjoint and Limits

Limits and colimits, if they exist for all functors \( J \to C \), provide respectively right and left adjoints for the diagonal functor \( \Delta : C \to C \):

\[
\begin{array}{ccc}
\text{Lim} & \text{Colim} \\
\downarrow & \downarrow \\
C & C' \end{array}
\]

(= right adjoint of \( \Delta \)).

Conversely, left adjoints can be interpreted as limits. First note that an initial object in any category \( C \) is a limit:

\[
\text{Initial object } C = \text{Colim}(\mathcal{O} \to C) = \text{Lim}(1_C : C \to C).
\]

where \( \mathcal{O} \) denotes the empty category (the ordinal 0) and \( \mathcal{O} \to C \) is the empty functor. The definition of the initial object \( \mathcal{O} \) states exactly that it is the colimit of the empty functor. Moreover, the unique arrows \( \mu_c : e \to c \), one for each \( c \), define a cone \( e \to \text{Id}_C \). If \( \lambda : d \to \text{Id}_C \) is a cone from some other vertex, then there is a unique \( f : d \to e \) with all \( \mu_c \circ f = \lambda_c \); indeed, this equation for \( c = e \) shows that \( f \) must be \( \lambda_e \), and for \( f = \lambda_e \), this equation \( \mu_c \circ \lambda_e = \lambda_c \) does hold because \( \lambda \) is a cone over \( \text{Id}_C \). This
proves \( e = \text{Lim } \text{Id}_C \). The converse property, that any limit of \text{Id} is initial, is a special case of

**Lemma 1.** If \( \lambda : d \rightarrow \text{Id}_C \) is a cone over the identity functor and \( F : J \rightarrow C \) is a functor such that \( \lambda F : d \rightarrow F \) is a limiting cone for \( F \), then \( d \) is initial in \( C \).

**Proof.** Since \( \lambda \) is a cone, the triangles

\[
\begin{array}{ccc}
& d & \\
\lambda_f & \downarrow & \lambda_g \\
F_i & \\
\end{array}
\]

commute for each \( i \in J \) and each arrow \( f \) in \( C \). But \( \lambda F \) is a limiting cone, so the first two triangles prove \( \lambda_g = 1 \). Then by the third triangle, \( f = \lambda_i \); there is a unique arrow \( f \) from \( d \) to each \( c \), and \( d \) is indeed initial.

This result reduces initial objects to limits. Now a functor \( G : A \rightarrow X \) has a left adjoint precisely when for each \( x \) the comma category \((x \downarrow G)\) of all pairs \( \langle g : x \rightarrow Ga, a \rangle \), has an initial object. In this way we can express the left adjoint by limits. Recall that \( \langle g, a \rangle \mapsto a \) defines the (second) projection \( Q : (x \downarrow G) \rightarrow A \) of the comma category.

**Theorem 2.** (Formal criterion for the existence of an adjoint.) A functor \( G : A \rightarrow X \) has a left adjoint if and only if both

(i) \( G \) preserves all limits which exist in \( A \);

(ii) For each \( x \in X \), \( \text{Lim}(Q : (x \downarrow G) \rightarrow A) \) exists in \( A \).

When this is the case, a left adjoint \( F \) is given on each \( x \in X \) as

\[
Fx = \text{Lim}(Q : (x \downarrow G) \rightarrow A),
\]

and the left adjoint of each arrow \( g : x \rightarrow Ga \) is the component \( \lambda_g : Fx \rightarrow Qa = a \) of the limiting cone \( \lambda \) for the limit (3).

**Proof.** Since right adjoints preserve all limits, (i) is necessary. Since a left adjoint \( F \) to \( G \) has each \( \langle \eta_x : x \rightarrow GFx, Fx \rangle \) an initial object in \((x \downarrow G)\), any functor on this comma category has a limit (namely, its value on that initial object). Hence (ii) is necessary.

Now we consider the converse. By hypothesis (ii) the composite functor

\[
(y \downarrow G) \xrightarrow{\text{Id}} (y \downarrow G) \xrightarrow{Q} A
\]

has a limit in \( A \) for each \( y \in X \). By hypothesis (i), \( G \) preserves all limits; hence, using the Lemma of §V.6, \( Q \) creates all limits. Therefore \text{Id} has a limit on \((y \downarrow G)\). This limit is, by (2), an initial object there, say \( y \rightarrow Ga \). But then \( a \) is a value \( a = Fy \) for a left adjoint \( F \), and

\[
Fy = Q[\text{Lim}(y \downarrow G)] = \text{Lim}(Q : (y \downarrow G) \rightarrow A)
\]

(since \( Q \) preserves this limit which it has created!). This is the desired formula; the rule for finding the left adjoints follows at once.

### Weak Universality

Given a functor \( G : A \rightarrow X \) and an object \( x \in X \), a weak universal arrow from \( x \) to \( G \) is a pair \( \langle r, w : x \rightarrow Gr \rangle \) consisting of an object \( r \in A \) and an arrow \( w \) of \( X \), as indicated, such that for every arrow \( f : x \rightarrow Ga \) there exists an arrow \( f' : r \rightarrow a \) with \( f = Gf' \cdot w \). This is just the definition of universal arrow, except that \( f' \) is not required to be unique. By the same device (Freyd) we can modify all the various types of universals, defining weak products, weak limits, weak coproducts (existence but not uniqueness in each case).

As an application, we give a second proof of the Freyd existence theorem for an initial object (Theorem V.6.1).

**Theorem 1.** If \( D \) is a small complete category with small hom-sets, then \( D \) has an initial object if and only if it has a small set \( S \) of objects which is weakly initial: For every \( d \in D \) there exists \( s \in S \) and an arrow \( s \rightarrow d \).

**Proof.** Let \( S \) also denote the full subcategory of \( D \) with the objects \( s \); since \( D \) has small hom-sets, \( S \) is still small, so by completeness the inclusion functor \( F : S \rightarrow D \) has a limiting cone \( \mu : v \rightarrow F \). We shall prove \( v = \text{Lim } F \) initial in \( D \).

### Exercises

1. State the dual of Theorem 2.
2. (Bénabou, formal criterion for representability.) Let \( C \) have small hom-sets, while \( * \) is the one-point set. Prove: A functor \( K : C \rightarrow \text{Set} \) is representable if and only if \( (q, a) \) is a limit, \( K \) preserves all limits which exist in \( C \), and \( \text{Id} : \text{Set} \rightarrow \text{Set} \) has a limit in \( (x \downarrow G) \) for each \( x \in X \) (since \( C \) is a category).
3. (Formal criterion for a universal arrow.) Let \( X \) have small hom-sets. Prove that there is a universal arrow from \( x \in X \) to \( G : A \rightarrow X \) if and only if \( \text{Lim} \ (x \downarrow G) \rightarrow A \) exists in \( A \).
4. (Refinements of formal existence criteria).
   (a) In the Theorem, show that condition (i) may be replaced by "\( G \) preserves the limits required to exist in (ii)."
   (b) In Theorem 2, show that condition (i) may be replaced by "\( K \) preserves the limit of \( Q \).
5. (Representable and adjoints; Bénabou.) Let \( C \) have small hom-sets, and construct from each \( K : C \rightarrow \text{Set} \) the category \( C_{\text{rep}} \) obtained by adjoining to \( C \) one new object \( \infty \) with new hom-sets \( C_{\text{rep}}(\infty, c) = \{Kc, C_{\text{rep}}(\infty, \infty) = 0\} \), the one-point set and \( C_{\text{rep}}(\infty, \infty) = 0 \), the empty set, with appropriate composition. Let \( J_{\text{rep}} : C \rightarrow C_{\text{rep}} \) be the inclusion. Prove that \( K \) is representable if and only if \( J_{\text{rep}} \) has a left adjoint.
First, for every $d \in D$ we choose $s \to d$ and define $\gamma_d$ as the composite $\gamma_d : t_{\Delta_d} s \to d$. We claim that $\gamma : v \to 1_{D_0}$ is a cone. For, take any arrow $f : d \to d'$ and form the diagram

Since $S$ is small complete, there is a pull-back of $s \to d' \to s'$ with vertex $p$; since $S$ is weakly initial, there is an arrow $v \to s' \to p$. The two composite arrows $s' \to p \to s$ and $s' \to p' \to s'$ are in $S$ because $S$ is full, so the two upper quadrilaterals commute (the $\mu$ is a cone), while the pentagon commutes because $p$ is a pull-back. This proves $\gamma$ is a cone.

If, in particular, we choose $v \to s \to s$ to be $\mu$, then $\gamma$ is a cone such that the composite $\gamma F : v \to F$ is the limiting cone $\mu$. By Lemma 1.1, $v$ is initial in $D_0$. q.e.d.

Carefully examined, this proof is just a refinement of the previous one (§ V.6), where we took first a product $\Pi S$ (to get a single weakly initial object) and then a suitable equalizer. In this proof, these operations are combined to one: $\operatorname{Lim} F$ for $F : S \to D$.

3. The Kan Extension

Given a functor $K : M \to C$ and a category $A$ we consider the functor category $A^C$, with objects the functors $S : C \to A$ and arrows the natural transformations $\sigma : S \to S'$, and we define the functor $A^K : A^C \to A^M$ by the assignments

$$\langle \sigma : S \to S' \rangle \mapsto \langle \sigma K : SK \to S'K \rangle.$$

The problem of Kan extension is to find left and right adjoints to $A^K$. We consider this problem first for right adjoints.

**Definition.** Given functors $K : M \to C$ and $T : M \to A$, a right Kan extension of $T$ along $K$ is a pair $R, \varepsilon : RK \to T$ consisting of a functor $R \in A^C$ and a natural transformation $\varepsilon$ which is universal as an arrow from $A^K : A^C \to A^M$ to $T \in A^M$.

As always, this universality determines the functor $R = \operatorname{Ran}_K T$ uniquely, up to natural isomorphism. In detail, this universality means that for each pair $S, \alpha : SK \to T$ there is a unique natural transformation $\sigma : S \to R$ such that $\alpha = \varepsilon \cdot \sigma K : SK \to T$. The diagram is

$$\begin{array}{c}
\text{Nat}(S, R) \cong \text{Nat}(SK, T),
\end{array}$$

natural in $S$; again, this natural bijection determines $\operatorname{Ran}_K T$ from $K$ and $T$. It is a right Kan extension because it appears at the right in the hom-set "Nat" (But note that some authors call this $R$ a "left" Kan extension).

By the general result that universal arrows from the functor $A^K$ to all the objects $T$ together constitute a left adjoint to the functor $A^K$, it follows that if every functor $T \in A^M$ has a right Kan extension $\langle R, \varepsilon : RK \to T \rangle$, then $T \to R$ is (the object function of) a right adjoint to $A^K$ and $\varepsilon$ is the unit of this adjunction. In the sequel, we shall construct right Kan extensions for individual functors $T$, which may exist when (the whole of) the right adjoint of $A^K$ does not exist.

A useful case is that in which $M$ is a subcategory of $C$ and $K : M \to C$ the inclusion $M \subseteq C$; in this case, $A^K$ is the operation which restricts the domain of a functor $S : C \to A$ to the subcategory $M$. Conversely, for given $T : M \to A$ we consider extensions $E : C \to A$ of $T$ to $C$. Then $Ec \in A$ must have for each arrow $f : c \to m$ in $C$ an arrow $Ef : Ec \to Tm$ in $A$, and these arrows must constitute a cone from the vertex $Ec$ to the base $T$, where $T$ is regarded as a functor on the category of arrows $f : c \to m$ (fixed $c$ to variable $m$). These arrows $f$ are the objects of the comma category $(c \downarrow K)$, so a natural choice of $Ec$ is the limit (with $Ef$ the limiting cone) of the functor $T : (c \downarrow K) \to A$.

$$\begin{array}{c}
\text{This procedure (compare (1.3)) works in general. For each } c \in C,
\text{the comma category } (c \downarrow K) \text{ has the objects } \langle f, m \rangle, \text{ written } f \text{ for short,}
\text{where } f : c \to Km \in C, \text{ while } \langle f, m \rangle \mapsto m \text{ is (the object function of) the projection functor }
Q : (c \downarrow K) \to M.
\end{array}$$

**Theorem 1.** (Right Kan extension as a point-wise limit). Given $K : M \to C$, let $T : M \to A$ be a functor such that the composite
(c\mid K)\to M\to A has for each c\in C a limit in A, with limiting cone \lambda, written
\[ Rc = \lim_{c\in (c\mid K)}(c\mid K)-\Delta M\to A = \lim_f Tm, \quad f\in (c\mid K). \]
(3)
Each g: c\to c' induces a unique arrow
\[ Rg: \lim T Q \to \lim T Q' \]
(4)
commuting with the limiting cones. These formulas define a functor R: C\to A, and for each n\in M, the components \lambda_{1Kn} = \epsilon_n of the limiting cones define a natural transformation \varepsilon: RK\to T, and R\epsilon is a right Kan extension of T along K.

Proof. First, Rg is defined in (4) by the fact that the limit is a functor of (c\mid K) and hence of c. Specifically, given g: c\to c' and the projection Q': (c'\mid K)\to A, each \tilde{f}' : c'\to Kn determines \tilde{f} : c\to Kn as (c\mid K), the components \lambda_{\tilde{f}g}: Rc\to Tm form a cone from Rc, and since the cone \lambda' is universal, there is a unique arrow Rg which makes
\[ Rc' = \lim T Q' \to Tm \]
commute for all \tilde{f}'. (Actually, \tilde{f}' \to \tilde{f}g defines the functor (g\mid K) : (c'\mid K)\to (c\mid K), so that TQ' = TQ : (g\mid K), and Rg is the canonical comparison (cf. "final functors").) This choice of Rg clearly makes R a functor.

For each n\in M, 1Kn is an object of (Kn\mid K), so the limiting cone \lambda has a component \lambda_{1Kn}: RKn\to Tn, called \epsilon_n. For each h: n\to n' form the diagram
\[ RKn \quad \lambda_{1Kn} \rightarrow \quad \epsilon_n \quad \rightarrow \quad \lambda_{1Kn'} \rightarrow \quad RKn' \]
(6)
the lower triangle commutes by the definition of Rg for g = Kh, and the upper triangle commutes because \lambda is a cone. Therefore the square commutes; this states that \varepsilon: RK\to T is natural.

Now let S: C\to A be another functor, with \alpha: SK\to T natural. We construct \sigma: Sc\to Rc from the diagram
\[ Rc = \lim_f Tm \quad \tilde{\to} \quad Tm \quad \tilde{\to} \quad Tm' \]
\[ Sc \quad \tilde{\to} \quad SKm \quad \tilde{\to} \quad SKm' \]
(7)
for each f': c'\to Kn in (c'\mid K). The right-hand square and the outer square commute by the definition of \sigma, and the top box by the definition (5) of Rg. Therefore the left-hand (inner) square commutes after both legs are composed with \lambda_{f'} — and this for all f'. But \lambda' is a limiting cone, so the left-hand square commutes. Therefore \sigma is natural.

The definition (7) of \sigma for c = Kn, f = 1Kn, and m = n shows that \alpha_n = \lambda_{1Kn} = \epsilon_n, hence that \alpha = \varepsilon \cdot \sigma K. This proves that \varepsilon: RK\to T gives every \sigma as \sigma = \varepsilon \cdot \sigma K for some \sigma. The diagram (8) shows that \sigma is unique with this property. Indeed, this property determines the components \sigma_{Kn} of \sigma; to determine other components, set c = Kn, f = 1Kn, and m = n in (8). The lefthand square commutes if \sigma is natural, and then \lambda_{f'} = \sigma_n is determined for all g: c\to Kn. But \lambda is a limiting cone, so \sigma_n is determined. This shows that \varepsilon is universal, q.e.d.

Corollary 2. If M is small and A complete, any functor T: M\to A has a right Kan extension along any K: M\to C, and A^{K} has a right adjoint.

This applies in particular when A = Set; this is the case originally studied by Kan [1958].

Corollary 3. If the functor K in the theorem is full and faithful, then the universal arrow \varepsilon: RK\to T for the Kan extension R of T along K is a natural isomorphism \varepsilon: RK \cong T.

Proof. For n\in M, RKn is obtained from a limit over the comma category (c\mid K) with c = Kn. Because K is full and faithful, every object f: Kn\to Km in this comma category can be written as f = Kh for a unique h: n\to m. This states that 1: Kn\to Kn is an initial object in this comma category and hence that RKn = \lim_f TQ can be found by evaluating TQ just at this initial object: Thus RKn = Tn, \epsilon_n = 1, q.e.d.

This also gives a case in which a Kan extension is an actual extension:

Corollary 4. If M is a full subcategory of a category C and T: M\to A is a functor such that each composite (c\mid K)\to M\to A has a limit in A.

The Kan Extension 235
then there is a functor \( R : C \to A \) with \( RK = T \) (i.e., \( R \) extends \( T \)) such that the identity natural transformation \( 1 : RK \to T \) makes \( R \) the right Kan extension of \( T \) along the insertion \( K : M \to C \).

Proof. Apply Corollary 3 to the insertion \( M \to C \).

The left Kan extension \( L = \text{Lan}_K T \) is described similarly, as a pair \( L, \eta : T \to LK \) with \( \eta \) universal from \( T \) to \( A^K \); this gives a bijection

\[
\text{Nat}(\text{Lan}_K T, S) \cong \text{Nat}(T, SK)
\]

natural in \( S \in A^K \). When the requisite colimits exist, \( L \) is given by

\[
L \cong \text{Colim}\{(K \mid c) \to M \to A\},
\]

where \( P \) is the projection \( \langle m, km \to c \rangle \mapsto m \).

Exercises

Exercises 1–4 refer to the data for a Kan extension:

\[ K : M \to C, \quad T : M \to A. \]

1. If \( A \) is the arrow category, and \( M \) and \( C \) are sets, then a functor \( T : M \to 2 \) can be regarded as a subset of \( M \). Show that \( 2^M \) is the contravariant power set \( \mathcal{P} M \), that \( \text{Lan}_K T \) is the direct image of \( TCM \) under the function \( K \), and describe \( \text{Ran}_T \).

2. (Kan extensions of representable functors.) If \( A = \text{Set} \), and \( M, C \) have small hom-sets, show that the left Kan extension of \( M(m, -) \) is \( C(Km, -) \) with unit \( m; \eta : I_m \to C(Km, K -) \) given by \( \eta m = 1_{km} \).

3. If \( M, C \), and \( A \) are all sets, while \( A \) has at least two elements and \( K \) is not surjective, prove that neither \( \text{Lan}_K T \) nor \( \text{Ran}_T \) exists.

4. (Ulmer.) Show that Corollary 3 still holds if the hypothesis “\( K \) is full and faithful” is replaced by “\( K \) is full, and \( T \)” as “\( T \)”. Here \( K \) is as faithful as \( T \)” when, for arrows \( h, h' : m \to n \) in \( M \), \( Kh = Kh' \) implies \( Th = Th' \).

5. For any category \( M \), let \( M^{\text{op}} \) be the category formed by adding to \( M \) one new object \( \infty \), terminal in \( M^{\text{op}} \). For \( T : M \to A \), prove (from first principles) that a colimiting cone for \( T \) is a left Kan extension of \( T \) along the inclusion functor \( M \subset M^{\text{op}}, \) and conversely.

4. Kan Extensions as Coends

The calculus of coends gives an elegant formula for Kan extensions; for variety we treat the left Kan.

Theorem 1. Given functors \( K : M \to C \) and \( T : M \to A \) such that for all \( m, m' \in M \) and \( c \in C \) the copowers \( C(Km, c) \cdot Tm \) exist in \( A \), then \( T \) has a left Kan extension \( L = \text{Lan}_K T \) along \( K \) if for every \( c \in C \) the following coend exists, and when this is the case, the object functor of \( L \) is this coend

\[
Lc = (\text{Lan}_K T)c = \int_c C(Km, c) \cdot Tm, \quad c \in C.
\]

Proof. By the parameter theorem, we may regard this coend as a functor of \( c \). Compare it with any other functor \( S : C \to A \). Then

\[
A(Tm, SKm) \cong \text{Nat}(C(Km, -), A(Tm, -))
\]

by the Yoneda lemma and the representation of the set of natural transformations as an end (in a sufficiently large full category \( \text{Ens} \) of sets). Now we can write down in succession the following isomorphisms

\[
\text{Nat}(L, S) \cong \int_c \sum A(Lc, Sc) \quad \text{(end formula for Nat)}
\]

\[
\cong \int_c \sum A(c(Km, c) \cdot Tm, Sc) \quad \text{(Definition of } L)\]

\[
\cong \int_c \sum A(c(Km, c) \cdot Tm, Sc) \quad \text{(Continuity of } A(-, Sc))
\]

\[
\cong \int_c \sum \text{Ens}(C(Km, c), A(Tm, Sc)) \quad \text{(Definition of copowers)}
\]

\[
\cong \int_c \sum \text{Ens}(C(Km, c), A(Tm, Sc)) \quad \text{(Fubini)}
\]

\[
\cong \int_c \sum A(Tm, SKm) \quad \text{(by (2) above)}
\]

\[
\cong \text{Nat}(T, SK) \quad \text{(end formula for Nat)}.
\]

Here the Fubini theorem (interchange of ends) applies because both indicated ends \( \sum \) and \( \int \) exist, while \( \text{Ens} \) must be a sufficiently large category of sets (to contain all hom-sets for \( A \) and \( C \) and all sets \( \text{Nat}(L, S), \text{Nat}(T, SK) \) for all \( S : C \to A \)). Since each step is natural in \( S \), the composite isomorphism is natural in \( S \) and proves that \( L = \text{Lan}_K T \).

Note that we do not assert the converse: That if \( \text{Lan}_K T \) exists, it must be given for each \( c \) as the coend \( (1) \).

The unit \( \eta \) of this Kan extension is obtained by setting \( S = L \) and following the chain of isomorphisms. We record the result:

Theorem 2. (Kan extensions as coends, continued.) For the Kan extension \( (1) \) above the universal arrow \( \eta : T \to LK \) is given for each \( m \in M \) as the composite of an injection \( i_m \) of the copower \( (f = 1_K : Kn \to Kn) \) with a component of the ending wedge \( \omega \):

\[
Tm \xrightarrow{i_m} C(Kn, Kn) \cdot Tn \xrightarrow{\omega_m, \cdot} C(Km, Kn) \cdot Tm = (\text{Lan}_K T)(Kn).
\]
For the left Kan extension we thus have two formulas – (1) above by coends, and (3.10) by colimits. They are closely related, and simply constitute two ways of organizing the same colimit information (see Exercise 1 below). The corollaries of §3 can be deduced from either formula. Also right Kan extensions are given by a formula

\[(\text{Ran}_K T)c = \int_n Tm^{c_i}(c_n)\]

valid when the indicated power (powers \(X, a \mapsto a^X\) in \(A\)) and its end exist.

Consider additive Kan extensions: \(M, C\), and \(A\) are Ab-categories and the given functors \(K\) and \(T\) are both additive. Then we can describe a right Ab-Kan extension of \(T\) along \(K\) as an additive functor \(R: C \to A\) with a bijection (3.2) given and natural for additive functors \(S\). This functor \(R\) need not agree with the ordinary right Kan extensions \(\text{Ran}_K T\) obtained by forgetting that \(K, T\) (and \(S\)) are additive. However, \(R\) can still be given by a formula (3) with an end, provided the power \(a^c\) involved (for \(a \in A, C \in \text{Ab}\)) is replaced by a “cotensor” \(a^c\) defined by the adjunction

\[A(b, c^c) \cong \text{Ab}(C, A(b, c))\]

for all \(b \in A\) (see Day and Kelly [1969], Dubuc [1970]). For example, if \(A = \text{R-Mod}\), this makes \(a^c = \text{Ab}(C, a)\) with the evident module structure (induced from that of \(a \in \text{R-Mod}\)).

Derived functors are an example. If \(T: \text{R-Mod} \to \text{Ab}\) is right exact, its left derived functors \(T_\ast: \text{R-Mod} \to \text{Ab}\) come equipped with certain connecting morphisms, which make them what is called a connected sequence of functors (Mac Lane [1963a], Cartan-Eilenberg [1956]): basic example: If \(A\) is a right \(R\)-module, the left-derived functors of the tensor product \(A \otimes_R K =: \text{R-Mod} \to \text{Ab}\) are the torsion products \(\text{Tors}_m(A, -): \text{R-Mod} \to \text{Ab}\).

The left-derived functors \(T_\ast\) of \(T\) can be described by the following “universal” property: \(T_0 = T\), and if \(S_\ast\) is any connected sequence of (additive) functors, each natural transformation \(S_\ast \Rightarrow T_0\) extends to a unique morphism \(\{S_\ast | n \geq 0\} \to \{T_\ast | n \geq 0\}\) of connected sequences of functors.

This property may be rewritten thus. Embed \(\text{R-Mod}\) in a larger \(\text{Ab}\)-category \(E\) with objects \(\langle C, n \rangle\), \(C\) an \(R\)-module and \(n\) a nonnegative integer, while the hom-groups are \(E(\langle C, n \rangle, \langle B, m \rangle) = \text{Ext}^{n m}_R(C, B)\), with composites given by the Yoneda product. Then \(C(\langle C, 0 \rangle)\) is a functor \(K: \text{R-Mod} \to E\). A connected sequence of additive functors \(\{T_\ast | n \geq 0\}\) is then the same thing as a single additive functor \(T_\ast: E \to \text{Ab}\) with \(T_\ast(C, n) = T_\ast(C)\), while \(T_\ast\) on the morphisms of \(E\) gives the connecting morphisms. The universal property stated above for the sequence \(T_\ast\) of left derived functors of \(T\) now reads:

\[\text{Nat}(S_\ast K, T) \cong \text{Nat}(S_\ast, T_\ast)\].

Pointwise Kan Extensions

This states exactly that \(T_\ast\) is the right \(\text{Ab-Kan}\) extension of \(T = T_0\) along \(K: \text{R-Mod} \to \text{E}\) (and that the unit \(e: T_\ast K \to T\) of this Kan extension is the identity).

For details we refer to Cartan-Eilenberg or to Mac Lane [1963a] (where the category \(E\) is treated in a different but equivalent way, as a “graded \(\text{Ab}\)-category”).

Exercises

1. If the coends in Theorem 1 exist, prove that these coends do give the colimits required in the formula (3.10) for \(\text{Ran}_K\).
2. For fixed \(K\), describe \(\text{Ran}_K T\) and \(\text{Ran}_K T\), when they exist, as functors of \(T\).
3. (Dubuc) If \(\text{Ran}_K T\) exists, while \(L: C \to D\) is any functor, prove that \(\text{Ran}_{LK} T\) exists if and only if \(\text{Ran}_K \text{Ran}_{LT} T\) exists and that then these two functors (and their universal arrows) are equal.
4. (Ulmer; Day-Kelly; Kan extensions as a coend in a functor category \(A^C\))
   If \(C(Km', c) \cdot Tm\) exists for all \(m', c \in M\) and all \(c \in C\), show that \(\langle m', c \rangle \cdot C(Km', c) - Tm\) is the object function of a functor \(M^\ast \times M^\ast \to A^C\). Prove that \(T\) has a left Kan extension along \(K\) if and only if this bifunctor has a coend, and that then this coend is the \(\text{Kan}\) extension

\[\text{Ran}_K T = \int C(Km, c) \cdot Tm\]

Describe the universal arrow for \(\text{Ran}_K T\) in terms of the coend.
5. (Ulmer). As in Ex. 4, obtain a necessary and sufficient condition for the existence of \(\text{Ran}_K T\) in terms of the limit formula, interpreted in the functor category \(A^C\).

5. Pointwise Kan Extensions

Given functors

\[C \leftarrow K\]

and a right Kan extension \(\text{Ran}_K T\) with counit \(e: (\text{Ran}_K T) \Rightarrow T\), we say that \(G\) preserves this right Kan extension when \(G \circ \text{Ran}_K T\) is a right Kan extension of \(GT\) along \(K\) with counit \(G\hat{e}: G(\text{Ran}_K T) \Rightarrow GT\). This implies (but is stronger than)

\[G \circ \text{Ran}_K T \cong \text{Ran}_K (GT)\]

We already know that right adjoints \(G\) preserve limits. We now show that they also preserve Kan extensions.

Theorem 1. If \(G: A \Rightarrow X\) has a left adjoint \(F\), it preserves all right Kan extensions which exist in \(A\).

Proof. First a preliminary, for an adjunction

\[A(Fx, a) \cong X(x, Ga), \quad x \in X, \ a \in A\].
Kan Extensions

If in place of $x$ we have a functor $H : C \to X$ and in place of $a$ a functor $L : C \to A$, then applying this adjunction at every $Lc$ and $Hc$ gives a bijection,

$$\text{Nat}(FH, L) \cong \text{Nat}(H, GL).$$

(2)

(As usual the adjunction switches $F$ on the left to $G$ on the right.)

Now assume the adjunction and a right Kan extension $\text{Ran}_K T$ for some $K$ and $T : M \to A$. Then for any functor $H : C \to X$ we have the following bijections

$$\text{Nat}(H, G \circ \text{Ran}_K T) \cong \text{Nat}(FH, \text{Ran}_K T) \cong \text{Nat}(HK, GT),$$

natural in $H$; the first and third are instances of (2), and the second is the definition of the right Kan extension. The composite bijection (for all $H$) shows that $G = \text{Ran}_K T$ is the right Kan extension $\text{Ran}_K G T$. To get its counit, we set $H = G \circ \text{Ran}_K T$ and take the image of the identity; we get $G$, where $\varepsilon : (\text{Ran}_K T)K \to T$ is the counit of the given Kan extension.

**Corollary 2.** If $R, e : RK \to T$ is a right Kan extension and $A$ has small hom-sets and all small copowers, then for each $a \in A$, $A(a, K -) : C \to \text{Set}$, is the right Kan extension of $A(a, T -) : M \to \text{Set}$, with counit $A(a, -)$.

**Proof.** The functor $A(a, -) : A \to \text{Set}$ has the left adjoint $X \to X \cdot a$, the copower.

**Definition.** Given $C \to M \to A$, where $A$ has small hom-sets, a right Kan extension $R$ is point-wise when it is preserved by all representable functors $A(a, -) : A \to \text{Set}$, for $a \in A$.

**Theorem 3.** A functor $T : M \to A$ has a pointwise right Kan extension along $K : M \to C$ if and only if the limit of $(c \downarrow K) \to M \to A$ exists for all $c$. When this is the case, $\text{Ran}_K T$ is given by the formulas of Theorem 3.1.

**Proof.** Since $A(a, -)$ preserves limits, any Kan extension given by the limit formula is pointwise.

Conversely, suppose for each $a \in A$ that $A(a, T -) : M \to \text{Set}$ has a right Kan extension $R^a = A(a, R -)$, as in the figure

```
  C
  | v
  | k
  M --\- A(a, -) \to \text{Set}
    | \n    | \n    \--\- R^a
```

Then for each functor $V$, as shown, there is a bijection

$$\text{Nat}(V, R^a) \cong \text{Nat}(VK, A(a, T -)),$$

natural in $V$. This holds in particular when $V = C(c, -)$ for some $c \in C$, so

$$\text{Nat}(C(c, -), A(a, R -)) \cong \text{Nat}(C(c, -), A(a, T -)).$$

Density

We reduce the left hand side by the Yoneda Lemma and the right hand side by the lemma below to get

$$A(a, R c) \cong \text{Cone}(a, TQ : c \downarrow K) \to A).$$

This states that the set of cones is representable, hence that the limit of $TQ$ exists, q.e.d.

The missing lemma is

**Lemma.** Given $K : M \to C$, there is a bijection

$$\text{Cone}(a, (c \downarrow K) \to M \to C) \cong \text{Nat}(C(c, K -), A(a, T -)).$$

**Proof.** A cone $\tau : a \to TQ$ assigns to each $f : c \to Km$ an arrow $\tau(f, m) : a \to Tm$ subject to the cone conditions; for each $h : m \to m'$,

$$\tau(Kh \circ f, m') = Th \circ \tau(f, m).$$

A natural transformation $\beta : C(c, K -) \to A(a, T -)$ assigns to each $m \in M$ and to each $f : c \to Km$ an arrow $\beta_m f : a \to Tm$, subject to the naturality condition, for each $h : m' \to m$, that

$$\beta_m(Kh \circ f) = Th \circ \beta_m f.$$

The bijection $\tau \mapsto \beta$ is now evident.

This proof of the theorem also shows

**Corollary 4.** $R, e : RK \to T$ is a pointwise Kan extension of $T$ along $K$ if and only if, for all $a \in A$ and $c \in C$,

$$A(a, R c) \to \text{Nat}(C(c, K -), A(a, T -))$$

sending $g : a \to R c$ to the transformation with the component

$$C(c, Km)^{-} \to A(Rc, R K m)^{-} G \circ \text{Ran}_K a, A(a, T m)$$

at $m \in M$ is a bijection.

**Exercises.**

1. In the situation (1), if $\text{Ran}_K T$ and $\text{Ran}_K G T$ both exist, with counits $\varepsilon$ and $\epsilon'$, prove that there is a unique natural transformation (the canonical map)

$$w : G = \text{Ran}_K T \to \text{Ran}_K G T$$

with $\varepsilon' \cdot w = G \varepsilon$, and prove that $G$ preserves $\text{Ran}_K T$ if and only if $w$ is an isomorphism.

6. Density

A subcategory $M$ of $C$ is said to be dense in $C$ if every object of $C$ is a colimit of objects of $M$; more exactly, a colimit in a canonical way, for which the colimiting cone consists of all arrows $m \to c$ to $c$ from an $m \in M$. More generally, density can be defined not only for an inclusion
$M \subset C$, but for any functor $K : M \to C$. The arrows $m \to c$ are then replaced by the objects $(m, f : K m \to c)$ of the comma category $(K \downarrow c)$. Recall that the projections $P^c, Q^c$ of this comma category are given by $P^c(m, f) = m, Q^c(m, f) = f$, and observe that the object function of $Q^c$ may also be regarded as a cone $Q^c : K P^c \to c$.

**Definition.** A functor $K : M \to C$ is dense when for each $c \in C$

$\text{Colim (} c \downarrow K^{-1} \text{)} = P^c \longrightarrow \text{M} \longrightarrow K^{-1} \longrightarrow c,$

(1)

with colimiting cone the "canonical cone" $Q^c$. In particular, a subcategory $M$ of $C$ is dense in $C$ when the inclusion functor $M \to C$ is dense in the sense just defined.

The definition (1) is sometimes phrased, "The canonical map $\text{Colim } K P^c \to c$ is an isomorphism"; here the canonical map is the unique arrow $k : \text{Colim } K P^c \to c$ which carries the colimiting cone to $Q^c$.

For example, the one-point set * is dense in Set: For each set $X$, the comma category $(A \times X)$ is just the set (discrete category) of elements $x \in X$, each regarded as a function $x : * \to X$, while (1) becomes statement that each $X$ is the coproduct $\Pi X$ of its elements (i.e., that a function $f$ with domain $X$ can be uniquely determined by specifying the value $f(x)$ at each $x \in X$).

Dually, a functor $K : M \to C$ is codense when for each $c \in C$

$\text{Lim (} c \downarrow K \text{)} \longrightarrow K \longrightarrow \text{M} \longrightarrow \text{Lim (} c \downarrow K \text{)} = Q^c,$

(2)

with limiting cone the canonical cone sending $(f : c \to K m, m)$ to $f$. But this limit is precisely the one involved in the definition of $\text{Ran}_K K$.

**Proposition 1.** The functor $K : M \to C$ is codense if and only if $1_{dc}$, together with the identity natural transformation $1_{dc} : K \to K$, is the pointwise right Kan extension of $K$ along $K$.

In this case Corollary 5.4 simplifies (e is the identity) to the correspondence sending each $f : a \to c$ to the natural transformation $f^* = C(f, K-)$:

$C(c, K-) \longrightarrow C(a, K-) \longrightarrow C(f, K-)$

(3)

(the transformation $f^*$ is "composition with $f$ on the right". Hence

**Proposition 2.** The functor $K : M \to C$ is codense if and only if the correspondence $f \to C(f, K-)$ above is for all $a$ and $c \in C$, a bijection

$C(a, c) \cong \text{Nat}(C(c, K-), C(a, K-))$;

(4)

that is, if and only if the functor $C^P \to \text{Ens}^M$ defined by

$c \mapsto C(c, K-) : M \to \text{Ens}$

is full and faithful.

**Corollary 3.** If the hom-sets of $M$ lie in a full category $\text{Ens}$ of sets, then Yoneda embedding $Y : M \to (\text{Ens}^{\text{Mpp}})$, given by $Ym = M(m, -)$ is codense.

**Proof.** By the Yoneda Lemma itself, for each $F : M \to \text{Ens}$,

$(\text{Ens}^{\text{Mpp}})(F, Ym) = \text{Ens}^M(Ym, F) \cong Fm.$

Thus the right side of (4) above, with $C = (\text{Ens}^{\text{Mpp}}), a = F$ and $c = G$ becomes

$\text{Nat}(G, F) = (\text{Ens}^{\text{Mpp}})(F, G) = C(F, G),$

and (4) becomes an identity.

This result is often stated thus: Any functor $M \to \text{Ens}$ is a canonical limit of representable functors.

The dual of Proposition 2 states that $K : M \to C$ is dense if and only if $c \mapsto C(K- - c)$ is a full and faithful functor $C \to \text{Ens}^{\text{Mpp}}$. As an application, we show that the full subcategory of finitely generated abelian groups is dense in $\text{Ab}$. We need only show that for any two abelian groups $A$ and $B$ the map

$\text{Ab}(A, B) \to \text{Nat}(\text{Ab}(K- - A), \text{Ab}(K- - B))$

is a bijection. First, it is injective: Two homomorphisms $f, g : A \to B$ which agree on cyclic subgroups of $A$ must agree everywhere. Also, it is surjective: Given $\tau : \text{Ab}(K- - A) \to \text{Ab}(K- - B)$, we define a function $a : A \to B$ by taking $fa$ for each $a \in A$ to be the value of $\tau$ on the map $Z \to A$ taking 1 to $a$. Because $Z \oplus Z$ is a finitely generated group, this function must be a homomorphism. Its image under the map in question agrees with $\tau$; the proof is complete. Note that the argument proves more: The full subcategory with one object $Z \oplus Z$ is dense in $\text{Ab}$. (There are two summands $Z$ required because abelian groups are algebraic systems defined by binary operations.)

**Exercises**

1. In $R^{-}\text{Mod}$, show that the full subcategory with one object $R \oplus R$ is dense.
2. Show that the full subcategory with one object $Z$ is not dense in $\text{Ab}$.
3. Show (Urysohn Lemma) that the closed unit interval is a cogenerator in the category of all normal topological spaces (a full subcategory of $\text{Top}$). Is it codense?
4. Let $l$ be the image category $KM$ for $K : M \to C$ be the subcategory of $C$ with objects all $K m$ for $m \in M$ and arrows all $K k, h$ in $M$. Prove that $K$ dense implies $KM$ a dense subcategory of $C$.
5. Prove that the objects of a subcategory $M$ generate $C$ if and only if the functor $C \to \text{Ens}^{\text{Mpp}}$ given by $c \mapsto C(K- - c)$, is faithful.
6. If all copowers $\text{C}(Km, c) : Km \text{ exist in } C$, prove that $K : M \to C$ is dense if and only if each object $c \in C$ is the coend

$c = \int C(Km, c) \cdot Km$

with coending wedge $\omega ^c_{Km} : C(Km, c) \cdot Km \to c$ given on the injections $i_x$ of the copower as $\omega ^c_{Km} i_x = f : Km \to c$. 

7. All Concepts are Kan Extensions

The notion of Kan extensions subsumes all the other fundamental concepts of category theory.

**Theorem 1.** A functor $T : M \to A$ has a colimit if and only if it has a left Kan extension along the (unique) functor $K_c : M \to 1$, and then $\text{Colim} T$ is the value of $\text{Lan}_{K_c} T$ on the unique object of $1$.

**Proof.** A functor $S : 1 \to A$ is just an object $a \in A$, and a natural transformation $\tau : T \to SK_a$, for $K_c : M \to 1$, is just a cone with base $T$ and vertex $a$. Since the left Kan extension $L = \text{Lan}_{K_c} T$ is constructed to provide the universal natural $\eta : T \to LK_c$, it also provides the universal cone with base $T$, and hence the colimit of $T$.

Dually, right Kan extensions along the same functor $K_c$ give limits.

**Theorem 2.** (Formal criteria for the existence of an adjoint.) A functor $G : A \to X$ has a left adjoint if and only if the right Kan extension $\text{Ran}_G 1_A : X \to A$ exists and is preserved by $G$; when this is the case, this right Kan extension is a left adjoint $F = \text{Ran}_G 1_A$ for $G$, and the counit transformation $\varepsilon : (\text{Ran}_G 1_A)G \to 1_A$ for the Kan extension is the counit $\varepsilon : FG \to 1$ of the adjunction.

**Proof.** If $G$ has a left adjoint $F$, with unit $\eta : 1_A \to GF$ and counit $\varepsilon : FG \to 1_A$, then we can construct for all functors $H : A \to C$ (in particular, for the identity functor $1_A$) a bijection

$$\text{Nat}(S, HF) \cong \text{Nat}(SG, H),$$

natural in $S : X \to C$, by the assignments

$$(\sigma : S \to HF) \mapsto (SG \xrightarrow{\sigma G} HFG \xrightarrow{HF} H),$$

$$(\tau : SG \to H) \mapsto (S \xrightarrow{S\eta} SG \xrightarrow{\tau} HF).$$

The first followed by the second is the identity $\sigma \mapsto \sigma$, because the diagram

$$\begin{array}{c}
S \\
\downarrow S\eta
\end{array}$$

$$\begin{array}{c}
HF \\
\downarrow HF\eta
\end{array}$$

$$\begin{array}{c}
SGF \xrightarrow{\sigma GF} HFGF \xrightarrow{HF} HF
\end{array}$$

is commutative (the first square represents the horizontal composite $\sigma \eta$ in two ways, and the second square is $H$ applied to one of the two triangular identities for $\eta$ and $\varepsilon$). The composite in the other order is also an identity, by a similar diagram. Hence we have the asserted bijection, clearly natural in $S$. If we take $H = 1_A$, this bijection shows that $F = \text{Ran}_G 1_A$, its unit is the image of $\sigma = 1_\varepsilon$, so is $\varepsilon$. If we take $H = G$,

this bijection shows that $GF = \text{Ran}_G G$, with unit $G\varepsilon$. Hence $G$ preserves the right Kan extension $\text{Ran}_G 1_A$. We have proved the first half of the theorem.

We have proved more: For any $H, HF = \text{Ran}_G H$, with unit $H\varepsilon$. Thus $\text{Ran}_G 1_A$ is preserved by any functor whatever (it is an absolute Kan extension).

**Proposition 3.** If $G : A \to X$ has a left adjoint $F$ with counit $\varepsilon : FG \to 1$,

then $\text{Ran}_G 1_A$ exists, is equal to $F$ with counit $\varepsilon$, and is preserved by any functor whatever.

Now suppose conversely that $1_A$ has a right Kan extension $R$ along $G$. Then we have bijections

$$\varphi = \varphi_\varepsilon : \text{Nat}(S, R) \cong \text{Nat}(SG, H), \quad \varphi(S \xrightarrow{\varepsilon} R) = \varepsilon \cdot qG,$$

$$\psi = \psi_\eta : \text{Nat}(H, GR) \cong \text{Nat}(HG, G), \quad \psi(H \xrightarrow{\eta} GR) = G\varepsilon \cdot \sigma G,$$

natural in $S : X \to A$ and $H : X \to X$, with counit $\varphi_\eta 1 = \varepsilon : RG \to 1_A$ and $\psi_\eta 1 = G\varepsilon : GFR \to 1_G$. Define $\eta : 1 \to GR$ to be $\psi_\eta^{-1}(1 : G \to G)$. Then

$$G\varepsilon \cdot \eta \eta = 1_G.$$

This is one of the two triangular identities for the proposed adjunction $\varepsilon : RG \to 1_A; \eta : 1 \to GR$. The other would be $\varepsilon \cdot R\eta = 1_\varepsilon$. Applying the bijection $\varphi_\eta$, it will suffice to prove instead $\varphi(\varepsilon \cdot R\eta) = \varepsilon$. Putting in the definition of $\varphi$ in terms of $\varepsilon$, we are to prove the following square commutative:

$$\begin{array}{c}
RG \xrightarrow{\varepsilon} RFR \\
\downarrow R\eta \\
1_A \xleftarrow{1} GR
\end{array}$$

Insert the dotted arrow at the top and use $R$ of the (known) triangular identity $G\varepsilon \cdot \eta \eta = 1$. The square then reduces to the equivalence of two expressions for $\varepsilon$:

$$\varepsilon = G\varepsilon \cdot R\eta = 1_\varepsilon.$$ q.e.d.

The arguments so far in this section have not used either formula for Kan extensions. We now examine the meaning of these formulas in the simple case of Kan extensions along the identity functor $1 : C \to C$.

The universal property defining Kan extensions shows at once for each $T : C \to A$ that

$$\text{Lan}_T T = T.$$ 

Consider in particular $T : C \to \text{Set}$, and assume that $C$ has small homsets. Then, in the formula for $\text{Lan}_T$ as an end, all the powers involved exist, so for every $c \in C$

$$Tc = (\text{Lan}_T T)c = \int_m Tm^c(c, m).$$
But in $\text{Set}$, $X^Y = \text{Set}(Y, X)$, and by (IX.5.2) the end reduces to a set of natural transformations

$$Tc = \int \text{Set}(C(c, m), Tm) \cong \text{Nat}(C(c, -), T).$$

The result is just the Yoneda Lemma.

**Exercises**

1. Show that the bijection (1) (and (5.2) as well) is a special case of a bijection defined for an adjoint square (Exercise IV.7.4)

   $$\text{Nat}(HG, GK) \cong \text{Nat}(FH, KG).$$

2. Obtain the Yoneda Lemma from the limit formula for $\text{Ran}_p T$. (This gives an independent proof of the Yoneda Lemma, which was not used in the proof of §3).

3. (a) If $K : M \rightarrow C$ has a right Kan extension $R$, along itself, $\varphi : \text{Nat}(S, R) \cong \text{Nat}(SK, K)$, prove that $\langle R, \eta, \mu \rangle$ is a monad in $C$, where $\eta = \varphi^{-1}(\text{Id}_K), \mu = \varphi^{-1}(e \cdot R)$. (This is called the codensity monad of $K$).

(b) Show that $K$ is codense if and only if $\eta$ is an isomorphism.

(c) If $G : A \rightarrow C$ has a left adjoint $F : X \rightarrow A$ with unit $\eta : \text{Id} \rightarrow GF$ and counit $e : FG \rightarrow \text{Id}$, then its codensity monad exists and is $\langle GF, \eta, GeF \rangle$. (The monad defined by the adjunction).

**Notes**

The formal criteria for adjoints are due to Bénabou [1965]. Kan extensions by limits and colimits, in the critical case when the receiving category $A$ is $\text{Set}$, was achieved by Kan in [1960]. The impact of this construction was not realized until 1963, when Lawvere introduced these extensions in functorial semantics. Ulmer emphasized their importance, and in an unpublished paper gave the coend formula (without the name coend) for $\text{Lan}_p T$. Bénabou (unpublished) and Day-Kelly [1969] describe Kan extensions in relative categories (including Ab-categories). This idea is further developed by Dubuc [1970]; here the coend formula for Kan extensions plays a central role.

The Cartan-Eilenberg notion of derived functors is, as noted in §4, the original and decisive example of a Kan extension. Verdier, by embedding each abelian category in a suitable derived category, has achieved an elegant form of this interpretation of derived functors by Kan extensions. For an exposition, see Quillen [1967].

Isbell, in a pioneering paper [1960], defined a functor $K : M \rightarrow C$ to be "left adequate" when $c \mapsto C(K(-, c)$ is full and faithful. This assignment is the functor of the dual of Proposition 6.2; hence by that theorem "left adequate" and "dense" agree. Isbell has developed the ideas further in characterizing categories of algebras [1964].

The ubiquity of Kan extensions has developed gradually; I have learned much in this chapter from my student Eduardo Dubuc; and Max Kelly has suggested major improvements, notably the use of pointwise Kan extensions.
Bibliography


Index

Ab-category 17, 24, 29, 190, 194
Abelian
— categories 194 ff.
— groups 24
Absolute
— coequalizer 145
— Kan extension 245
— limit 145
Action
— group 137
— left of a monoid 170
— of operators 120
Addition
— ordinal 171
— of arrows 192
Additive
— category 192
— functor 28, 83, 193, 238
— of operators 238
Adjoint
— 2, 85
Freyd’s — functor theorem 117
Left 2, 38, 70, 85, 230
Left — left-inverse 92
Right 2, 79, 82, 230
— equivalence 91
— functor 38, 79
— pairs 93
— square 101 ex.
Adjunction 2
Adjunct 79
Adjunction 78, 80, 81
— category of — 99
— count of — 81
— front and back — 81
— map of — 97
— monad defined by — 135
— unit of — 81
— with a parameter 100
Algebraic system 120

Algebras
— morphisms of $T$ — 136
— structure map of $T$ — 137
$T$ — 136
— Variety of — 120
— Amalgamated product 66
— Arity 120
Arrows
— addition of — 194 ex.
— canonical — 74, 101 ex., 165
— category of — 40
— composable pair of — 9, 10, 49, 196
— connecting — 203
diagonal — 84 ex., 189
epi — 19, 190
— factorization of — 195
— idempotent — 20
— identity — 7, 8, 12, 190
— invertible — 19
— kernel of — 187
— monic — 19, 190
— regular — 21
— universal — 8, 55, 58, 231 ex.
— weak universal — 231
— zero — 20, 74, 187, 190
— function 13
— only-meta-category 9
— Associative law 8
— general — 167, 168, 171
— for $T$-algebra 136
— for monad 134
— for monoid 159
— for monoidal categories 162
— Associativity 7
— Atomic statement 31
— Augmentation 46, 175
— Augmented simplicial object 175
— Barycentric coordinates 174
— Based category 180
— Base point 26, 184

Basic
— arrows 162
— graphs 162
— Beck’s theorem 147
— Bifunctor 37, 191, 210
— Bilinear composition 28
— Binary
— relation 26
— words 161
— Biproduct 190 ff.
— Bound
— greatest lower — 110, 122
— least upper — 110, 122
— variable 31, 219
— Boundary
— homomorphism 175
— of tetrahedron 174

Cancellable (left or right) 19
Canonical
— arrow 74, 110 ex., 167, 189, 211
— map 165, 167
— presentation 149
— Cartesian
— closed category 95
— product 2, 21
— square 71
— Category 7, 10, 27
— $Ab$ — 28
— abelian — 194
— additive — 192
— based — 180
— cartesian closed — 95
closed — 180
— comma — 47
— complete — 105, 106
— concrete — 26
— connected — 86
— co-well-powered — 126
discrete — 11
double — 44
dual — 33
empty — 229
— enriched — 181
— equivalence of — 20, 91
— filtered — 33
— filtered — 207
— free — 50
— functor — 40, 44, 115
— image — 243 ex.
— isomorphism of — 90
— large — 12, 23
— locally small — 127

Category
— monoidal — 157, 158, 167, 176
— free — 166 ex.
— strict — 157, 160, 171
— symmetric — 180
— opposite — 33
— preadditive — 28
— product — 36
— pseudo-filtered — 212
— relative — 180
— simplicial — 12, 171
— subgroup — 220
— super-comma — 111 ex.
two-dimensional — 44, 102
— well-powered — 126
— of adjunctions 99
— of algebras 124
— of arrows 40
— of diagrams 53
— of small sets 12, 24

Chain complex 177, 198
Character group 17
Chase, diagram 75, 200
Class 23
— equation 120
— Closed
cartesian — categories 95
— of arrows 180
Closure operation 135, 137, 153
Cochain complex 179
Cocomponents of a map 74
Codense functor 242
Codensity monad 246
Codomain 7, 10
Coend 212, 236
Coequalizers 64
— absolute — 145
— creation of — 147, 152
— split — 146
Cogenzerating set 123
Cogenzerator 123
Coherence theorem 157, 161
Coimage 196
Cokernel 64
— pair 66
Colimit 67, 208, 210, 229, 236
— filtered — 208
— reflection of — 150
— Commutative diagram 4, 8
— }
Index

Commomrator 17
— subgroup 14
Compact Hausdorff space 153
Compactification 121, 127
Compactly generated spaces 181
Compact-open topology 181
Comparison
— functor 139, 150, 152
— theorem 144
Complete category 105
Component 69, 214
connected — 88 ex.
matrix of — 192
— of natural transformation 16
— of wedge 219
Composable pair 9, 10, 13, 49
Composite 7, 14
horizontal — 42, 102, 133
vertical — 42, 99, 102
— function 9
— function 14, 43
— of paths 20
— of transformations 40, 102
Composition 7, 27
Comprehension principle 21
Concrete category 26
Cone 67, 71, 106
canonical — 242
colimiting — 67, 210
limiting — 67, 69, 211
universal — 67
Congruence 52
Conjugate natural transformation 99, 102
Conjugation 18 ex., 20
Connected
— category 86
— component 88 ex.
— groupoid 20
— sequence of functors 238
Connecting homomorphism 202, 238
Connection, Galois 93, 94
Continuous
— functor 112
— map 153, 181
Contractible parallel pair 146 ex.
Contravariant
— functor 17, 33
— hom-functor 34
Coordinates, barycentric 174
Copowers 64, 237
Coproduct 62
denumerable — 168
Coproduct
finite — 208
infinite — 64, 208
injections of — 63
— diagram 62
— object 63
Coreflective 89, 181
Count of adjunction 81, 85
Covariant
— functor 34
— hom-functor 34
— power-set functor 138 ex.
Co-well-powered category 126
Creation
— of coequalizers 147, 152
— of ends 221
— of limits 108
CTT-Crude tripleability theorem 151 ex.
Degeneracy 175
Dense
— functor 242
— subcategory 241
Derived
— functor 238
— operator 120
Determinant 16
Diagonal
— arrow 84, 189
— functor 67
— map 73, 192
Diagram 2, 3, 4, 5, 71, 196
biproduct — 190
category of — 53
commutative — 3, 8, 161
coproduct — 62
limit — 69, 71
product — 69
— chase 75, 200
— scheme 10, 49, 51
Difference
— kernel 70
— member 204
Dinatural transformation 214
Direct
— product 69
— sum 191
Directed
— preorder 207
— set 207
Disjoint
— hom-sets 27
— union 63
Domain 7, 10
Double
— category 44
— end 226
Dual 31
— category 33
Duality principle 32
Dummy 215
Eilenberg-Moore category of a monad 135
Element, universal 57
Embedding 15
Empty
— category 10, 229
— functor 229
End 218
creation of — 221
double — 226
interchange of — 237
preservation of — 221
— of natural transformation 224
Ending wedge 219
Endofunctor 133
Enriched category 181
Epi 19
split — 20
— monic factorization 190, 195
Equalizer 70, 109, 129
Equational class 120
Equivalence
— adjoint — 91
— natural — 16
— of categories 18, 91
ETAC 31
Euclidean vector spaces 216
Evaluation map 96, 216
Exact 196
left — functor 197
right — sequence 197
short — sequence 196
— functor 197
Extensions 229
absolute Kan — 245
Kan — 229
left — 238
right — 232, 233
Extranatural transformation 215
Face operator 175
Factorization of arrows 190, 195
Faithful functor 15
Fiber map 71
Fibered
— category 35
— product 71
— sum 66
Fields of quotients 56
Filtered
— category 207
— colimit 208
— set 207
Final
— functor 213, 234
— subcategory 213
Finite
— limit 109
— product 72
Five Lemma 198, 201
Forgetful functor 14, 85, 116, 140,
153, 170, 208
Fork, split 145
Formal criteria
— for existence of adjoint 230, 244
— for representability 231
— for a universal arrow 231
Free
— category 50
— monoid 51
— monoidal category 162
— product 124 ex., 210
— T-algebra 136
Freyd
— adjoint functor theorem 117
— existence theorem for an initial
— object 116, 231
Full functor 14
Function 8
arrow — 13
composite — 9
identity — 9
inclusion — 9
insertion — 9
monotone — 172
object — 13
order preserving — 94
— set 40
Functor 3, 13
additive — 29, 83, 193, 238
adjoint — 79
codense — 242
comparison — 138
composite — 14
continuous — 112
contravariant — 17, 35
covariant — 34, 138
Index

Functor

dense — 242
derived — 238
diagonal — 67
empty — 229
exact — 197
faithful — 15
final — 213, 234
forgetful — 14, 85, 116, 140, 153, 170, 208
full — 14
identity — 14, 165
inclusion — 15
left adequate — 246
left exact — 197
morphism of — 16
power-set — 13, 33
representable — 61
underlying — 14
Yoneda — 61
— category 40

Fundamental groupoid 20

Galois connections 93 ff.
General linear group 14
Generating
— object 123
— set 121
Generators of a category 52
Geometric realization 223
Gödel-Bernays axioms 23
Graded set 120
Graph 10, 48, 75
Group 11, 20, 75
fundamental — 20
small — 22
— actions 137
— in a category 3
Groupoid 20

Hausdorff spaces 121, 131
compact — 121, 153
compactly generated — 181
Hom-functor 35, 193
contravariant — 34
covariant — 34
Hom-object 180
Hom-sets 10, 27, 180
disjoint — 27
Homology 175, 198
singular — 175
Homomorphism (see also morphism)
boundary — 175
connecting — 202

Homomorphism (see also morphism)
crossed — 179
Homotopy class
— of maps 12, 13, 25
— of paths 20
Horizontal composite 42, 102, 133
Idempotent 20
split — 20
Identities (for algebras) 120
Identity 7 (see also unit)
— triangular — 83
— arrow 7, 8, 10
— function 9
— functor 14, 43, 165
— natural transformation 43 ff.
Image 196
Inclusion
— function 9
— functor 15
Induced map 34
Infinite
— coproduct 64
— product 69
Initial functor 214
Initial object 20, 229
element of — 116, 231
Injection 15, 19
— of coproduct 63
Injective
— monotone function 172
— object 114
Insertion function 9
Integral 219
double — 226
iterated — 226
Interchange
— law 44, 134
— of ends (Fubini) 237
Internal hom functor 180
Intersection of subobjects 122
Intertwining operator 41
Inverse 14
left adjoint-left — 92
left or right — 19
two-sided — 14
— limit 68
Invertible arrow 19
Isomorphism 14
natural — 16
reflection of — 150
of categories 14, 90
of objects 19
Iterated integral 226

Join 110, 122

Kan extensions 229 ff.
absolute — 245
additive — 238
left — as coends 238
pointwise — 240
right — 232 ff., 240
Kellerification 182 ff.
Kelley spaces 181
Kernel 187
difference — 70
pair 71
Kleisli category of a monad 143

Large category 23
Least upper bound 63
Left
— action 170
— adequate functor 246
— adjoint 38, 79, 83, 229
— adjoint — inverse 92
— adjoint-right-inverse 130
— adjunct 79, 230
— cancellable 19
— exact functor 197
— inverse 19
— Kan extensions 238, 244
— regular representation 170

Lemma
Five — 201
Short five — 198
snake — 202
Yoneda — 61

Length of words 161
Limit 68, 76, 109, 211, 229, 234
creation of — 108
direct — 67
filtered — 212
finite — 109, 211
inductive — 67
inverse — 68
pointwise — 112, 233
preservation of — 112
projective — 68
— object 68
— of a natural transformation 224

Limiting cone 67, 68, 114, 232
Linear order 11
Locally small category 127
Loop space 185

Map (see also arrow)
canonical — 165, 211, 242
continuous — 13, 153, 181
diagonal — 73, 192
evaluation — 96
fiber — 71
homotopic — 13
structure — of algebras 136, 139, 148
— of adjunctions 97
Matrices 11, 74, 192
Matrix multiplication 74, 192
Meet 110, 122
Member 200
Metacategory 7 ff.
Metagraph 7
Modules 12, 28, 138
Monad 133, 135, 176
Codensity 246 ex.
free group — 135
multiplication of — 134
unit of — 134
— defined by adjunction 135
Monadic 139
Monic
split — 20
— arrow 19
Monad 2, 4, 11, 75, 134, 166
free — 51, 168
universal — 157
Monoidal categories 4, 157, 158, 165
free — 168
strict — 157, 171
symmetric — 180
Monotone function 15 ex., 172
Morphisms 9
— of arrows 9
— of categories 13
— of functors 16
— of graphs 48
— of monoidal categories 162
— relaxed — 160
strict — 160
— of short exact sequences 198
— of simplicial objects 174
— of T-algebras 136
Multiplication
— in a monad 134
— of monoidal categories 158, 161

Natural 38, 217, 234
components of — transformation 16
Saunders Mac Lane

Saunders Mac Lane, born in Connecticut in 1909, studied mathematics at Yale, Chicago, and Göttingen and has taught mathematics at Harvard, Cornell, and the University of Chicago. He is now the Max Mason Distinguished Service Professor of Mathematics at the University of Chicago. With Samuel Eilenberg he discovered the theory of categories and some of the basic ideas of homological algebra. His book *Homology*, covering the latter subject, was published by Springer-Verlag in 1963. The present book, summarizing the ideas of category theory, is based on his lectures at Chicago, Canberra, Bowdoin, and Tulane.